

2024OPPO: 混合多场景网络 (HMDN)：基于层次化统一模型的CTR预估方法



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

26 人赞同了该文章

收起

Introduction

在线推荐⁺系统通过准确预测点击率（CTR）来增加平台收入至关重要。随着业务领域的扩展，它面临着多样化的群体、场景和目标，每种都需要定制化的推荐策略。然而，以多场景为例，为每种分布构建单独的模型成本高昂，因此需要一种能够同时容纳多种场景的模型。所有这些都可以概括为一个统一的多分布建模范式，旨在捕捉分布之间的共性和差异。

近年来，在多场景、多目标、多群体和多兴趣等领域取得了显著进展。然而，大多数方法只考虑了单一的多分布建模方式。实际上，可能存在多种多分布问题同时并存，并形成层次关系，我们称之为“混合多分布”。例如，在大型电商平台中，不同群体在不同场景下的兴趣和意图往往不同。因此，设计以捕捉混合多分布内部层次关系的模型是必要的。

一些最近的研究工作开始同时处理多场景和多目标问题，但通常使用复杂的堆叠结构，直接将底层多场景模块与顶层多任务模块堆叠在一起。尽管这些方法在实践中是有效的，但也存在两个局限性：1) 同时建模超过两个多分布问题时，简单堆叠会导致过度复杂化，从而降低灵活性；2) 它们未能考虑混合多分布内部的层次关系，导致性能不佳。

在本文中，我们提出了一种灵活的建模范式，名为“混合多分布网络”（HMDN），它有效地建模了混合多分布内部的层次关系，并能与现有的多分布建模方法**无缝集成⁺**。通过利用这些层次关系，我们可以将参数关联到不同的数据分布，构建一个更高效的统一结构。具体来说，我们首先引入了一个“层次多分布嵌入细化”（HMDRR）模块，该模块在多个级别上递归应用量化残差，以获得从粗到细的粒度层次关系的量化嵌入。然后，我们将提取的量化嵌入相加，以得出整体的层次表示。这种层次表示随后可以集成到现有的单多分布模型，如专家混合（MoE）模型和动态权重模型（DW）中，使它们能够同时处理多场景和多目标问题，并捕捉混合多分布内部的层次关系。

Methodology

Problem Formalization

本文主要探讨了在混合多场景数据下的推荐系统统一建模问题。给出 N 种不同的多分布 $\{B_1, B_2, \dots, B_N\}$ ，对于第 n 种多分布，比如多场景，其分布为 $\{B_{n1}, B_{n2}, \dots, B_{nM}\}$ ，其中 M 表示多场景问题中的场景数量。对于一组混合多场景数据 $\{B_{nm}\} \ m \in [1, M] \ n \in [1, N]$ ，它们共享一个共同的**特征空间⁺** \mathcal{X} 和一个共同的标签空间 \mathcal{Y} 。目标是构建一个统一的CTR**预测函数⁺**： $f: \mathcal{X} \rightarrow \mathcal{Y}$ 能够充分探索混合多分布内的层次关系，并准确预测点击率结果。令输入特征为 x ，从中提取混合多场景特征 x_b 来指导统一建模的构建。

Architecture Overview

▲ 赞同 26 ▼ ● 添加评论 ↗ 分享 ❤ 喜欢 ★ 收藏 📄 申请转载 ...

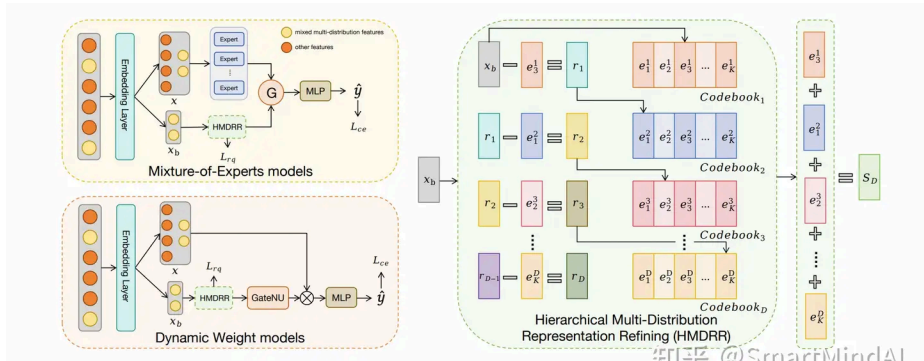


Figure 1: Illustration of the proposed HMDM framework. (\otimes) denotes the element-wise multiplication operation.

图展示了新型混合多分布建模框架HMDN的完整过程。具体来说，对于一批具有各种特性的输入，我们首先通过嵌入层将输入特性和混合多分布特性映射到变量 \mathbf{x} 和 \mathbf{x}_b 。然后，设计了层次多分布表示 (HMDRR) 模块，通过多层次残差量化提取层次表示。

Hierarchical Multi-Distribution Representation Refining

Embedding Layer

在嵌入层中，我们最初将与多分布相关的特征映射到低维稠密向量，并将它们拼接为 \mathbf{x}_b 。然后，我们将整个输入特征映射到低维稠密向量 \mathbf{x} 。这些表示 (\mathbf{x} 和 \mathbf{x}_b) 能够捕捉到整体和特定分布的属性，从而使得模型能够增强处理多样性和复杂分布的能力。

Residual Quantization Layer

在 \mathbf{x}_b 上利用多级残差量化，以获得精细层次的多分布表示。这一过程允许我们捕捉多分布数据中的细微变化和依赖关系，从而产生更精确、更详细的表示。在向量量化 (VQ) 中，codebook \mathcal{C} 包含代码 \mathbf{c} 及其嵌入 \mathbf{e}_c ，形成有限集合 $\{\mathbf{c}, \mathbf{e}_c\}_{c \in [K]}$ ，其中 K 是codebook的大小 $\mathbf{c}_e \in \mathbb{R}^z$ 是嵌入维度。给定一个向量 $\mathbf{s} \in \mathbb{R}^z$ ，向量 \mathbf{s} 的量化，表示为 $VQ(\mathbf{s}; \mathcal{C})$ ，找到codebook \mathcal{C} 中与 \mathbf{s} 最近的嵌入的代码，表示为：

$$VQ(\mathbf{s}; \mathcal{C}) = \underset{c \in [K]}{\operatorname{argmin}} \|\mathbf{s} - \mathbf{e}_c\|_2^2$$

因此，受到近期相关研究的启发，我们采用残差量化技术，RQ-VAE，来捕捉这些层次关系。设量化深度为 D 。给定混合多分布表示 \mathbf{x}_b ，其残差量化可以表示为：

$$\text{残差量化}(\mathbf{x}_b) = \mathbf{x}_b - \text{量化}(\mathbf{x}_b, D)$$

通过残差量化技术，我们能够更准确地处理不同层次的表示，避免了单纯合并不同层次的表示可能会忽视的层次关系信息的丢失，以及混合多分布中的细微关系：

$$RQ(\mathbf{x}_b; \mathcal{C}, D) = (\mathbf{c}_1, \dots, \mathbf{c}_D)$$

其中 \mathbf{c}_d 表示第 d 级的代码。将第0级的残差表示为 $\mathbf{r}_0 = \mathbf{x}_b$ ，RQ递归地计算表示第 $d-1$ 级残差的代码 \mathbf{c}_d ，并计算下一个残差 \mathbf{r}_d ：

$$\mathbf{c}_d = VQ(\mathbf{r}_{d-1}; \mathcal{C}), \quad \mathbf{r}_d = \mathbf{r}_{d-1} - \mathbf{e}(\mathbf{c}_d)$$

我们将每个级别的量化嵌入求和，以得出层次表示向量 \mathbf{s}_D 。 $\mathbf{s}_D = \sum_{i=1}^D \mathbf{e}(\mathbf{c}_i)$ 通过引入这种层次结构，我们使其更适合学习具有依赖性的混合多分布表示。不同于@Tiger和@UIST，我们的目的是推导出层次表示向量 \mathbf{s}_D ，而不是代码元组。

Dual-View Hierarchical Representation

我们根据输入的分区设计两种表示提取实现，隐式和显式：

隐式表示提取 它直接利用混合多分布特征 \mathbf{x}_b 来建立所有codebook。这有利于自主捕获隐含的复杂层次关系

显式表示提取 它包括按照分布类型将 \mathbf{x}_b 划分为特定分布的特征, 例如, 将与人口相关的特征和与场景相关的特征分别分离到不同的codebook 层级。

HMDRR模块的**损失函数**⁺计算方式如下:

$$L_{rq} = \sum_{d=1}^D \|e(c_d) - \text{sg}[\mathbf{x}_b^{(d)}]\|_2^2 + \beta \sum_{d=1}^D \|\text{sg}[e(c_d)] - \mathbf{x}_b^{(d)}\|_2^2$$

其中, $\mathbf{x}_b^{(d)}$ 代表第d层码本的多分布嵌入输入 $\text{sg}[\cdot]$ 代表停止梯度操作。随着 d 的增加 $\mathbf{x}_b^{(d)}$ 与 $e(c_d)$ 之间的量化误差逐渐减小, 使得RQ能够在从粗级别到细级别的过程中逐步逼近混合多分布特征, 并同时更新码本 C 。

Hierarchical Multi-Distribution Modeling

我们基于专家混合 (MoE) 模型和分布式权重 (DW) 模型实现基于层次多分布模型的构建。 **专家混合 (MoE)** MoE-based models 由多个专家网络和控制网络组成。控制网络根据输入的特性选择最适合的专家网络进行预测, 并通过权重结合每个专家的输出, 从而获得最终结果。整体表达式如下:

$$\hat{y} = \sum_{i=1}^n G(\mathbf{x})_i E_i(\mathbf{x})$$

$$G(\mathbf{x}) = \text{softmax}(\mathbf{x} \cdot \mathbf{W}_g)$$

其中 $G(\cdot)$ 表示门控网络, 而 $E_i(\cdot)$ 表示第 i 个专家网络。

在MoE中选择门控网络的适当输入至关重要。直观的选择是多分布表示 \mathbf{x}_b , 但这种选择过于粗粒度, 可能会导致层次分布信息的丢失。同时, 孤立的分布表示在面对新分布时可能表现出较差的一般化能力。因此, 我们采用之前推导出的层次表示 \mathbf{s}_D 作为门控网络的输入, 而不是 \mathbf{x}_b , 以促进精细粒度的混合多分布建模。等式可以重新表述为:

$$\hat{y} = \sum_{i=1}^n G(\mathbf{s}_D)_i E_i(\mathbf{x})$$

通过动态权重 (DW) 的模型, 采用**门控机制**⁺ (如GateNU) 来动态调整下层嵌入, 生成适应不同分布的个性化的嵌入:

$$\hat{y} = \delta_b \otimes \mathbf{x}$$

$$\delta_b = G_{NU}(\mathbf{x}_b)$$

$$L = L_{ce} + \alpha L_{rq} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) + \alpha L_{rq}$$

第一项是CTR任务的**交叉熵损失**⁺, 表示为 L_{rq} 。符号 α 用于调整该损失的重要性。样本数量用符号 N 表示, 其中 $y_i \in \{0, 1\}$ 是第 i 个样本的真实值 \hat{y}_i 是对应的预测概率。

EXPERIMENTS

Dataset Description

阿里-CCP数据集是从淘宝的工业推荐系统收集的**真实世界推荐系统流量日志**的公开数据集。它最初包含三个领域。我们遵循先前工作的**数据分割**⁺方式, 利用`user_gender`, `user_city`和`domain_id`来分割样本, 产生了总共超过四千二百万个分割。训练集和**测试集**⁺的大小分别超过训练集四千二百万和测试集四千三百万。

工业数据集是从OPPO在线广告平台收集的工业用户日志数据, 时间跨度从2024年1月2日到2024年1月9日。我们按照时间顺序分割训练集和测试集, 前七天用于训练, 最后一天用于测试。我们利用`domain_id`, `is_new_user`和`ad_source`作为分割依据, 产生了总共12个分割。`is_new_user`表示在特定领域内是否为新用户, `ad_source`则代表广告的来源, 是OPPO还是第三方。

Experimental Results

表列出了三

Model	Production		Ali-CCP	
	AUC	RelaImpr	AUC	RelaImpr
DNN	0.8206	+0.00%	0.6183	+0.00%
DCN	0.8196	-0.12%	0.6231	+0.78%
Wide&Deep	0.8328	+1.48%	0.6156	-0.43%
Shared Bottom	0.8401	+2.38%	0.6188	+0.08%
PEPNet	0.8415	+2.54%	0.6246	+1.01%
HMDN (PEPNet)	0.8421	+2.62%	0.6250	+1.08%
MMoE	0.8427	+2.69%	0.6237	+0.87%
HMDN (MMoE)	0.8434	+2.78%	0.6253	+1.13%

首先，单一多分布建模方法，如共享底部、MMoE 和 PEPNet，比传统方法表现出更优的性能，这表明对CTR预测，区分地建模各种分布是必要的。其次，对比MMoE 和 PEPNet 与共享底部，我们发现以更精细的粒度捕捉分布之间的关系能带来更多的提升。最后但最重要的是，HMDN 在公共数据集和工业数据集上均取得了最佳性能。

值得注意的是，HMDN 对 MMoE 和 PEPNet 的改进完全归因于其对混合多分布中的层次关系的全面捕捉。这只需要对现有的单一多分布建模方法进行少量的调整，证明了HMDN的有效性和灵活性。综上所述，HMDN 的成功不仅在于其对复杂多分布的高效处理，更在于它能够适应并优化现有的单一多分布建模方法，展现出其在CTR预测领域的卓越性能和广泛适用性。

原文《HMDN: Hierarchical Multi-Distribution Network for Click-Through Rate Prediction》

发布于 2024-09-04 11:10 · IP 属地北京

[工业级推荐系统](#) [ctr预估](#) [MOE](#)



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