arXiv:2305.03953v2 [cs.IR] 9 May 2023

用于点击率预测的跨域增强网络

Xu Chen、Zida Cheng、Shuai Xiao✉、Xiaoyi Zeng、Weilin Huang

摘要——数据稀疏性是点击率（CTR）预测中的一个重要问题，尤其是在用户-物品交互过于稀疏、难以训练出可靠模型时。近年来，为了利用相关领域中的有效数据，许多跨领域点击率预测（CDCTR）研究相继开展。然而，现有的大多数 CDCTR 研究都有一个不切实际的局限：要求不同领域的输入同质（即共享特征字段）。而针对不同领域输入异构（即特征字段不同）的 CDCTR 尚未得到广泛探索，却是一个紧迫且重要的研究问题。本文提出跨域增强网络（CDAnet），能够在输入异构的两个领域之间进行知识迁移。具体而言，CDAnet 包含一个转换网络和一个增强网络，两者按顺序训练。转换网络通过设计两个独立分支，分别计算来自两个异构输入领域的特征，再利用专门设计的交叉监督特征转换器学习有意义的跨域知识。随后，增强网络通过在潜在空间中进行特征转换来编码所学的跨域知识，并对模型进行微调，以完成最终的点击率预测。通过在两个公开基准数据集和一个工业生产数据集上进行大量实验，我们证明 CDAnet 能够学习有意义的转换后特征，并显著提升点击率预测的性能。CDAnet 已在淘宝应用的图像到商品检索中进行了为期 20 天的线上 A/B 测试，使点击率绝对提升**0.11 个百分点**，商品交易总额相对**增长 1.26%**。

索引词—特征转换、跨域点击率预测、异构输入特征

1 引言

点LICK-点击率（CTR）预测用于估计用户点击候选商品的概率，在推荐、检索和广告等在线服务中发挥着重要作用 [1], [2], [3]。例如，作为全球最大的电子商务平台之一，淘宝拥有一系列应用领域，如文本到商品检索和图像到商品检索。每个领域都有自己的点击率预测模型，称为单领域点击率预测模型

预测。已有研究指出，数据稀疏性是严重限制单领域点击率模型提升的关键问题 [4], ，近期的研究致力于利用其他相关领域的数据来改进点击率模型 [4], [5], [6], [7], [8]。

近年来，各种跨领域点击率预测（跨域点击率预测，CDCTR）方法相继被提出，大致可分为两类：联合训练和预训练与微调。联合训练方法通过将不同领域的多个点击率目标结合到单个优化过程中来实现。该方法通常使用共享的网络参数建立连接，并在不同领域之间迁移所学习的知识，例如MiNet [5]、DDTCDR [9] 和STAR [8]。然而，由于两个领域通常具有不同的优化目标，这些方法往往会受到优化冲突问题的影响，进而可能导致负迁移结果 [10],[11]。为了解决这一问题，近期已有一些方法



(a) 文本到商品检索　　(b) 图像到商品检索

图 1：淘宝应用中文本到商品检索和图像到商品检索的示例。

已有研究提出了用于联合训练多个点击率目标 [12], [13], [14] 的方法。另一方面，预训练与微调方法通常分为两个阶段：首先在源域中依次训练点击率模型，然后在一个.....

• *Xu Chen*、*Zida Cheng*、*Shuai Xiao*、*Xiaoyi Zeng* 和 *Weilin Huang* 均就职于阿里巴巴集团。电子邮箱：{*huaisong.cx*、*chengzida.czd*、*shuai.xsh*、*weilin.hwl*}*@alibaba-inc.com*，*yuanhan@taobao.com*。 ✉ 表示通讯作者

目标域，在目标域中，通常可以通过利用从源域预训练得到的模型参数来提升性能。请注意，每个训练阶段仅使用一个目标进行优化，这显著降低了负迁移的影响。因此，该方法已广泛应用于淘宝平台 [6], [7]。

值得注意的是，近期大多数跨域点击率预测方法都致力于探索具有同构输入特征 [5], [6], [7], [8], [12], [13], 的额外跨域数据，这意味着不同领域之间共享特征字段。例如，MiNet [5] 和 MMOE [12]都具有用于在不同领域之间迁移知识的共享输入嵌入层。特别是， [6], [7] 中提出的近期预训练与微调方法也要求使用同质输入，使其能够将目标域数据直接应用于预训练源模型，从而生成用于微调目标模型的额外特征。然而，对同质输入的要求可能会严重限制其在实践中的应用。许多跨域场景通常存在异构输入特征，这意味着两个领域具有不同的特征字段。例如，在淘宝的图像到商品检索中，图像是重要的查询信息，而文本则是文本到商品检索中的关键查询信息。图像检索系统高度依赖图像特征，而文本检索系统则非常注重文本特征。在淘宝应用中，文本到商品检索已经发展了多年，所产生的数据量比图像到商品检索多出一个数量级。理想情况下，可以执行知识迁移，从文本到商品域中学习用户行为，然后将其应用于提升图像到商品检索的性能。

这启发我们开发一种新的跨域点击率预测方法，支持异构输入特征。这一方向尚未得到广泛研究，但在工业应用中亟待解决，且颇具挑战性。一种直接的解决方案是在近期工作 [9], [12], [13], 的基础上，将共享嵌入层替换为多个域专属层，分别处理异构输入；然而，这不可避免地会破坏专为知识迁移设计的关键模块，导致性能不尽如人意。最新的 STAR [8]能够处理异构输入，但它主要依靠参数共享来迁移知识，同时在学习过程中涉及多个点击率目标，这可能导致知识迁移效率低下。

本文提出了一种新颖的跨域增强网络（CDAnet），由转换网络和增强网络组成。它依次执行：首先学习跨域知识，然后通过跨域增强将所学知识隐式编码到目标模型中。具体而言，转换网络旨在分别处理来自不同领域的输入（从而支持异构输入特征），并通过设计的交叉监督特征转换器，在潜在特征空间中学习知识转换。随后，增强网络通过在潜在空间中为目标领域样本添加额外的转换后潜在特征，执行跨域增强。这一过程将从源域学到的知识隐式编码其中，为目标模型的微调提供多样且有意义的额外信息。图 2 展示了 CDAnet 与其他方法的学习过程对比。实验结果表明，CDAnet 能显著提升点击率预测性能；该方法已在淘宝应用的图像到商品检索中进行了线上 A/B测试，使点击率绝对提升**0.11 point**，商品交易总额相对增长**1.26%** 。该方法已成功部署上线，服务数十亿消费者。总而言之，本文的贡献总结如下：

- 我们提出一种支持异构输入的新型跨域点击率预测任务，使目标模型能够从不同领域学习有意义的补充知识。这能有效缓解点击率预测中的数据稀疏问题，也使其区别于大多数仅支持同质输入的现有跨域方法。

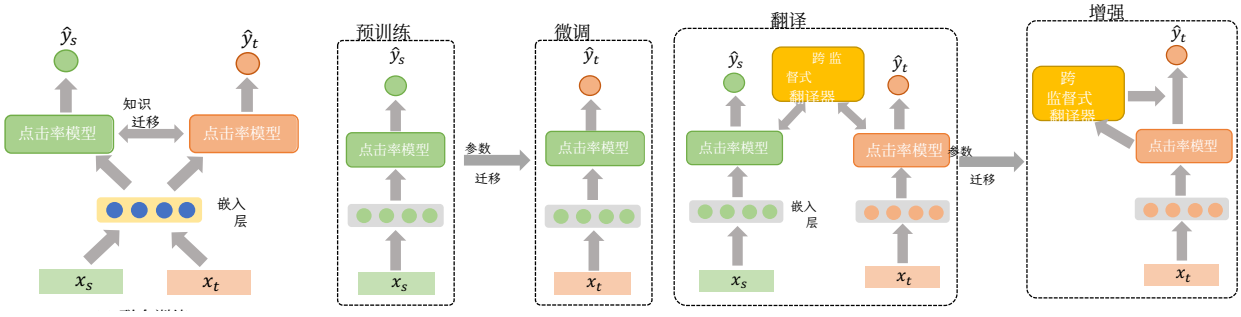
- 我们提出跨域增强网络（CDAnet），由专门设计的转换网络和增强网络组成。转换网络能够通过分别计算两个领域的特征，从异构输入中学习跨域知识。随后，在目标模型微调期间，它通过特征增强在潜在空间中隐式编码所学知识，从而充当高效的领域转换器。

- 在多个数据集上进行的大量实验验证了所提出方法的有效性。我们的 CDAnet 已部署于淘宝应用的图像到商品检索中，并显著提升了点击率和商品交易总额。此外，通过实证研究，我们表明 CDAnet 能够学习有意义的翻译后的特征，从而进一步提升点击率。

2 相关工作

2.1 点击率预测

单领域点击率：点击率（CTR）预测在各种在线服务中发挥着至关重要的作用，例如现代搜索引擎、推荐系统和在线广告。以往的研究通常将逻辑回归 [15]与特征工程结合起来进行点击率预测。这些方法往往缺乏建模复杂特征交互的能力，而且严重依赖人工设计特征。得益于深度神经网络（DNN）出色的特征学习能力，深度学习方法已被广泛用于点击率预测，近期的研究则聚焦于应用 DNN 学习特征交互，例如 Wide&Deep [1], 、DeepFM [16] 和 DCN [17]。例如，在 [1], 中，Cheng 等人结合浅层线性模型与深层非线性网络，以捕捉低阶和高阶特征；而在 [16]中，则将因子分解机 [18] 与深度网络相结合来进行点击率预测。在 DCN [17],中，Wang 等人设计了深度交叉网络，以学习有界阶数的特征交互。此外，深度模型在为点击率任务建模更丰富的信息方面也展现出强大的能力。例如，DIN [2] 和 DIEN [19] 旨在根据历史点击行为捕捉用户兴趣，



（a）联合训练　　(b) 预训练与微调　　(c) CDAnet

图 2：联合训练、预训练与微调以及我们的 CDAnet 的比较。a) 是联合学习方案，其中嵌入层通常是共享的，并在两个点击率领域之间采用知识迁移技术。b) 展示了预训练与微调的学习方式：首先训练一个源域点击率模型，然后目标域模型加载预训练的源域模型参数，并根据目标领域目标函数进行微调。c) 展示了 CDAnet 的学习方式。首先，转换网络通过交叉监督翻译器学习如何将目标域的潜在特征映射到源域。然后，增强网络复用预训练参数，并将转换后的潜在特征作为额外信息，用于对目标领域点击率预测进行跨域增强。

而 DSTN [3] 则在对用户行为建模时纳入了上下文广告。

跨领域点击率预测：单域点击率预测面临数据稀疏问题，因为现实系统中的用户行为通常极为稀疏。因此，跨域点击率预测应运而生，旨在利用相关但数据丰富的领域中的用户行为，促进当前领域的学习。近期，一种联合学习方法得到发展，并成为跨领域点击率预测的代表性方法。例如，STAR [8]是一种星型拓扑模型，包含一个由不同领域共享的中心网络，以及多个针对各领域定制的领域专属网络。在 MiNet [5], 中，Ouyang 等人尝试利用源域中的辅助数据（例如历史用户行为和广告标题）来提升目标域的性能。与此同时，许多跨域推荐（CDR）方法也已得到发展 [9], [20], [21], [22],，并可自然地应用于跨域点击率预测问题。例如， [20], 提出了一种深度交叉连接网络，能够在不同领域之间迁移用户评分模式。在 DDTCDR [9], 中，Li等人提出了一种深度双向迁移网络，能够以迭代方式在领域之间双向迁移信息。然而，这类模型在联合学习期间有两个不同的点击率目标，这可能导致梯度干扰问题 [10], [23],[24]。为克服这些局限，MMOE [12] 应运而生，它通过采用共享混合专家模型和任务专属门控网络，为不同任务学习自适应特征选择。该模型显式地动态建模任务关系，从而自动分配模型参数，缓解优化过程中的任务冲突。为了更明确地解耦任务专属信息与任务共享信息，PLE [13]将网络划分为任务共享组件和任务专属组件，并采用渐进式路由机制，逐步提取和分离更深层的语义知识。

除联合训练外，预训练与微调也是一种广泛应用的两阶段学习范式。在预训练阶段，模型首先在源域中进行训练。随后，在微调阶段，目标模型加载预训练模型的参数，再针对目标领域点击率预测进行微调。在每个阶段，优化时仅使用一个目标，因此可以在一定程度上缓解梯度干扰问题。研究表明，这种方法在工业系统中可能更高效、更有效 [25]。近期，Zhang等人提出了 KEEP [7] ，这是一种两阶段框架，由一个基于网络规模的长期数据运行的监督式预训练知识提取模块，以及一个将提取的知识融入下游微调模型的插件网络组成。在 CTNet [6],中，Liu 等人关注时序演变场景下的跨领域点击率预测问题。此外，近期还有许多工作（例如 [26], [27]）研究网络的不同层，并采用多种迁移方式（例如线性探测和微调），以更高效地进行知识迁移。

大多数跨域点击率预测工作都专注于在不同的用户行为分布之间迁移知识，却忽略了输入特征的差异，因此通常在跨域同构输入 [6], [7], [12], [13]方面存在明显局限。相比之下，本文提出的 CDAnet 在学习过程中同时考虑用户行为分布和输入特征的差异。它允许异构输入，并设计了新颖的翻译网络和增强网络，以便在这种场景下实现更有效的知识迁移。

2.2 翻译学习

翻译是一项将数据从源域翻译到目标域，同时保留内容信息的任务，其中有两个热门研究方向：图像到图像翻译 [28] 和神经机器翻译 [29]。由于神经机器翻译处理序列数据，并

与我们的工作并不完全相关，因此我们将在此主要介绍图像到图像翻译。在早期的图像到图像翻译研究中，研究人员通常使用对齐的图像对来训练翻译模型。Pix2pix [30]是首个基于条件 GAN [31]的统一图像到图像翻译框架。然而，配对图像难以获取，因此研究者提出了非配对图像到图像翻译网络 [32], [33], [34], [35]。在非配对图像到图像翻译中，一些工作提出了保留源域数据图像属性的约束，例如语义特征 [36] 和类别标签 [37]。另一种广为人知且有效的约束是循环一致性损失 [33], [34], [35]：如果将输入图像翻译到目标域，再翻译回来，就可以得到原始输入图像。尽管图像到图像翻译领域已有多种研究，但其中不可或缺的一部分是跨域翻译，即将源域图像转换为目标域图像。我们的 CDAnet 也采用了类似的交叉监督翻译器。不过，关键区别在于，CDAnet 并不旨在获得数据空间中的转换结果，而是学习映射后的潜在特征，以增强目标域中的点击率预测。

3 问题形式化

在具有异构输入特征的跨域点击率预测中，给定源域 $\mathcal{S}$，其训练样本为$(x_s,\ y_s)$，其中 $x_s \in \mathbb{R}^{F_s \times 1}$ 表示输入特征，$y_s \in \{0,1\}$ 表示点击标签（即 1 表示点击，0 表示未点击）。同理，我们有目标域 $\mathcal{T}$及其训练样本$(x_t,\ y_t)$，其中 $y_t \in \{0,1\}$是目标域中的点击标签，$x_t \in \mathbb{R}^{F_t \times 1}$ 是输入特征。F_s 和 F_t 分别表示源域和目标域的输入特征维度。当 x_s 和 x_t具有不同的特征字段时，我们将它们定义为异构特征。例如，在文本到商品检索中，文本查询非常重要且应用广泛；而在图像到商品检索中，我们没有文本查询，但有图像查询，且图像特征与文本特征差异很大。当两个领域具有异构输入特征时，许多近期研究无法妥善处理这种情况，因此如何弥合异构差距以实现高效的知识迁移颇具挑战性。

4 方法

4.1 概述

使用异构输入特征进行跨域点击率预测建模的关键，在于确保模型既能嵌入异构输入，又能具备良好的知识迁移能力，即使不采用近期研究中广泛使用的共享嵌入层技术 [5], [6], [12]。因此，我们提出了 CDAnet，它由两个按顺序训练的网络组成：转换网络和增强网络。首先，转换网络使用解耦嵌入层嵌入异构输入特征，并通过设计的跨监督式翻译器学习潜在特征之间如何相互转换。随后，将转换网络预训练得到的参数迁移至增强网络。接下来，增强网络将转换后的潜在特征与目标域样本的原始潜在特征结合起来，以微调目标域模型。模型架构如图3所示。以下部分将介绍各个模块的详细信息。

4.2 转换网络

不同于旨在数据空间中转换图像的图像到图像翻译，我们的转换网络学习在不同领域之间转换潜在特征。它包括四个部分：解耦嵌入层、基于MMOE的特征提取器、交叉监督翻译器和预测塔。

4.2.1 解耦嵌入层

由于输入特征具有异质性，因此有必要将两个领域的嵌入层解耦，使嵌入层能够适应其所属领域中的输入模式。给定源域的输入 $x_s \in \mathbb{R}^{F_s \times 1}$ 和目标域的输入 $x_t \in \mathbb{R}^{F_t \times 1}$，以及嵌入层 H_s和 H_t，我们有：

$$h_s = H_s(x_s),\ h_t = H_t(x_t) \tag{1}$$

其中，$h_s \in \mathbb{R}^{d \times 1}$ 和 $h_t \in \mathbb{R}^{d \times 1}$ 分别是 x_s 和 x_t的嵌入特征。d 是潜在特征维度。H_s 和 H_t 是特定领域嵌入层。

4.2.2 基于*MMOE*的特征提取器

在转换网络中，源域和目标域的两个点击率目标会联合优化。由于数据分布不同，它们的优化方向也不同。若忽略这一点，可能会导致优化冲突问题，并造成负迁移。在此，我们引入一个共享的基于MMOE的网络，为每个领域提取有用特征，使每个领域都能自动拥有自己的参数。我们将 F_i 记为具有 L层非线性多层感知机的第 i个专家，则有：

$$f_s^i = F_i(h_s),\ f_t^i = F_i(h_t) \tag{2}$$

其中，$f_s^i \in \mathbb{R}^{d \times 1}$ 和 $f_t^i \in \mathbb{R}^{d \times 1}$ 分别是源域和目标域中第 i个专家的输出。令 G_s 和 G_t 分别为源域门控和目标域门控，它们是具有 $\mathbb{R}^d \to \mathbb{R}^d$的 L 层非线性多层感知机，则有：

$$g_s = \text{softmax}(W_s^{gate} G_s(h_s)),\ g_t = \text{softmax}(W_t^{gate} G_t(h_t)) \tag{3}$$

其中，$W_s^{gate} \in \mathbb{R}^{K \times d}$，$W_t^{gate} \in \mathbb{R}^{K \times d}$为仿射变换，$K$ 表示专家数量。$g_s \in \mathbb{R}^{K \times 1}$和 $g_t \in \mathbb{R}^{K \times 1}$ 分别为源域门控和目标域门控的输出。根据公式 3，源域和目标域的点击率目标可以自动选择各自的参数进行优化。为此，我们可以得到两个领域的潜在特征，分别为 $z_s \in \mathbb{R}^{d \times 1}$ 和 $z_t \in \mathbb{R}^{d \times 1}$:

$$z_s = \sum_{i=1}^{K} g_s^i f_s^i,\ z_t = \sum_{i=1}^{K} g_t^i f_t^i \tag{4}$$

借助这个基于MMOE的特征提取器，一方面，参数共享有助于知识迁移。另一方面

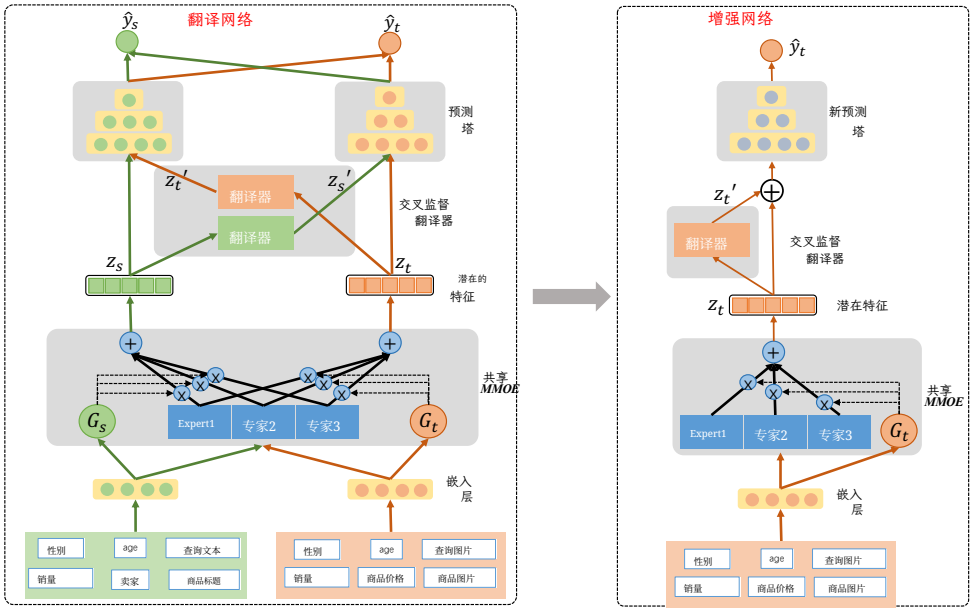


图 3：我们提出的跨域增强网络（CDAnet）的架构。首先，转换网络主要通过交叉监督来学习翻译器。然后，将转换网络中除塔层之外的参数迁移到增强网络。之后，增强网络将目标域样本及其转换后的潜在特征结合起来，用于增强目标域模型的微调。

，其架构能够有效缓解优化冲突问题 [12]。该基于MMOE的网络被置于嵌入层之后，这是因为我们认为，低层中的特征更加丰富且具有更强的泛化性，而高层中的特征则更加紧凑和专门化。前者包含更多可迁移的模式，能够使下游任务受益。

4.2.3 交叉监督翻译器

在获得各领域的潜在特征后，我们提出一种交叉监督翻译器来学习潜在特征转换，其灵感来自图像到图像翻译。在图像到图像翻译中，目标域中的图像被转换为源域中的图像，并由源域中的真实图像进行监督。而在跨域点击率预测中，目标域中的样本（例如，组合后的用户特征和物品特征）可能与源域中的样本不共享用户行为标签，因为这两个样本并不存在配对关系。相反，跨域点击率预测领域普遍认为，如果用户喜欢目标域中的某个物品，那么当用户和物品被映射到源域中对应的特征时，这种喜好行为也会保留下来。例如，如果用户喜欢科幻电影，那么他或她也往往会喜欢科幻小说。

换句话说，给定目标领域的潜在特征 z_t，我们的目标是将其转换为源域中的 $z_t^{'}$，同时保留 z_t 的语义。然后，可以将 $z_t^{'}$ 输入源域的预测塔，并使用目标领域标签 y_t 进行监督。设 $W_t^{tran} \in \mathbb{R}^{d \times d}$ 为目标领域的翻译器，BCE 为二元交叉熵损失，则目标领域的交叉监督翻译器通过以下方式进行优化：

$$\min \mathcal{L}_t^{cross} = BCE(\sigma(R_s(z_t^{'})), y_t), \quad z_t^{'} = W_t^{tran} z_t \tag{5}$$

其中，σ 是 sigmoid 函数，R_s 是源域的预测塔。为稳定训练，我们对源域采用对称形式，如公式 5 所示：

$$\min \mathcal{L}_s^{cross} = BCE(\sigma(R_t(z_s^{'})), y_s), \quad z_s^{'} = W_s^{tran} z_s \tag{6}$$

其中，R_t 是目标域的预测塔，$W_s^{tran} \in \mathbb{R}^{d \times d}$ 是源域的翻译器。需要注意的是，R_s 和 R_t 的网络架构可以根据各自领域的特点采用不同的架构。

为了更好地进行潜在特征转换，除上述交叉监督外，我们还对转换器 W_s^{tran}和 W_t^{tran}添加了正交映射约束：

$$\begin{aligned} \min \mathcal{L}^{orth} = & ||\mathbb{T}(W_s^{tran})(W_s^{tran} z_s) - z_s||_F^2 + \\ & ||\mathbb{T}(W_t^{tran})(W_t^{tran} z_t) - z_t||_F^2 \end{aligned} \tag{7}$$

其中，$\mathbb{T}$ 表示转置操作。数学上的正交变换具有保持向量长度和夹角的特性。因此，式 7 中的正交映射约束有助于使不同样本的潜在特征z_t 保持相似性，并避免多个 z_t 在转换后坍缩到同一个点。

4.2.4 翻译中的目标函数

除了上述学习翻译器的目标外，我们仍需采用常规目标来优化各领域中的点击率任务。具体而言，源域和目标域的常规点击率目标表示如下：

$$\min \mathcal{L}_s^{vani} = BCE(\sigma(R_s(z_s)), y_s), \mathcal{L}_t^{vani} = BCE(\sigma(R_t(z_t)), y_t) \tag{8}$$

这两个目标有助于模型学习用于点击率预测的网络参数（例如塔式网络）。与此同时，当塔式网络经过优化时

对于每个领域的点击率预测，转换器可以适应塔式网络，并沿着有意义且正确的方向学习特征转换。总而言之，转换网络的目标函数为：

$$\min \mathcal{L}_{trans} = \underbrace{\mathcal{L}_s^{vani} + \mathcal{L}_t^{vani}}_{\mathcal{L}^{vani}} + \alpha(\underbrace{\mathcal{L}_s^{cross} + \mathcal{L}_t^{cross}}_{\mathcal{L}^{cross}}) + \beta\mathcal{L}^{orth} \tag{9}$$

其中， α 和 β 是损失权重的超参数。

4.3 增强网络

完成翻译学习后，我们首先将网络参数迁移到增强网络中，包括嵌入层、共享的 MMOE 模块和翻译器。为了给增强网络提供足够的学习灵活性，如图 3 所示，预测塔采用全新初始化，而非沿用翻译网络中的参数。

4.3.1 跨域增强

与此同时，在转换网络中学习到的翻译器可以将潜在特征 z_t 映射到另一个空间，得到 $z_t^{'}$，从而提供更多有用信息。也就是说，我们可以将目标域样本的翻译后特征用作额外信息，以增强目标域模型的训练。目标域的增强后潜在特征表示如下：

$$z_t^{aug} = z_t \oplus z_t^{'} \tag{10}$$

其中， $\oplus$表示拼接操作。对于源域中的增强，也可以通过类似的公式得到增强后的特征。

4.3.2 增强目标函数

利用增强后的潜在特征 z_t^{aug}，我们将其输入新的预测塔 R_t^{aug}，并使用标准点击率目标进行微调。目标函数定义如下：

$$\mathcal{L}_{aug} = BCE(\sigma(R_t^{aug}(z_t^{aug})), y_t) \tag{11}$$

在这一增强阶段，模型只有一个点击率目标，因此可以避免多目标模型中的优化冲突问题。若关注源域的性能，其增强网络也可以采用类似方式进行优化。式 7 中的约束在此处同样适用，其中 β 与式 9 中相同。

从知识迁移技术的角度来看，联合训练方法 [8], [12], [13], [20] 和微调方法 [5], [6], [7] 主要依赖参数共享，而我们的 CDAnet 提出了一种全新的先翻译再增强的思路。此外，CDAnet 支持异构输入特征，而大多数现有的跨域点击率预测模型都要求使用同质输入。

5 实验与分析

5.1 实验设置

5.1.1 数据集

我们首先在两个公开基准数据集上进行实验，这两个数据集的领域具有异构输入特征：亚马逊[1] 和淘宝[2][3]。我们选择亚马逊中规模最大的两个领域——电影和图书进行实验。在电影域中，我们拥有用户 ID、电影 ID、电影类型、电影导演和电影名称信息。而在图书域中，我们拥有用户 ID、图书 ID、图书类别、图书作者和图书名称信息，这些信息与电影域中的信息不同。此外，每个领域都有各自的用户行为。原始用户行为是 0 至 5 分的评分，我们将高于 3 分的评分作为正向反馈，将其他评分作为负向反馈，用于点击率预测。用户 ID 和物品 ID 均嵌入为 64 维特征。电影名称和图书名称均使用 word2vec Glove-6B 模型[4]处理为向量。电影类型、电影导演、图书类别和图书作者分别映射为各领域的独热特征。淘宝数据集包含广告（广告）域和推荐域中的用户-物品交互。我们将购买行为类型视为正向反馈，将其他类型视为负向反馈。在广告域中，我们获取用户 ID、年龄、性别、职业以及其他一些用户资料信息。对于广告侧，我们拥有广告 ID、广告类别和广告品牌信息。在推荐域中，我们获取用户 ID、物品 ID 和物品类别信息，这些特征与广告域中的特征截然不同。对于淘宝数据集，用户 ID 和物品 ID 均嵌入为 64 维特征。其他离散特征处理为独热或多热特征。对于这两个数据集，我们还应用 k-核过滤，以确保每个用户或物品至少有 k 次交互。k 对于亚马逊的电影域和图书域分别为 5 和 10。对于淘宝，k 在广告域和推荐域也分别为 5 和 10。数据集统计信息汇总于表1。

表1：数据集统计信息。

数据集	亚马逊（电影和图书）		淘宝（广告和推荐）	
领域	电影	book	ad	rec
#用户	29,680	52,690	141,917	186,731
项目数	16,494	47,302	165,689	379,817
输入特征维度	8,044	24,466	44,897	5,004
训练样本数	1,685,836	7,133,107	3,576,414	12,168,878
#验证样本	210,730	891,638	447,052	1,521,110
#测试样本	210,730	891,639	447,052	1,511,110
#正样本	351,216	1,486,064	234,736	855,362

此外，由于我们针对的是淘宝场景下具有异构输入特征的真实世界跨域点击率预测问题，因此还在阿里巴巴生产数据以及淘宝移动应用上的在线实验中评估了模型。具体而言，我们收集了文本到商品检索中六个月的用户行为作为源域数据，并收集了图像到商品检索中一年的用户行为作为目标域数据。这两个领域均包含数千亿条样本，以及丰富的用户侧和商品侧输入特征，这些特征均应用于在线生产系统。由于这两种场景各自的特点，两个领域中的输入特征存在显著差异。

1. https://jmcauley.ucsd.edu/data/amazon/ 2 56
3. https://tianchi.aliyun.com/dataset/ .
https://tianchi.aliyun.com/dataset/649 4.
https://nlp.stanford.edu/projects/glove/

5.1.2 基线方法

现有的预训练与微调方法（例如 Keep [7],CTNet [6]）需要共享特征字段来进行知识迁移，因此我们不将它们纳入基线方法。相反，我们采用多种方法进行比较，包括单域方法和联合训练方法：

- MLP。深度多层感知机模型是在线搜索和推荐系统中常见且高效的排序模型。此处，MLP 作为每个领域的单领域点击率模型。
- ShareMiddle。考虑到不同领域的输入特征具有异质性，我们将嵌入层分开，同时借鉴 Share-Bottom [12]技术，共享各领域嵌入层之后的中间层。
- STAR [8]。STAR 是一种星型拓扑模型，它利用所有领域的数据，同时训练一个模型来服务所有领域。
- DDTCDR [9]。这是一种深度双重迁移学习模型，以迭代方式在相关领域之间迁移知识。
- MMOE [12]。MMOE 通过共享的混合专家模块和任务专属门控，对多任务学习中的任务关系进行隐式建模。此处，我们在跨域点击率预测中使用 MMOE，并为不同领域设置专属嵌入层。
- PLE [13]。PLE 是一种多任务学习模型，它将共享组件与任务专属组件明确分离，并采用渐进式路由机制来提取和迁移知识。

5.1.3 参数设置

在公开基准数据集上的实验中，我们按照时间顺序，采用常见的 8:1:1 比例将数据划分为训练集、验证集和测试集。实验重复进行多次，并取平均值作为模型性能。所有模型的潜在特征维度均设为 64。训练轮数设为 200，以确保模型收敛。为进行公平比较，我们对所有模型的超参数和层数进行网格搜索。对于 CDAnet，在亚马逊数据集的两个领域中，α 为 0.01，β 为 0.1，专家数量为 2，且每个专家有 2 层；预测塔有 2 层，其中包括 logit映射层。在淘宝数据集的两个领域中，α为 0.03，β 为 0.1，专家数量为 2，每个专家有 2 层，预测塔有 3 层，其中包括 logit映射层。在我们的生产数据集上，文本到商品域的预测塔是一个 DCN-v2 [38]模型[5] ，而图像到商品领域的预测塔则是一个在线服务模型。根据在公开基准数据集上的经验，我们还将生产数据集中的专家数量设为 2，每个专家设为 2 层，α 和 β 均设为 0.01[6]。

5.我们无法访问文本到商品域的在线生产模型，因此根据经验选择 DCN-v2 作为预测塔网络。

6. 由于时间有限，在该生产数据集上调整超参数会消耗过多计算资源，因此我们没有进行调整。未来的实验将调整不同的超参数。

表2：不同模型在亚马逊和淘宝上的曲线下面积比较。列出了两个领域的结果。

数据集	亚马逊		淘宝	
领域	电影	Book	Ad	Rec
MLP	0.6595	0.7604	0.6161	0.6865
ShareMiddle	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	**0.7225**	0.7811	**0.6200**	**0.7034**

表3：图像到商品领域在我们的生产数据集上的曲线下面积对比结果。这里我们将文本到商品作为源域，并将我们的图像到商品领域作为目标域。

Task	CTR
Base	0.7845
CDAnet	0.7866(**+0.21%**)

5.2 总体比较

5.2.1 离线比较

在本部分中，我们展示了数据集两个领域上的比较结果。遵循已有工作 [6], [8],，我们将曲线下面积作为评估指标。结果如表2和表3所示。

从表2中可以观察到，与各种模型相比，CDAnet通常能够提升点击率性能。联合训练模型（例如DDTCDR、MMOE、PLE）原本不支持具有异构输入特征的跨域点击率预测，而我们通过为其修改并加入特定领域嵌入层，使其能够运行。这些模型的性能不如CDAnet，表明仅通过使用特定领域嵌入层修改现有模型，无法很好地实现跨领域知识迁移。尽管STAR能够处理异构输入，但它主要依赖中心化的领域共享网络来迁移知识，根据表2中的结果，这种方式并不十分有效。相比之下，我们的先翻译后增强理念更具优势，能够带来更好的性能。此外，结合表3中的结果可以看出，即使面对如此大规模的数据，CDAnet仍然能够使点击率任务受益。它使点击率的曲线下面积绝对提升了0.21%。这可能是因为CDAnet能够将文本到商品域中的大规模高质量数据知识很好地迁移到图像到商品领域。

5.2.2 在线 *A/B* 测试

我们进一步在淘宝应用的图像到商品检索 A/B测试框架中进行了为期20天的在线实验。基线模型是仅使用图像到商品数据训练的在线服务模型。在线评估指标为实际点击率和商品交易总额。在线 A/B测试表明，我们的 CDAnet 使点击率绝对提升了0.11个百分点，商品交易总额相对提升了1.26%。此外，我们比较了基线模型与 CDAnet 的一些在线排序示例。结果如图4所示。图中给出了两个示例，用以说明 CDAnet 具有更强的排序能力。在（a）中，如果用户上传一张旺仔牛奶的图片，我们可以看到，在线基线模型将一瓶视觉上相当相似的旺仔糖排在第2位，这相当



（a）旺仔牛奶示例　　（b）连衣裙示例

图4：在线基线模型与我们提出的 CDAnet 的排序示例。给定一个具有相同用户的查询，我们列出不同模型的前5个排序项。（a）是旺仔牛奶的示例，（b）是连衣裙的示例。

相似的商品排在第2位。相比之下，我们的 CDAnet 能够排除这种情况，排名前5的商品全都是旺仔牛奶。一个可能的原因是，文本领域中牛奶和糖果的查询差异很大。文本到商品检索中的点击率模型能够很好地捕捉这种差异，而图像到商品领域中的点击率模型则容易受到视觉模式的干扰。我们的 CDAnet 能够将文本到商品域的知识迁移到图像到商品领域，从而改善排序结果。在（b）中，我们展示了一个连衣裙的排序示例。虽然我们的数据库中没有与查询完全相同的商品，但 CDAnet 更擅长对类似连衣裙的商品进行排序，而在线基线模型将一条裙子排在第5位。这些示例有助于我们更好地理解 CDAnet 的优势。

5.2.3 不同稀疏数据

为了研究 CDAnet 在更稀疏条件下的表现，我们开展了一项实验，以考察训练数据在不同稀疏度下的模型性能。具体而言，在固定验证集和测试集的情况下，我们改变原始训练数据与新训练集的比例。

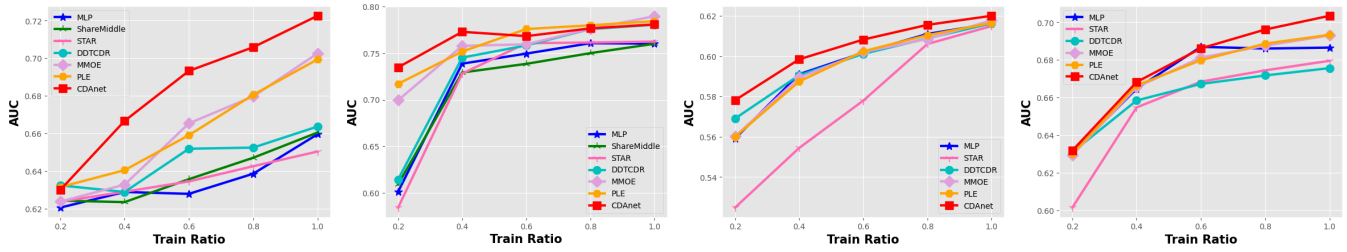
结果如图5所示。我们可以看到，与其他跨领域点击率预测方法相比，CDAnet 在几乎所有稀疏度下都能取得更好的结果。CDAnet 采用了更先进的知识迁移方式，同时考虑了异构输入和优化冲突问题，因此能够展现出更强的泛化性能。

5.3 迁移特征的语义

CDAnet 学习如何在不同领域之间转换潜在特征，并利用转换后的潜在特征作为额外信息来提升目标领域点击率预测，因此值得探究转换后的特征对于目标领域点击率预测是否具有实际意义。在本部分中，我们设计了一项实验，以考察潜在特征在转换前后是否具有相关语义。具体而言，对于亚马逊数据集，给定一个在两个领域中都具有共享正标签训练实例的用户[7]，，我们可以获得图书域的潜在特征矩阵 $Z_b \in \mathbb{R}^{N_b \times d}$ 和电影域的潜在特征矩阵 $Z_m \in \mathbb{R}^{N_m \times d}$。然后，我们可以通过学习得到的翻译器转换 Z_m 的特征，并将其记为 $Z_m^{'} \in \mathbb{R}^{N_m \times d}$。接下来，对于转换后特征 $Z_m^{'}$中的一行，我们在 Z_b中查找其 k个最近邻。在这种情况下，我们将这些邻居的书名作为转换后特征所对应的信息。最后，对于 $Z_m^{'}$中的一行，我们可以检查图书空间中邻居的书名是否与电影空间中的原始电影名称具有相关语义。结果汇总于表4中。

从该表中可以看出，图书域中的5个最近邻项目与电影域中对应项目的语义相近。例如，对于用户“21778”，其互动过的电影为“8329”，电影名称为“亚当斯一家”，讲述了一个关于情感与爱情的故事。同时，对于对应的翻译后特征，图书域中的5个最近邻项目也是关于情感与爱情的小说。这表明，翻译后的特征能够捕捉从电影域到图书域的相关语义。这一结果符合跨域点击率预测中的一种普遍观点：如果用户喜欢某一领域中的某个项目，那么用户也可能喜欢其他领域中语义相关的项目 [6], [9], [22]。值得注意的是，我们的模型并未使用跨领域项目的任何对齐信息，却能自动捕捉这种关联，并提供更多有用信息来提升点击率

7.请注意，我们的方法并不要求用户在两个领域都拥有训练样本。我们在此提出这一要求，只是为了更好地分析翻译后的特征。为了便于分析，我们让两个领域中的样本共享正标签，因为点击率中的“负”样本通常是采样得到的，其置信度低于正样本。



(a) 亚马逊电影 (b) 亚马逊图书 (c) 淘宝广告 (d) 淘宝推荐

图5：不同稀疏度对亚马逊和淘宝的影响。训练比例指原始训练数据的比例。

表4：结果表明，图书空间中的翻译后特征具有与其在电影空间中的原始项目语义相关的项目语义。标题为电影名称或图书名称。加粗的图书标题表示其与电影语义相近。

	电影领域中的物品		来自图书域的 5个最近邻项		
用户ID	物品ID	标题	ItemId	标题	物品类型
21778	8329	亚当斯一家	39040	**心碎之处**	情感、爱情
			33024	计划	
			29368	**舔（舞台跳水小说）**	
			37926	**雪覆**	
			44092	**群星依旧闪耀之处**	
4842	13887	《神秘科学剧院3000》	10505	**《寻羊冒险记：一部小说》**	悬疑、惊悚小说
			7713	**《韦尔德姐妹》**	
			8205	**《北极星》**	
			135765	**《哈利·波特与密室》**	
			6374	**《安布勒警告》**	
2078	4977	《星际之门：SG-1》第二季	43606	《无耻伯爵夫人》	奇幻、爱情
			16339	**《赤色之城》**	
			35940	**非同寻常的狼女孩**	
			10612	**反抗者**	
			12671	**匕首之星（帕林史诗，第1卷）**	

表5：CDAnet不同模型部分的消融实验。图中数值为曲线下面积结果。“w/o”表示不包含相应的模型部分。“w/o translation network”表示我们直接训练随机初始化的增强网络。“w/o augmentation network”表示我们仅使用训练好的转换网络进行评估。

数据集	亚马逊		淘宝	
领域	电影	book	ad	rec
不使用 MMOE	0.7186	0.7648	0.6199	0.7012
不使用 $\mathcal{L}^{orth}$	0.7153	0.7734	0.6190	0.7022
不使用 $\mathcal{L}^{cross}$	0.7188	0.7757	0.6191	0.7003
不含转换网络	0.6464	0.7649	0.6075	0.7001
不含增强网络	0.6818	0.7686	0.6168	0.6811
CDAnet	**0.7225**	**0.7811**	**0.6200**	**0.7034**

预测性能。该结果进一步证明了我们的转换网络的价值。

5.4 消融研究

5.4.1 不同模型部分的影响

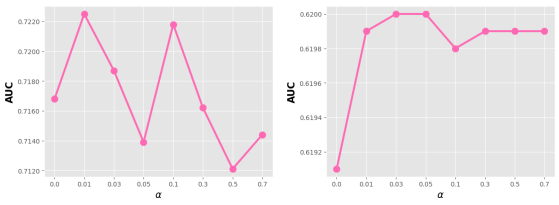
为了评估模型不同组件所带来的影响，我们进行了实验，分析移除相应模块对性能的影响。结果如表 5 所示。

比较 CDAnet 与不使用 MMOE 的结果可以看出，基于MMOE的特征提取器对提升模型性能具有积极作用。该MMOE 模块可能有助于转换网络缓解优化冲突问题，并更好地迁移知识。通过比较不使用 $\mathcal{L}^{orth}$ 和不使用 $\mathcal{L}^{cross}$ 的结果可以看出，无论是移除正交约束损失 $\mathcal{L}^{orth}$，还是移除交叉监督损失 $\mathcal{L}^{cross}$，都会导致性能下降。正交约束损失 $\mathcal{L}^{orth}$ 有助于 CDAnet 在转换后保持潜在特征之间的相似性，并避免模式坍塌问题。交叉监督损失 $\mathcal{L}^{cross}$ 在学习两个转换器 W_s^{trans} 和 W_t^{trans} 的过程中发挥着至关重要的作用。没有 $\mathcal{L}^{cross}$，转换器就无法学习如何将潜在特征转换到另一个空间。

此外，移除转换网络或增强网络都会导致性能较差。转换网络负责学习如何在不同领域之间迁移知识。移除该网络后，后续增强网络便无法获取有用的知识。增强网络复用转换网络预训练得到的参数，并利用额外的转换后的潜在特征来实现最终的点击率预测目标。移除增强网络后，我们只剩下目标为转换的转换网络，因此无法保证良好的点击率预测性能。

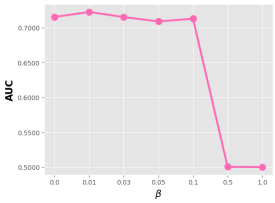
5.4.2 参数敏感性

在 CDAnet 中，α 和 β 分别控制跨域监督损失和正交约束损失的权重。K 控制共享的基于 MMOE 的特征提取器中的专家数量。在此，我们研究参数 sen-

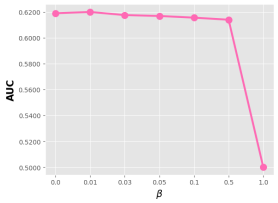


(a) α 在亚马逊电影上

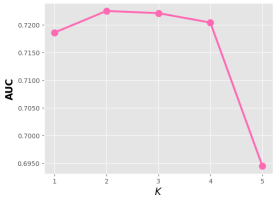
(b) α 在淘宝广告上



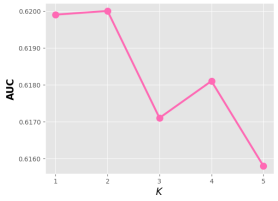
(c) β 在亚马逊电影上



(d) β 在淘宝广告上



(e) K 在亚马逊电影上



(f) K 在淘宝广告上

图6：超参数对不同数据集的影响。 α 和 β 是不同损失部分的损失权重。K 表示专家数量。

这些超参数的敏感性，以研究它们对模型性能的影响。结果如图6所示。

在此图中， α 控制交叉监督对翻译器学习的影响强度。 α 过大可能会主导 $\mathcal{L}^{vani}$ 的学习，并对交叉监督翻译器的学习产生不良影响。 β 控制 $\mathcal{L}^{orth}$ 的权重， β 过大可能会限制翻译器的学习灵活性。 K 是共享的基于MMOE的特征提取器中的专家数量，而 $K = 2$ 表现最佳。一个可能的原因是，过大的 K 会涉及过多参数，而数据不足以支撑模型的复杂度。

6 结论与未来工作

在具有异构输入特征的真实世界系统中，跨域点击率（CDCTR）预测是一项重要且实用的问题。本文提出了一种名为 CDAnet 的新模型，该模型包含转换网络和增强网络，可有效迁移具有异构输入特征的 CDCTR 知识。通过大量实验，我们表明 CDAnet 能够学习有意义的转换后潜在特征，并提升点击率预测性能。在淘宝应用的大规模生产数据集和在线系统上的结果表明，该模型在真实世界应用中具有优越性。

尽管 CDAnet 的性能优于现有模型，但在知识迁移方面仍存在不足。两个领域的潜在特征通常无法完全重叠，这意味着潜在特征中的某些成分可能无法翻译，并可能产生负面影响。未来，我们将研究更有效的潜在特征转换技术，以提升目标域的点击率预测效果。

参考文献

[1] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Arad-hye, G. Anderson, G. Corrado, W. Chai, M. Ispir 等，“用于推荐系统的宽深度学习”，载于第*1*届推荐系统深度学习研讨会论文集，2016年，第7–10页。[2] G. Zhou, X. Zhu, C. Song, Y. Fan, H. Zhu, X. Ma, Y. Yan, J. Jin, H. Li和K. Gai，“用于点击率预测的深度兴趣网络”，载于第*24*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，2018年，第1059–1068页。[3] W. Ouyang, X. Zhang, L. Li, H. Zou, X. Xing, Z. Liu和Y. Du，“用于点击率预测的深度时空神经网络”，载于第*25*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，2019年，第2078–2086页。[4] C. Li, Y. Lu, Q. Mei, D. Wang和S. Pandey，“Twitter时间线中的广告点击预测”，载于第*21*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，2015年，第1959–1968页。[5] W. Ouyang, X. Zhang, L. Zhao, J. Luo, Y. Zhang, H. Zou, Z. Liu和Y. Du，“MiNet：用于跨域点击率预测的混合兴趣网络”，载于第*29*届*ACM*国际信息与知识管理会议论文集，2020年，第2669–2676页。[6] L. Liu, Y. Wang, T. Wang, D. Guan, J. Wu, J. Chen, R. Xiao, W. Zhu和F. Fang，“淘宝中的跨域持续迁移学习与点击率预测”，*arXiv*预印本*arXiv:2208 05728*，2022。[7] Y. Zhang, Z. Chan, S. Xu, W. Bian, S. Han, H. Deng和B. Zheng，“KEEP：通过知识提取与插入实现在线推荐的工业级预训练框架”，载于第*31*届*ACM*国际信息与知识管理会议论文集，2022年，第3684–3693页。[8] X.-R. Sheng, L. Zhao, G. Zhou, X. Ding, B. Dai, Q. Luo, S. Yang, J. Lv, C. Zhang, H. Deng等，“一个模型服务所有领域：用于多域点击率预测的STAR拓扑自适应推荐模型”，载于第*30*届*ACM*国际信息与知识管理会议论文集，2021年，第4104–4113页。[9] P. Li和A. Tuzhilin，“DDTCDR：深度双重迁移跨域推荐”，国际网络搜索与数据挖掘会议，2020年。[10] O. Sener和V. Koltun，“多任务学习作为多目标优化”，神经信息处理系统进展，第31卷，2018年。[11] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai和L. Van Gool，“密集预测任务的多任务学习：综述”，*IEEE*模式分析与机器智能汇刊，2021年。[12] J. Ma, Z. Zhao, X. Yi, J. Chen, L. Hong和E. H. Chi，“使用多门专家混合模型对多任务学习中的任务关系进行建模”，载于第*24*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，2018年，第1930–1939页。[13] H. Tang, J. Liu, M. Zhao和X. Gong，“渐进式分层提取（PLE）：一种用于个性化推荐的新型多任务学习（MTL）模型”，载于第*14*届*ACM*推荐系统会议，2020年，第269–278页。[14] X. Lin, H. Chen, C. Pei, F. Sun, X. Xiao, H. Sun, Y. Zhang, W. Ou和P. Jiang，“电子商务推荐中的多目标优化帕累托高效算法”，载于第*13*届*ACM*推荐系统会议论文集，2019年，第20–28页。[15] M. Richardson, E. Dominowska和R. Ragno，“预测点击：估计新广告的点击率”，载于第*16*届国际万维网会议论文集，2007年，第521–530页。[16] H. Guo, R. Tang, Y. Ye, Z. Li和X. He，“DeepFM：基于因子分解机的点击率预测神经网络”，载于第*26*届国际人工智能联合会议论文集，IJCAI’17系列，AAAI出版社，2017年，第1725–1731页。[17] R. Wang, B. Fu, G. Fu和M. Wang，“用于广告点击预测的深度交叉网络”，载于*ADKDD’17*论文集，2017年，第1–7页。[18] S. Rendle，“因子分解机”，载于*2010*年*IEEE*国际数据挖掘会议。IEEE，2010年，第995–1000页。

[19] G. Zhou, N. Mou, Y. Fan, Q. Pi, W. Bian, C. Zhou, X. Zhu和K. Gai，“用于点击率预测的深度兴趣演化网络”，载于*AAAI*人工智能会议论文集，第33卷，第01期，2019年，第5941–5948页。[20] G. Hu, Y. Zhang和Q. Yang，“Conet：用于跨域推荐的协同跨网络”，载于国际信息与知识管理会议。ACM，2018年，第667–676页。[21] F. Yuan, L. Yao和B. Benatallah，“Darec：通过迁移评分模式实现跨域推荐的深度领域适应”，国际人工智能联合会议，2019年。[22] X. Chen, Y. Zhang, I. W. Tsang, Y. Pan和J. Su，“迈向跨域推荐中用户偏好的等价转换”，*ACM*信息系统汇刊（*TOIS*），2020年。[23] Z. Wang, Y. Tsvetkov, O. Firat和Y. Cao，“梯度疫苗：研究并改进大规模多语种模型中的多任务优化”，*arXiv*预印本*arXiv:2010.05874*，2020年。[24] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman和C. Finn，“多任务学习的梯度手术”，神经信息处理系统进展，第33卷，第5824–5836页，2020年。[25] L. Chen, F. Yuan, J. Yang, X. He, C. Li和M. Yang，“面向跨域推荐的用户特定自适应微调”，*IEEE*知识与数据工程汇刊，2021年。[26] U. Evci, V. Dumoulin, H. Larochelle和M. C. Mozer，“Head2toe：利用中间表示实现更好的迁移学习”，载于国际机器学习会议。PMLR，2022年，第6009–6033页。[27] Y.-L. Sung, J. Cho和M. Bansal，“Lst：阶梯式侧向微调，实现参数与内存高效迁移学习”，*arXiv*预印本*arXiv:2206.06522*，2022年。[28] Y. Pang, J. Lin, T. Qin和Z. Chen，“图像到图像转换：方法与应用”，*IEEE*多媒体汇刊，2021年。[29] D. W. Otter, J. R. Medina和J. K. Kalita，“深度学习用于自然语言处理的应用综述”，*IEEE*神经网络与学习系统汇刊，第32卷，第2期，第604–624页，2020年。[30] P. Isola, J.-Y. Zhu, T. Zhou和A. A. Efros，“使用条件对抗网络进行图像到图像转换”，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第1125–1134页。[31] M. Mirza和S. Osindero，“条件生成对抗网络”，*arXiv*预印本*arXiv:1411.1784*，2014年。[32] M.-Y. Liu, T. Breuel和J. Kautz，“无监督图像到图像转换网络”，神经信息处理系统进展，第30卷，2017年。[33] Z. Yi, H. Zhang, P. Tan和M. Gong，“Dualgan：用于图像到图像转换的无监督对偶学习”，载于*IEEE*国际计算机视觉会议论文集，2017年，第2849–2857页。[34] J.-Y. Zhu, T. Park, P. Isola和A. A. Efros，“使用循环一致对抗网络进行无配对图像到图像转换”，载于*IEEE*国际计算机视觉会议论文集，2017年，第2223–2232页。[35] T. Kim, M. Cha, H. Kim, J. K. Lee和J. Kim，“利用生成对抗网络学习发现跨域关系”，载于国际机器学习会议。PMLR，2017年，第1857–1865页。[36] Y. Taigman, A. Polyak和L. Wolf，“无监督跨域图像生成”，*arXiv*预印本*arXiv:1611.02200*，2016年。[37] K. Bousmalis, N. Silberman, D. Dohan, D. Erhan和D. Krishnan，“使用生成对抗网络进行无监督像素级域适应”，载于*IEEE*计算机视觉与模式识别会议论文集，2017年，第3722–3731页。[38] R. Wang, R. Shivanna, D. Cheng, S. Jain, D. Lin, L. Hong和E. Chi，“DCN v2：改进的深度交叉网络及网络规模学习排序系统的实践经验”，载于*2021*年万维网会议论文集，2021年，第1785–1797页。