

2024-阿里巴巴：CDAnet一种基于显式特征增强的跨域CTR预测模型



SmartMindAI

专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注

已关注

28 人赞同了该文章

Introduction

点击率（CTR）预测是在线服务中的关键环节，涉及评估用户对候选内容的点击概率。对于提升单域CTR模型，数据稀疏性是一个主要挑战。因此，目前研究致力于使用相关领域数据以增强目标域的CTR预测能力。近年来，跨域点击率预测（CDCTR）取得了进展，主要分为两类方法：联合训练和预训练与微调。联合训练通过整合不同域的CTR目标，共用参数并共享知识，如MiNet、DDTCDR、STAR等。尽管这些方法旨在促进知识转移，但存在优化阶段目标冲突的问题，可能导致负向效果。

预训练与微调方法涉及两步流程：首先在源域⁺上分阶段训练模型，然后在目标域上继续训练，利用预训练模型参数提升目标域性能。此方法有助于减少负向转移效应，且在资源利用方面更高效。在工业场景中，这种方法尤其适用于淘宝平台等大型在线服务。

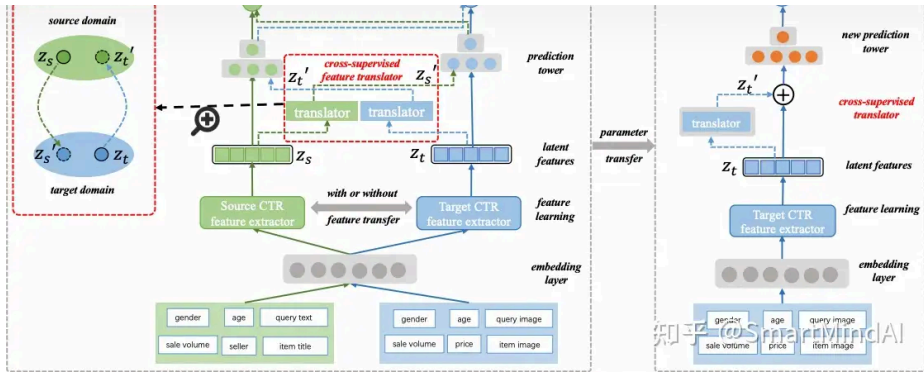
在KEEP方法中，首先从“超级域”（包含目标域用户或商品的极大规模域）中提取用户兴趣知识，然后通过知识提取模块将这些知识传递至目标点击预测模型⁺。KEEP有效利用了大量超级域知识，并在实际应用中取得了显著的性能提升。首先，通过设计的转换网络捕获不同域间的输入特征，包括用户特征和内容特征等，并在潜在空间中学习特征转换。随后，跨域增强网络通过增强目标域样本的额外转换潜在特征，将源域知识编码进入目标模型。这一过程确保了从源域学习到的知识被明确定义、多样化地应用，极大地丰富了目标域输入信息的上下文意义，从而在目标模型的微调阶段实现了优越的性能提升。

相较于其他方法，CDAnet在执行转换学习时对用户或内容不存在重叠需求，因此无需庞大的超级域数据集，展现出了高度的灵活性。此外，CDAnet还实现了高效的知识转移，通过共同转换目标域的各类特征，确保了用户兴趣在微调过程中保持敏感性。实验结果证实，CDAnet显著提升了点击率预测性能，在淘宝APP的在线A/B测试中，用于图像到商品检索任务，实现了绝对0.11个百分点的点击率提升、相对0.64%的交易量增长和相对1.26%的GMV增长，其成果已成功应用于为数亿消费者服务的在线环境中。

Method

Preliminary

知乎



在CDCTR预测任务中，源域 \mathcal{S} 与目标域 \mathcal{T} 各拥有各自的一组训练数据，包含特征 \mathbf{x}_s 与 \mathbf{x}_t ，以及点击标签 y_s 和 y_t 。特征 \mathbf{x}_s 与 \mathbf{x}_t 分别属于 $\mathbb{R}^{F_s \times 1}$ 与 $\mathbb{R}^{F_t \times 1}$ 标记为 $\{0, 1\}$ 的点击标签用于区分是不是真正的点击行为。

在针对此问题设计的CDAnet架构中，包含两个按序列运作的子网络-----翻译网络与增广网络。飞行网络首先将输入数据编码为潜在特征，并学习用于跨域转换的特性翻译机制。在翻译网络完成预训练后，其参数被应用至后续的增广网络中。增广网络接下来通过结合转换后的潜在特征与目标域样本的原始潜在特征，来加强目标域模型的训练与预测能力。

Translation Network

嵌入层、特征提取器、跨监督特征翻译器以及预测塔。

嵌入层:

此层负责将多样性的输入特征编码为有效的嵌入向量。为了增加效率与处理多种特征类型，嵌入层中定义了共享与非共享的子嵌入表示。共享子嵌入被用于处理多个域中可能共有的特征字段；而对于非共有的特征字段，则采用非共享子嵌入来进行编码。合并这些嵌入操作后的形式化表示被定义为 E 。给定源域的输入 \mathbf{x}_s 和目标域的输入 \mathbf{x}_t ，通过嵌入层，它们将分别被转换为对应领域的嵌入表示形式。

$$\mathbf{e}_s = E(\mathbf{x}_s), \mathbf{e}_t = E(\mathbf{x}_t)$$

在其中 $\mathbf{e}_s \in \mathbb{R}^{d \times 1}$ 和 $\mathbf{e}_t \in \mathbb{R}^{d \times 1}$ 分别代表源域输入 \mathbf{x}_s 和目标域输入 \mathbf{x}_t 对应的 d 维嵌入向量。

特征提取器:

在经过嵌入层的转换后，通常伴随着一个用于提取高级特征以进行用户行为建模的模块。在跨域点击率转移（CDCTR）中，这种模块同样是必不可少的组件，它能够与一些转移学习策略结合使用，例如在文中提及的基本共享多层感知器（MLP）和混合专家方法。这些策略在某种程度上促进了领域间知识的转移。

对于基本共享MLP，其主要优点在于其简洁性和计算效率，但也存在当优化不同目标时可能出现的梯度干扰问题。而混合专家方法能有效解决这一问题，但计算成本相对较高。所提出的跨域网络（CDAnet）在选择特征提取模块的灵活性上与其他领域兼容，它可以选择基本共享MLP或混合专家方法。这就使得它能够与现有的隐式知识转移技术相结合。为了一目了然，我们将源域和目标域的特征提取器分别表示为 F_s 和 F_t 。这样，得到的结果为：

$$F_s(\mathbf{e}_s), F_t(\mathbf{e}_t)$$

$$\mathbf{z}_s = F_s(\mathbf{e}_s), \mathbf{z}_t = F_t(\mathbf{e}_t)$$

用户在特定内容上的偏好行为在跳转到源域后会得以保留。例如，如果一位用户热爱科幻电影的类型，那么他在科幻小说领域也表现出相似的偏好。这就要求目标域样本，通过其特征转换到源域的方式保持原有标签，确保其在源域中的行为预测能够反映出与目标域的一致性。

$$\min_{W_t^{tran}} BCE(\hat{y}_t, y_t)$$

知乎

$$\min \mathcal{L}_t^{cross} = BCE(\sigma(R_s(z'_t)), y_t), \quad z'_t = W_t^{tran} z_t$$

$$\min \mathcal{L}_s^{cross} = BCE(\sigma(R_t(z'_s)), y_s), \quad z'_s = W_s^{tran} z_s$$

其中, R_t 是目标领域的预测模块, 而 $W_s^{tran} \in \mathbb{R}^{d \times d}$ 是源领域的转换器。 R_s 和 R_t 的网络结构均采用常见的MLP层。为了更有效地进行潜在特征转换, 除了上述的交叉监督外, 我们还对转换器 W_s^{tran} 和 W_t^{tran} 施加了正交映射约束:

$$\min \mathcal{L}^{orth} = \|\mathbb{T}(W_s^{tran})(W_s^{tran} z_s) - z_s\|_F^2 + \|\mathbb{T}(W_t^{tran})(W_t^{tran} z_t) - z_t\|_F^2$$

在其中 \mathbb{T} 代表转置运算。于数学里, **正交变换⁺** 所具备的功能是维持向量间的长度与角度不变。因此, 方程式中所加入的正交映射限制, 对潜在特征 z_t 进行翻译操作时, 有助于维持各样本特征间的相似性, 防止在翻译后使原本分散的 z_t 聚集到单一位置。

Objective Function in Translation

除学习转换器的目标外, 我们还需在每个领域中优化原始的点击率 (CTR) 任务。具体而言, 源领域和目标领域的原始CTR优化目标可表示为:

$$\min \mathcal{L}_s^{vani} = BCE(\sigma(R_s(z_s)), y_s), \mathcal{L}_t^{vani} = BCE(\sigma(R_t(z_t)), y_t)$$

我们为了学习用于CTR预测的网络参数 (比如预测塔的参数), 同时优化了源域和目标域的预测。当对每个域的CTR预测进行网络优化时, 翻译器能够适应并学习合理且精确的特征转换方向。因此, 翻译网络的核心目标在于

$$\min \mathcal{L}_{trans} = \underbrace{\mathcal{L}_s^{vani} + \mathcal{L}_t^{vani}}_{\mathcal{L}^{vani}} + \alpha \underbrace{(\mathcal{L}_s^{cross} + \mathcal{L}_t^{cross})}_{\mathcal{L}^{cross}} + \beta \mathcal{L}^{orth}$$

其中 α 和 β 被视为损失权重的**超参数⁺**, 用于调整模型优化过程中不同损失组件的重要性。

Augmentation Network

$$\text{增强后的潜在特征表示为 } z_t^{aug} = z_t \oplus z'_t$$

在这里, 连接操作作用 \oplus 表示。当关注目标域的增强时, 增强后的特征能通过相似的公式来获得。强化的目标函数 通过将增强后的潜在特征 z_t^{aug} 输入到增强后的预测塔 R_t^{aug} 中, 我们随后使用通常的点击率 (CTR) 目标进行微调。这个目标函数具体表示为: $\min_{\theta} L(R_t^{aug}(z_t^{aug}), y)$

其中 θ 代表预测塔的**参数集合⁺** L 是损失函数 z_t^{aug} 是增强后的特征表示 y 是真实的点击率标签。此目标旨在最小化预测的点击率与实际点击率之间的差异, 从而优化我们的目标模型, 使其在目标域上的性能得到提升。

$$\mathcal{L}_{aug} = BCE(\sigma(R_t^{aug}(z_t^{aug})), y_t)$$

在增强阶段, 模型专注于单一的点击率 (CTR) 目标, 这有助于避免多目标优化过程中可能出现的冲突问题。对源域进行增强时, 采取与目标域相同的方法进行优化。在知识转移的过程中, 应用均使用相同的 "beta" 参数以保持一致性。

Experiments and Analysis

Datasets

我们首先在Amazon和Taobao两个公共数据集上进行了实验。在Amazon数据集上, 我们对电影和书籍两个最大的领域进行了实验。电影领域的关键数据包括用户ID、电影ID、电影类型、导演信息和电影名称; 书籍领域的关键信息包括用户ID、书籍ID、类别、作者和书籍名称。对于评分系统, 我们将评分大于3的视为正向反馈, 其余视为负向反馈, 用于点击率预测。电影和书名均通过Stanford大学的Glove-6B模型转换为向量。

息、广告ID、广告类别和品牌信息；而对于推荐领域，我们获得了用户ID、内容ID和类别信息。在Handler层，用户ID和项ID被嵌入为64维的特征，其他离散特征被转换为一次热编码或多元热编码。

Table 1: The statistics of datasets.

dataset	Amazon(movie and book)		Taobao(ad and rec)	
domain	movie	book	ad	rec
#users	29,680	52,690	141,917	186,731
#items	16,494	47,302	165,689	379,817
#input feature dim	8,044	24,466	44,897	5,004
#train samples	1,685,836	7,133,107	3,576,414	12,168,878
#val samples	210,730	891,638	447,052	1,521,110
#test samples	210,730	891,639	447,052	1,511,110
#positive samples	351,216	1,486,064	234,736	855,362

Overall Comparison

我们对几种基本模型在不同公共评估基准上的性能进行了比较，并在研究如何结合不同的特征提取器（如CDAnet，IndepMLP，CDAnet，SharedMLP/PLE/MMOE）时考虑了无参数共享的隐式知识转移技术。

Dataset	Amazon		Taobao	
Source→Target	Book→Movie	Movie→Book	Rec→Ad	Ad→Rec
MLP	0.6595	0.7604	0.6161	0.6865
ShareBottom	0.6604	0.7603	0.6160	0.5020
STAR	0.6503	0.7626	0.6149	0.6795
DDTCDR	0.6636	0.7807	0.6162	0.6756
MMOE	0.7025	0.7899	0.6177	0.6930
PLE	0.6993	0.7846	0.6164	0.6933
CDAnet+IndepMLP	0.7207	0.7735	0.6205	0.7047
CDAnet+SharedMLP	0.7203	0.7847	0.6195	0.7034
CDAnet+PLE	0.7213	0.7838	0.6201	0.7023
CDAnet+MMOE	0.7225	0.7811	0.6200	0.7034

通过表中的结果，我们发现运用不同特征提取器的CDAnet普遍表现出较高的性能。联合训练模型（如ShareBottom、DDTCDR、MMOE、PLE）主要依赖共享嵌入层来传递知识，但其在实际应用中的有效性受限于领域间特征字段重叠较少的情况。此类模型往往存在嵌入参数无法有效共享的问题，从而降低了知识转移的效率。相比之下，CDAnet通过明确的翻译过程，增强目标模型并促进知识传递，使之在性能上保持优越。

另一方面，STAR模型主要通过围绕领域中心的共享网络来转移知识，其效率远不及CDAnet的明确知识翻译理念。在结合隐式知识转移技术的实验中，CDAnet，IndepMLP在Taobao数据集上表现出色，而CDAnet，SharedMLP/MMOE/PLE则较强于Amazon数据集。引入隐式知识转移技术固然有助于约束模型训练，但也可能对适应更多特征的能力产生限制。模型性能的提升关键在于确保对数据场景的适应性，CDAnet则提供了通过灵活选择特征提取器来优化性能的途径。

Production Deployment

为了深入研究CDAnet的效能，我们将其集成至阿里巴巴+Taobao平台的图像产品检索系统。用于训练的集合并记录在之前的章节内。考虑到文本检索与图像检索+领域存在显著差异，尤其是特征描述的不重叠性，我们选择在嵌入层中仅部分共享两个领域间的参数。在嵌入层完成特征映射后，我们利用DCN-v2来提取文本领域的潜在特征，而图像领域则以当前的在线服务模型作为特征提取器。对于两个预测塔，我们都采用了包含两层多层感知机（MLP）的架构，并镶嵌了逻辑映射层。基于实际操作中的经验，我们设定参数 α 和 β 的值为0.01。

Offline Performance on Production Dataset

知乎

著提升7%的模型性能。然而，该过程忽略了目标域的上下文特征，且提取的用户兴趣为静态信息，不受输入变化影响。

Model	AUC	CTR	deal number	GMV
Base	0.7845	9.50%	★	★
KEEP	0.7858(+0.13%)	-	-	-
CDAnet	0.7866(+0.21%)	9.61%(+0.11point)	+0.64%	+1.26%

与此相比，CDAnet模型未出现这些缺点，不仅性能更优，还在Taobao的图像到产品检索任务中实现了绝对的AUC提升，达0.21%，展现显著进展。

Online A/B Test

2022年12月26日至2023年1月16日及2023年2月6日至2023年3月31日，我们在淘宝APP的A/B测试框架内进行了大规模在线实验。实验期间，基线模型仅利用了图像到产品领域的数据进行训练。在线实验的关键评估指标为实际点击率（CTR）、交易次数及商品总价值（GMV）。结果表明，相较于基线模型，CDAnet在CTR上实现了绝对提升**0.11点**，GMV相对增长了**0.64%**，而GMV的增加更是达到了**1.26%**。

View	User Group				
Type	G1	G2	G3	G4	G5
Improv.	+0.09	+0.11	+0.13	+0.13	+0.12

View	Query Category									
Type	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Improv.	+0.14	+0.14	+0.08	+0.12	+0.15	+0.13	+0.15	+0.15	+0.12	+0.14

为了全面评估CDAnet的性能差异，我们还基于不同用户群体和查询类别，在表 中列出了与在线基础模型的CTR提升差距。这些数据显示，CDAnet在所有评估视角下均展现出显著改进。自2023年4月起，CDAnet已被全面部署，服务于数百万消费者。我们观察到在GMV增长显著大于交易量增长的现象，这可能表明较高价格的产品销售占比在提升。为探究这一现象背后的机制，我们按产品价格将商品划分为五个不同的价格级别，并比较了CDAnet与基础在线模型在不同价格段的表现。数据如表所示，对比了CDAnet在曝光产品质量(EPQ)得分和不同价格级别的交易数量上的优势。

Price Level	L1	L2	L3	L4	L5
Rel. Improv. EPQ score(%)	+2.10	+3.62	+3.60	+3.43	+2.68
Rel. Improv. of deal(%)	+0.08	+0.12	+0.11	+0.15	+0.17

结果显示，CDAnet在高价格级别上提高了EPQ得分，并显著增加了高价值商品的交易频次。通过这一发现，我们识别出了图像到产品检索面临的挑战:高价值商品（如Beats耳机、iPhone）的搜索数据相比低价值商品（如杯子、T恤）要少，导致针对这些商品的用户行为数据稀疏。由于积累的大量和丰富数据在传统文本到产品检索中的优势，CDAnet使得我们能够在图像到产品的高价格级别上应用这种知识。因此，高价值商品的EPQ得分显著提高，这促进了更有竞争力的商品排名，从而鼓励更多的交易发生。

Online Ranking Example

我们通过几个在线排名案例对比了基线模型和CDAnet，以直观验证CDAnet在提升排名性能上的优势。

知乎



结果如附图所示。在图(a)中，用户上传了一张儿童牛奶的产品图片，基线模型将排名第二的商品误认为与儿童糖果（使用红色方块标记）非常相似，而非预期的产品。而CDAnet则能够准确地识别并呈现实际需求的产品，前五名排名均为儿童牛奶相关。我们推测，在文本域中，儿童牛奶与糖果有着明确的区分，但在图像处理中，这一区分可能变得更加模糊，从而影响了基线模型的识别。通过比较CDAnet与基线模型在案例(b)中的表现，在用户搜索连衣裙时，CDAnet能够更准确地识别相关商品，并显著提高了排序质量。反之，基线模型在第五位推荐的是一条裙子（同样使用红色方块标记），相较于目标产品存在较大偏差。通过这样的案例分析，我们可以清晰地看到CDAnet通过共享文本到产品域的知识，在图像到产品域中实现了更精确的排名，进而提升了用户体验和效率。

发布于 2024-07-01 11:09 · IP 属地北京

迁移学习 (Transfer Learning) ctr预估 阿里巴巴集团

赞同 28 添加评论 分享 喜欢 收藏 申请转载



理性发言，友善互动



还没有评论，发表第一个评论吧

推荐阅读

Figure 5: Illustration of adaptive activation in DIN. Behaviors with high relevance to candidate ad get high activation weight.

广告CTR预估 | KDD2018: 详解阿里DIN(Deep Interest...

魔法学院的... 发表于推荐系统...

Model for Personalized Click-Th

Lyu, Yu Dong, Chengfu Huo, Weijun F
Alibaba Group
:l.z, dongyu.dy, chengfu.huocf, afei}@alibaba-inc

排序专题 (一) [阿里DMR模型]Deep Match to Rank...
Upper 发表于广告排序算...

2018阿里CTR预估模型---DIN (深度兴趣网络), 附TF2.x...

潜心

阿里2022-CTR预测模型 (CTNet): 深度解析跨

SmartMindAI