

阿里2022-CTR预测模型（CTNet）：深度解析跨领域CTR预测的创新方法与拓展应用领域



SmartMindAI

专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注我

已关注

23 人赞同了该文章

论文：Continual Transfer Learning for Cross-Domain Click-Through Rate Prediction at Taobao

Introduction

本文总结了点击率⁺预测任务中跨域CTR预测的重要性，介绍了联合学习和预训练&微调方法两种主要方法，并提出了持续传输学习（CTL）作为更高效的跨域CTR预测策略。然而，CTL方法面临大规模历史目标域数据需求的问题。本工作的主要贡献包括：

- 研究了持续迁移学习（CTL）这一重要且较少研究的问题，将其应用于多个真实的工业应用中。CTL通过优化和持续迭代，可以在不同的任务间进行学习迁移。我们建立了一个基于强化学习的模型，以实现CTL。该模型使用了合适的策略来确保模型的持久性和可扩展性⁺。实验结果显示，CTL能够显著提高工业应用的性能。
- 提出了一种名为CTNet的CTL模型，用于跨领域持续的CTR预测，具有高效性。我们提供了在大型RS中部署CTNet的有用工业经验。模型在数学公式和符号上进行了详细的描述。
- 进行了广泛的离线和在线实验，CTNet显示出高效率 and 优越的性能，相比SOTA方法，它在淘宝网上全面部署带来了显著改进。

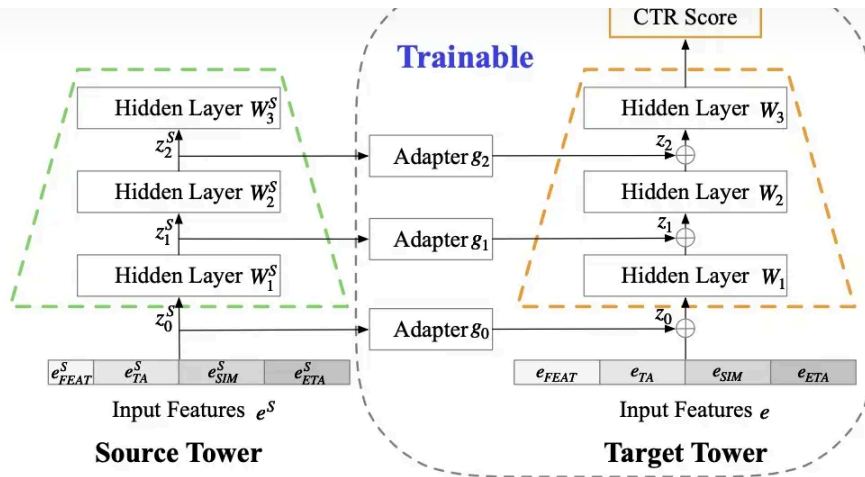


图 1: Architecture of the proposed CTNet.

Methods

Single-Domain Models at Taobao

淘宝推荐系统中的点击率预测模型⁺，每个领域都有自己的模型，通过输入用户-项目特征⁺预测用户点击项目的概率。数值特征经过离散化⁺处理成为分类特征，通过嵌入查找转换为密集嵌入。引入了用户历史行为序列，采用目标注意力、SIM（硬）和ETA对长期序列进行建模，最后将特征嵌入和注意力层输出连接输入到基于MLP的模型中。

CTNet: Continual Transfer Network

我们提出执行持续迁移学习（CTL），并在CTL设置下设计一个CTNet模型，以实现CTR预测。传统上，单域模型在时间段 $t \rightarrow t + 1$ 独立进行训练，使用新的用户反馈数据。在CTL下，模型通过离线增量学习或在线学习使用用户反馈数据。CTNet利用迁移学习策略，从大量预训练模型中学习特征表示，并将其应用于CTR预测任务。CTNet通过在线学习在新的用户反馈数据上进行微调，实现高效准确预测。

持续迁移学习（CTL）的目标是利用源域和目标域的历史和实时知识，以提高目标域上的预测性能。CTL给定一个随时间变化的源域⁺和一个随时间变化的目标域，并利用它们来提高预测性能。

CTNet将目标域模型扩展为双塔架构，通过逐层轻量级适配器连接两个塔，实现域适应。源塔使用源域知识构建，目标塔使用先前目标域模型。目标塔输出隐藏表示的计算公式为

$$\mathbf{z}_l = g(\mathbf{z}_{l-1}) \odot \mathbf{z}_{emb}$$

其中 \mathbf{z}_{emb} 是嵌入层的输出。

$$\mathbf{z}_l = \psi(\mathbf{W}_l \mathbf{z}_{l-1} + g_l(\mathbf{z}_l^S)), \quad l > 0 \quad \mathbf{z}_0 = \mathbf{e} + g_0(\mathbf{e}^S), \quad l = 0$$

本论文利用GLU门控线性单元实现了目标塔与源塔的信息流动控制，具体利用 ψ 表示激活函数⁺， \mathbf{W}_l 表示可训练权重矩阵⁺，并采用 \mathbf{z}_l^S 表示源塔输出隐藏表示。通过门控线性单元实现自适应控制⁺。

$$g_l(\mathbf{z}_l^S) = \mathbf{U}_l^1 \mathbf{z}_l^S \odot \sigma(\mathbf{U}_l^2 \mathbf{z}_l^S)$$

本文使用目标塔输出进行预测的问题，通过二元交叉熵⁺损失函数进行训练，将点击样本视为正样本，查看未点击样本视为负样本。具体地，使用Sigmoid激活函数和逐元素向量乘法，使用可训练矩阵 \mathbf{U}_l^1 和 \mathbf{U}_l^2 。

如何部署CTNet。在图中展示了单域持续学习和跨域持续学习。每个单域模型独立且持续地使用新反馈数据训练。CTNet部署前，目标塔由最新目标域模型初始化。双塔模型设计用于获取细粒度和实时的源域知识。详情见CTNet的描述。我们提出了一种新的方法，通过在线双塔架构来建模实时用户行为。这种方法不使用缓存嵌入，而是呈现完整的在线双塔架构，使用源域参数对用户行为进行建模。这种方法具有更好的适应性，能够适应不断演变的目标域用户行为，并能够编码用户-项目交互的细粒度知识。此外，这种方法还具有更好的性能，因为源域模型比目标域模型更轻、运行更快，不会因并行计算⁺而带来额外的推理延迟。本论文提出一种针对时间效率的特别设计，采用分层适配器实现并行计算，通过共享基于行为序列的特征进一步降低计算成本⁺。我们发现CTNet在与以前的目标域生产模型相比，不会导致额外的在线响应时间，并降低了大部分额外的计算成本。该设计有望提高训练和推理过程的效率。

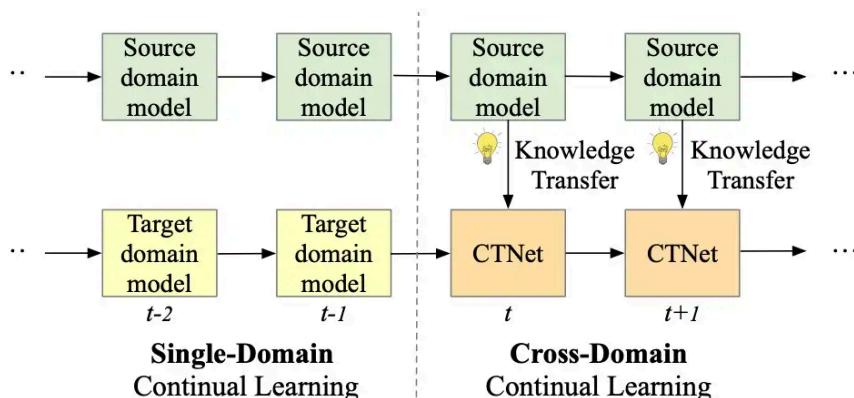


图 2: An illustration of CTNet deployment. 知乎 @SmartMindAI

Experiments

Experimental Details

论文评估了在淘宝生产数据上不同规模的推荐领域之间的持续跨领域CTR预测，并使用抽样流量日志来获得用户-物品交互数据。最后一天的数据用于测试，其他天的数据用于训练。数据集被组织成时间段，每个时间段持续6天，并使用迁移性能来评估模型⁺性能。领域A、B和C的数据集大小分别为约1500亿、20亿和10亿。

对比了各种方法，包括预训练⁺和微调方法，以及联合学习方法，并在生产模型上进行实验以确保公平性。同时，所有单域模型都在淘宝数据上进行了持续训练，并应用了特定的设置，如批次大小、隐藏维度、序列建模方法和优化器等。最后，GAUC被用于评估在线性能。

Offline Experimental Results

对CTNet进行了深入研究，并在大规模工业数据集上验证了其性能。CTNet在单域模型中表现出色，证明了跨域CTR预测的必要性。相比预训练和微调方法，CTNet重用了目标域训练有素的参数，从而减少了信息损失。此外，CTNet捕获了更细粒度和实时的源域知识，优于Extra Embedding。联合学习方法在淘宝工业数据集上的性能较弱，而CTNet利用了源域和目标域的长期知识。消融研究表明基于GLU的适配器性能优于线性层适配器，验证了门控单元特征选择⁺的有效性。CTNet在持续转移和一次性转移两种设置下的性能也得到了验证。结果表明CTNet具有优越的性能和潜力。

	Domain A to B		Domain A to C	
Model	AUC	GAUC	AUC	GAUC
Base (w/o Transfer)	0.7404	0.6788	0.7104	0.6677
Source Model	0.6640	0.6200	0.5783	0.5771
Finetune (Embeddings)	0.7439	0.6828	0.7190	0.6769
Finetune (All)	0.7368	0.6739	0.6727	0.6352
Extra Embedding	0.7402	0.6787	0.7137	0.6693
MLP++	0.7405	0.6764	0.7116	0.6710
Share Bottom	0.7417	0.6777	0.7119	0.6726
PLE	0.7395	0.6749	0.7103	0.6679
MiNet	0.7411	0.6765	0.7129	0.6712
DDTCDR	0.7408	0.6768	0.7118	0.6709
DASL	0.7406	0.6763	0.7132	0.6714
CTNet (w/o GLU)	0.7465	0.6877	0.7432	0.7023
CTNet	0.7474	0.6888	0.7451	0.7040

Production Deployment at Taobao

CTNet在两个大规模推荐系统上部署，分别提高了1.0%和3.6%的GAUC，显著提升了性能。在领域B中，CTNet带来了2.5%的点击率提升和7.7%的总销售额提升；在领域C中，CTNet带来了12.3%的点击率提升和31.9%的GMV提升，在线部署已经完全在线。自2021年12月以来，没有增加计算资源。显著提升了性能。

Table 2: The comparisons between one-time transfer and continual transfer.

Time	Model	AUC	GAUC
$t + \Delta t$	Base (w/o transfer)	0.7483	0.6931
	CTNet (one-time)	0.7548 (+0.65%)	0.7022 (+0.91%)
	CTNet (continual)	0.7548 (+0.65%)	0.7022 (+0.91%)
$t + 2\Delta t$	Base (w/o transfer)	0.7416	0.6788
	CTNet (one-time)	0.7462 (+0.46%)	0.6856 (+0.68%)
	CTNet (continual)	0.7486 (+0.70%)	0.6882 (+0.94%)
$t + 3\Delta t$	Base (w/o transfer)	0.7404	0.6788
	CTNet (one-time)	0.7436 (+0.32%)	0.6861 (+0.73%)
	CTNet (continual)	0.7474 (+0.70%)	0.6888 (+1.00%)

Related Work

点击率预测问题，介绍了深度学习时代提出的各种CTR预测模型。跨域CTR预测是重点，CTNet相较于KEEP具有更简单的架构，通过提取源域知识，节省了预训练的計算和存储资源，并提取更具表达能力的知识。CTNet的在线推理无需缓存源域知识，避免了额外的同步策略，部署更方便。CTNet能够提取源域用户项交互的细粒度和实时知识。持续学习被广泛应用于推荐系统以预测用户未来的兴趣。这个领域的研究主要集中在解决灾难性遗忘问题上，推荐系统+旨在逐步扩展所获得的知识。最近的尝试如持续预训练和持续领域自适应似乎与我们的工作类似，但本文中描述的持续迁移学习（CTL）中，两个域都是随时间变化的。CTL在跨领域设置下尚未进行充分研究。

Conclusions

CTNet，并见证了现有解决方案在系统性能和在线业务指标方面的重大改进。

编辑于 2023-11-05 17:27 · IP 属地北京

ctr预估 推荐系统 工业级推荐系统

▲ 赞同 23 ▼ ● 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

CTR的一些模型方案

Salon... 发表于机器学习的...

主流CTR预估模型的演化及对比

杨木易 发表于算法工程师...

基于深度学习的广告CTR预估算法

摘要：本文主要介绍了广告CTR预估算法在引入深度学习之后的基本演化过程及一些最新的进展，重点是从工业实现和应用的视角对Deep CTR模型进行剖析，探讨为什么这样设计模型、模型的关键要点...

阿里云云栖... 发表于程序员进修...

推荐算法之4——CTR预估

CTR预估模型可以广泛应用于化推荐、信息检索、在线广告域，用来学习和预测用户的行为。用户的反馈主要有点击、收藏、购买等。数据CTR预估模型的特征往往包含多个特征，这些特征...

做大饼馅儿... 发表于人工智能...