arXiv:2303.07607v2 [cs.IR] 8 Jun 2023

CoMeta：在推荐冷启动问题中利用协同信息增强元嵌入

胡浩南、荣大中、陈建海(✉)、何勤明和刘振光

浙江大学，中国杭州 {huhaonan, rdz98, chenjh919, hqm}@zju.edu.cn, {liuzhenguang2008}@gmail.com

摘要。冷启动问题对现有推荐模型而言颇具挑战。具体来说，对于只有少量交互的新物品，其 ID 嵌入训练不足，导致推荐性能不佳。近期一些研究引入元学习来解决冷启动问题，通过为新物品生成元嵌入，作为其初始 ID 嵌入。然而，我们认为这些方法的能力有限，因为它们主要利用仅包含少量信息的物品属性特征，却忽略了用户和旧物品的 ID 嵌入中蕴含的有用协同信息。为解决这一问题，我们提出 CoMeta，利用协同信息增强元嵌入。CoMeta 由两个子模块组成：B-EG 和 S-EG。具体来说，对于一个新物品：B-EG 计算旧物品 ID 嵌入的基于相似度的加权和，将其作为该新物品的基础嵌入；S-EG 不仅利用该物品的属性特征，还利用与其发生交互的用户的平均 ID 嵌入，生成其偏移嵌入。最终的元嵌入由基础嵌入与偏移嵌入相加得到。我们在两个公开数据集上进行了大量实验。实验结果表明，CoMeta 既有效又具有兼容性。

关键词：推荐系统 · 物品冷启动问题 · 元学习 · 物品ID嵌入 · 深度学习

1 引言

互联网的蓬勃发展催生了大量信息的日常产生。如今，信息过载已成为影响人们日常生活的严重问题。推荐系统通过为用户筛选有用信息，能够有效缓解信息过载问题。因此，推荐系统已被应用于各种网络服务中（例如，短视频 [3], 电子商务 [24] 和新闻门户 [23]）。协同过滤 [14],基于历史交互学习用户的潜在兴趣，是应用最广泛的传统推荐方法之一。除了协同过滤之外，许多近期研究（例如，Wide&Deep [2] 和 DeepFM [6]）还利用用户特征和物品特征，并采用深度神经

网络（DNN）来提升推荐性能。这些深度推荐模型通常由两部分组成：一个嵌入层和一个多层感知机（MLP）。嵌入层将原始特征转换为稠密向量，而多层感知机则捕捉更复杂的语义。

尽管这些模型取得了显著的成功，但它们仍然受到冷启动问题 [17]的困扰。在这些模型中，需要足够的交互数据，才能为每个物品训练出合理的 ID 嵌入。然而，在实际场景中，许多新物品的交互数量有限，因此其 ID 嵌入训练不足，导致推荐准确率下降。冷启动问题是推荐系统面临的主要挑战之一。

最近，一些方法被提出用于解决冷启动问题。DropoutNet [19] 和 MTPR [4] 通过对 ID 嵌入应用 dropout 来提升推荐模型的鲁棒性。MetaEmb [11]训练一个嵌入生成器，为每个新物品生成良好的初始 ID 嵌入。GME [10] 旨在构建物品图，并利用其属性特征以及其他相关旧物品的属性特征，为每个新物品学习理想的初始 ID 嵌入。MWUF [28] 提出了两个元网络，将冷物品 ID 嵌入转换到热特征空间。

用户和旧物品的 ID 嵌入经过了充分交互的充分训练，因此包含了极有用的协同信息。然而，上述所有方法都没有考虑协同信息。我们认为，这些方法的能力仍然有限。本文提出 CoMeta，利用协同信息增强新物品的元嵌入。CoMeta 包含两个组件：基础嵌入生成器（B-EG）和偏移嵌入生成器（S-EG）。在常见的研究设置中，冷启动问题分为两个阶段：冷启动阶段（新物品在训练数据集中没有任何交互）和预热阶段（新物品在训练数据集中有少量交互）。B-EG 中基础嵌入的计算和 S-EG 中交互用户 ID 嵌入的平均值的计算，都依赖于新物品的交互。因此，我们提出的 CoMeta 主要旨在为预热阶段的新物品学习理想的初始 ID 嵌入。此外，由于 CoMeta 是一个为新物品生成初始 ID 嵌入的额外模块，且不会影响旧物品，因此可以应用于各种推荐模型。

本文的主要贡献总结如下：

– 我们提出了一种基于元学习的方法 CoMeta，通过为新物品学习良好的初始 ID 嵌入来解决物品冷启动问题。CoMeta 可轻松应用于各种推荐模型。– 在 CoMeta 中，我们提出了两个子模块，分别称为 B-EG 和 S-EG，用于利用协同信息增强生成的元嵌入。具体而言，B-EG 通过旧物品的 ID 嵌入利用协同信息；S-EG 则通过交互用户的 ID 嵌入利用协同信息。

– 我们在两个大规模公开数据集上进行了大量实验。实验结果表明，CoMeta 优于现有的冷启动方法，并能显著提升广告冷启动问题的推荐性能。

2 相关工作

推荐系统旨在向用户推荐个性化物品。在早期研究中，许多研究者尝试使用传统机器学习方法进行推荐。FM [13], 能够对低阶特征交互进行建模，是应用最广泛的传统机器学习推荐方法之一。由于其简单有效，研究者提出了许多基于 FM 的方法。然而，上述方法无法捕捉高阶交互。近年来，研究者提出了各种深度学习模型来解决这一问题。Wide&Deep [2], 、DeepFM [6],、PNN [12]和 DCN [21] 会自动交叉高阶特征，以获取更多信息。DIN [27] 和 DIEN [26] 利用注意力机制，根据用户的历史交互捕捉用户兴趣。值得一提的是，也有一些研究关注推荐安全 [15,16]。

尽管这些模型取得了显著成功，但它们仍然受到冷启动问题的困扰。在推荐系统中，面对交互数据有限的新用户或新物品，要做出合理的推荐颇具挑战。为了解决冷启动问题，一些方法致力于学习更稳健的模型。DropoutNet [19] 和 MTPR [4]会对 ID 嵌入进行 dropout，以降低模型对它们的依赖。CC-CC [18] 则用随机向量替代 ID 嵌入。解决这一问题的另一种方法是使用变分自编码器 [7]。CVAR [25] 利用变分自编码器在潜在空间中对物品 ID 与物品特征之间的关系进行建模。

此外，近期有许多研究引入元学习来应对冷启动问题。元学习也称为学会学习，它利用多个相关任务中的先验知识，使模型仅凭少量样本就能快速适应新任务。MeLU [8]基于模型无关元学习（MAML） [5]框架，为所有用户生成模型的初始参数。PAML [20] 将社交网络引入 MeLU。MetaEmb [11]利用物品的属性特征，为新物品生成良好的初始 ID 嵌入。GME [10] 同时考虑新物品的属性特征和其他相关旧物品的属性特征，学习如何为新物品生成理想的初始 ID 嵌入。MWUF [28] 旨在训练元缩放网络和元偏移网络，以预热冷启动 ID 嵌入。

3 方法

在本节中，我们提出 CoMeta，利用协同信息为新物品 ID 生成理想的元嵌入。

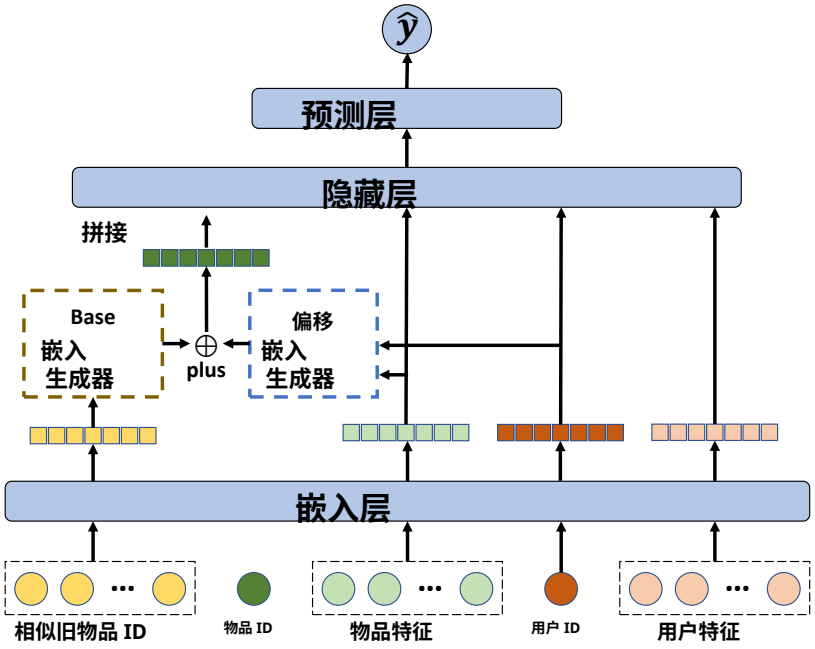

图 1. CoMeta 的框架。

3.1 概述

本文中，只有新物品 ID 嵌入由所提出的 CoMeta 生成，旧物品 ID 嵌入则可以直接使用。如图 1 所示，我们提出的 CoMeta 包含两个组件:

1）基础嵌入生成器（B-EG）。在 B-EG 中，对于每个新物品，我们首先计算它与旧物品之间的相似度分数。然后，根据相似度分数计算旧物品 ID 嵌入的加权和，并将结果作为生成的基础嵌入。2）偏移嵌入生成器（S-EG）。在 S-EG 中，对于每个新物品，我们首先计算与其发生交互的用户的平均 ID 嵌入。然后，我们将该物品的属性特征和交互用户 ID 嵌入的平均值作为输入，使用深度神经网络生成其偏移嵌入。

最终生成的新物品 ID嵌入可表示为:

$$\mathbf{v}_{new} = \mathbf{v}_{BEG} + \mathbf{v}_{SEG}, \tag{1}$$

其中 $\mathbf{v}_{BEG}$ 是基础嵌入，而 $\mathbf{v}_{SEG}$ 是偏移嵌入。

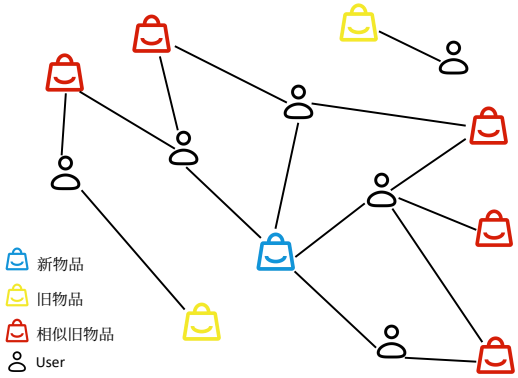

图 2. 为新物品查找相似旧物品的示意图。

3.2 B-EG：基础嵌入生成器

由于新物品的 ID 嵌入仅通过少量交互进行学习，因此缺乏足够的协同信息，难以很好地适配推荐模型。然而，通过充分交互训练得到的旧物品 ID 嵌入包含有用的协同信息。因此，我们提出 B-EG，以捕获旧物品 ID 嵌入中的协同信息，从而改进新物品的 ID 嵌入。

受 ItemCF [9], 启发，我们认为，如果两个物品与同一用户发生过交互，它们就可能相似。此外，[1] 表明，用户的交互越多，其对物品相似度的贡献就越小。因此，我们将物品 i 与物品 j 之间的相似度定义如下：

$$Sim\left(i,j\right)=\frac{\sum_{a\in\left(U\left(i\right)\bigcap U\left(j\right)\right)}\frac{1}{\log\left(1+\left|I\left(a\right)\right|\right)}}{\sqrt{\left|U\left(i\right)\right|\left|U\left(j\right)\right|}}, \tag{2}$$

其中，$U\left(i\right)$ 是与物品 i发生过交互的用户集合；$\left|U\left(i\right)\right|$ 是 $U\left(i\right)$ 中的用户数量；$\left|I\left(a\right)\right|$ 是与用户 a发生过交互的物品数量。我们将相似度最高的 K 个旧物品保留为物品 i的集合 $T\left(i\right)$，相似旧物品的权重可表示为：

$$\alpha_{ij}=\frac{Sim\left(i,j\right)}{\sum_{k\in T\left(i\right)}Sim\left(i,k\right)}. \tag{3}$$

图 2 展示了如何在 B-EG 中为新物品寻找相似的旧物品。

最后，B-EG 根据相似旧物品的 ID 嵌入为新物品生成基础嵌入。具体来说，对于新物品 i，我们通过对旧物品的 ID 嵌入进行加权求和，得到其基础嵌入 $\mathbf{v}^{BEG[i]}$：

$$\mathbf{v}_{BEG[i]}=\sum_{k\in T\left(i\right)}\alpha_{ik}\mathbf{v}_{k}. \tag{4}$$

3.3 S-EG：偏移嵌入生成器

现有模型（如 MetaEmb [11] ）使用物品属性特征为新物品 ID 生成元嵌入。然而，这些模型没有关注协同信息。与旧物品的 ID 嵌入一样，用户的 ID 嵌入也包含有用的协同信息。因此，在 S-EG 中，我们利用每个新物品的属性特征及其平均交互用户 ID 嵌入，为其生成偏移嵌入。具体来说，对于新物品 i，我们首先计算其精炼后的交互用户表示：

$$\mathbf{h}_i^u = \mathbf{W}_u Mean\left(\{\mathbf{u}_j, j \in U\left(i\right)\}\right), \tag{5}$$

其中，$\mathbf{W}_u$ 是一个可学习参数； $Mean\left(\cdot\right)$表示均值池化。

然后，我们计算其精炼后的属性特征表示：

$$\mathbf{h}_i^f = \mathbf{W}_f \mathbf{Z}_i, \tag{6}$$

其中，$\mathbf{W}f$ 是一个可学习参数。

最后，给定其精炼后的交互用户表示和属性特征表示，我们得到其偏移嵌入：

$$\mathbf{v}_{SEG[i]} = g_\omega\left(\mathbf{h}_i^u, \mathbf{h}_i^f\right), \tag{7}$$

其中 $g_\omega(\cdot)$ 是一个参数为 ω 的多层感知机。

S-EG 的训练过程。遵循冷启动研究 [11], ，我们将为每个物品学习一个偏移嵌入视为一项任务。为了在交互数据有限的情况下实现快速适应，我们采用了一种基于梯度的元学习方法，该方法是对 MAML 的泛化 [5]。训练 S-EG 时，我们只更新参数 $\{\mathbf{W}_u$、$\mathbf{W}f$，$\omega\}$ 并冻结其他参数。我们使用旧物品的交互数据训练 S-EG。给定一个物品 i，我们随机选取两个互不重叠的标注数据小批次 $\mathcal{D}_i^a$ 和 $\mathcal{D}_i^b$，每个小批次包含 M 个样本。

我们首先对第一个小批次 $\mathcal{D}_i^a$ 进行预测：

$$\hat{y}_{aj} = f_\theta\left(\mathbf{u}_{aj}, \mathbf{Z}_{aj}, \mathbf{v}_{BEG[i]} + \mathbf{v}_{SEG[i]}, \mathbf{Z}_i\right), \tag{8}$$

其中，下标 aj 表示来自 $\mathcal{D}_i^a$ 的第 j 个样本； $f_\theta\left(\cdot\right)$ 是推荐模型。然后，按如下方式计算这些样本的平均损失：

$$loss_a = -\frac{1}{M}\sum_{j=1}^{M} y_{aj}\log\hat{y}_{aj} + (1-y_{aj})\log(1-\hat{y}_{aj}). \tag{9}$$

我们通过计算 $loss_a$ 相对于初始偏移嵌入的梯度，并进行一步梯度下降，得到新的嵌入：

$$\mathbf{v}_{SEG[i]}^{'} = \mathbf{v}_{SEG[i]} - \eta\frac{\partial loss_a}{\partial \mathbf{v}_{SEG[i]}}, \tag{10}$$

其中， η 是梯度下降的步长。

接下来，我们使用新的嵌入，并在小批量 $\mathcal{D}_i^b$ 上进行预测：

$$\hat{y}_{bj} = f_\theta \left(\mathbf{u}_{bj}, \mathbf{Z}_{bj}, \mathbf{v}_{BEG[i]} + \mathbf{v}^{'}_{SEG[i]}, \mathbf{Z}_i \right). \tag{11}$$

然后，我们计算平均损失：

$$loss_b = -\frac{1}{M} \sum_{j=1}^{M} y_{bj} \log \hat{y}_{bj} + (1 - y_{bj}) \log(1 - \hat{y}_{bj}). \tag{12}$$

S-EG 的最终损失函数为：

$$loss_{SEG} = \beta loss_a + (1 - \beta)\, loss_b, \tag{13}$$

其中， $\beta \in [0, 1]$ 是一个用于平衡两种损失的超参数。该损失函数考虑了两个方面：1）对新物品的推荐误差应尽可能小。2）生成的嵌入应能利用有限的交互数据快速适应。

4 实验

4.1 数据集

我们在两个公开数据集 MovieLens-1M 和淘宝广告上评估 CoMeta。

MovieLens-1M 数据集[1]。 它是最广为人知的基准数据集之一，包含数千部电影和用户的 100 万条电影评分记录。每部电影都可视为一个物品，并具有 4 个特征：ID、标题、类型和发行年份。每位用户的相关特征包括 ID、年龄、性别和职业。我们将评分不低于 4 的值转换为 1，其余转换为 0。

淘宝广告数据集[2]。 该数据集采集自淘宝[3]的广告展示日志，包含 8 天内来自 114 万用户的 2600 万条点击记录。每条广告都可视为一个物品，其特征包括广告 ID、类别 ID、广告活动 ID、品牌ID和价格。每位用户有 9 个特征：用户 ID、微群组ID、CMS群组ID、性别、年龄、消费等级、购物深度、职业和城市等级。

4.2 骨干模型和基线方法

骨干模型。 所提出的 CoMeta 与模型无关，可应用于各种推荐模型。因此，我们在以下骨干模型上进行实验：

[1] http://www.grouplens.org/datasets/movielens/[2] https://tianchi.aliyun.com/dataset/dataDetail?dataId=56[3] https://www.taobao.com/

– **Wide&Deep** [2]。Wide&Deep 由逻辑回归（LR）和深度神经网络组成，可捕获低阶和高阶交互。– **DeepFM** [6]。DeepFM 将因子分解机（FM）与深度神经网络结合起来。FM 对一阶和二阶特征交互进行建模，而深度神经网络则对高阶特征交互进行建模。– **PNN** [12]。基于乘积的神经网络在深度神经网络中引入了乘积层。实验中，我们使用 PNN 的两种变体：IPNN（带有内积层的 PNN）和 OPNN（带有外积层的 PNN）。– **AFM** [22]。注意力因子分解机将注意力机制引入 FM，从而区分不同特征交互的重要性。

基线方法。 我们选择了一些用于冷启动问题的最先进（SOTA）元学习方法作为基线方法：

– **MetaEmb** [11]。Meta-Embedding 采用元学习方法，利用新物品的属性特征训练嵌入生成器，为新物品 ID 生成优质的初始嵌入。– **GME** [10]。图元嵌入基于元学习和图神经网络，为新物品 ID 生成理想的初始嵌入。我们使用在所有 GME 模型中性能最佳的 GME-A 作为基线方法。– **MWUF** [28]。元预热框架训练元缩放网络和元偏移网络，将冷物品 ID 嵌入转换到热特征空间。缩放函数将冷 ID 嵌入转换到热特征空间，而平移函数则能够生成更稳定的嵌入。

表 1. 数据集统计信息。

数据集	旧物品			新物品		
	# 物品 ID	# 预训练样本数 推荐模型	# 训练样本数 冷启动模型	# 物品 ID	# 个样本用于 预热训练	# 个样本用于 测试
MovieLens-1M 1	,058	765,669	84,640	1,127	67,620	123,787
淘宝广告	892	3,592,047	1,784,000	540	810,000	109,712

4.3 实验设置

数据集划分。 我们按照 [11] 和 [28]，根据物品的大小对其进行分组：

– 旧物品：我们将样本数量大于阈值 N_{old} 的物品视为旧物品。对于 MovieLens-1M 和淘宝广告数据集，我们分别将 N_{old} 设为 200 和 2000。

– **新物品**：我们将样本数量少于 N_{old}但大于 N_{new} 的物品视为新物品。我们按时间戳对每个新物品的样本进行排序，并将其分为 4 组。随后，我们将前 $3 \times K_{fold}$ 个样本用作 warm-a、warm-b 和 warm-c 集，其余样本用作测试集。对于这两个数据集， N_{new} 分别设为 $\{80, 1500\}$ 和 K_{fold} 设为 $\{20, 500\}$ 。

表 1 展示了数据集的详细信息。

实现细节。 我们将每个特征的嵌入维度设为 16，并将平衡超参数 β 设为 0.1。推荐模型中的多层感知机采用相同的结构，包含三个隐藏层（64-64-64）。此外，我们将学习率设为 0.001，并使用 Adam 算法优化模型参数。按照 [11], 以下步骤进行实验：

1) 使用旧物品交互数据预训练基础推荐模型。2) 训练额外模块（例如，嵌入生成器、元网络）。3) 使用额外模块初始化新物品 ID 嵌入，并在测试集上计算评估指标。在此步骤中，新物品没有任何交互。4) 为新物品 ID 生成更好的元嵌入。在此步骤中，新物品有少量交互。**注意，只有 CoMeta 需要此步骤**。5) 使用 warm-a 集更新新物品 ID 嵌入，并在测试集上计算评估指标。6) 使用 warm-b 集更新新物品 ID 嵌入，并在测试集上计算评估指标。7) 使用 warm-c 集更新新物品 ID 嵌入，并在测试集上计算评估指标。

我们将步骤 3 称为冷阶段，将步骤 5–7 称为 warm-a、warm-b 和 Warm-c阶段。在冷阶段，由于新物品没有任何交互，我们无法计算 B-EG 中的基础嵌入，也无法计算 S-EG 中交互用户 ID 嵌入的平均值。为应对这一挑战，我们分别将全局物品和全局用户的平均 ID 嵌入作为基础嵌入和交互用户 ID 嵌入的平均值。在 warm 阶段，不同于其他方法将冷阶段生成的嵌入用作新物品的初始 ID 嵌入，CoMeta 会根据 warm 阶段的交互数据重新生成元嵌入。遵循 [11], , 每个训练步骤中，我们仅训练 1 个 epoch 来更新模型参数。

评估指标。 我们采用两种广泛使用的评估指标来评估性能：1）**曲线下面积（AUC）**。ROC 曲线下的面积。数值越大越好。2）**对数损失**。 二元交叉熵损失。数值越小越好。

表 2. 两个公开数据集上的实验结果。最佳结果以粗体显示。

数据集	骨干网络	方法	冷阶段 曲线下面积	冷阶段 对数损失	Warm-a阶段 曲线下面积	Warm-a阶段 对数损失	Warm-b阶段 曲线下面积	Warm-b阶段 对数损失	Warm-c阶段 曲线下面积	Warm-c阶段 对数损失
MovieLens-1M	DeepFM	DeepFM	0.7043	0.6660	0.7311	0.6370	0.7520	0.6126	0.7679	0.5920
		MetaEmb	0.7120	0.6508	0.7380	0.6235	0.7577	0.6011	0.7724	0.5826
		GME	**0.7168**	0.6443	0.7405	0.6188	0.7588	0.5978	0.7727	0.5804
		MWUF	0.7099	0.6598	0.7471	0.6220	0.7686	0.5918	0.7821	0.5702
		CoMeta	0.7106	**0.6381**	**0.7789**	**0.5706**	**0.7861**	**0.5607**	**0.7917**	**0.5526**
	Wide&Deep	Wide&Deep	0.7026	0.6686	0.7308	0.6386	0.7522	0.6139	0.7678	0.5936
		MetaEmb	0.7108	0.6524	0.7376	0.6247	0.7574	0.6025	0.7718	0.5845
		GME	**0.7158**	0.6435	0.7397	0.6182	0.7580	0.5978	0.7716	0.5812
		MWUF	0.7067	0.6636	0.7441	0.6254	0.7664	0.5947	0.7797	0.5735
		CoMeta	0.7081	**0.6375**	**0.7789**	**0.5695**	**0.7854**	**0.5609**	**0.7903**	**0.5538**
淘宝广告	DeepFM	DeepFM	0.5740	0.2270	0.5910	0.2244	0.6025	0.2229	0.6124	0.2218
		MetaEmb	0.5770	0.2263	0.5934	0.2239	0.6042	0.2226	0.6138	0.2216
		GME	**0.5787**	0.2260	0.5935	0.2238	0.6034	0.2226	0.6125	0.2216
		MWUF	0.5779	0.2259	0.5932	0.2247	0.6020	0.2234	0.6107	0.2223
		CoMeta	0.5781	**0.2258**	**0.5982**	**0.2234**	**0.6078**	**0.2223**	**0.6165**	**0.2213**
	Wide&Deep	Wide&Deep	0.5952	0.2257	0.6098	0.2234	0.6195	0.2220	0.6281	0.2208
		MetaEmb	0.5970	0.2246	0.6111	0.2227	0.6204	0.2215	0.6287	0.2205
		GME	**0.6001**	0.2240	0.6134	0.2223	0.6220	0.2212	0.6297	0.2203
		MWUF	0.5974	0.2236	0.6102	0.2223	0.6188	0.2213	0.6271	0.2204
		CoMeta	0.5974	**0.2235**	**0.6153**	**0.2219**	**0.6231**	**0.2210**	**0.6304**	**0.2201**

4.4 实验结果

总体性能。 我们在两个公开数据集上进行实验，并选择 Wide&Deep 和 DeepFM 作为骨干网络。同时，我们将 CoMeta 与三种最先进的冷启动方法进行比较：MetaEmb、GME、MWUF。对于每种方法，我们运行五次，并在表 2 中报告平均结果。

冷启动阶段。MetaEmb 和 GME 的表现优于骨干模型，这表明属性特征能够为物品 ID 提供有用信息。此外，GME 的表现远优于 MetaEmb，因为邻居物品包含更多信息。MWUF 通过对全局物品的嵌入取平均值来初始化新物品 ID 嵌入，这也很有帮助。尽管我们关注的是预热阶段，所提出的 CoMeta 也能在冷启动阶段改善新物品 ID 嵌入。

预热阶段。 由于 MetaEmb 和 GME 在冷启动阶段都能为新物品 ID 学习到理想的初始嵌入，因此它们在预热阶段的表现仍优于骨干模型。MWUF 在 MovieLens-1M 上表现良好，但在淘宝广告上的预测效果较差。我们推测，原因在于淘宝广告中新物品的属性特征训练不足，导致 MWUF 无法将冷物品 ID 嵌入正确转换到热特征空间。可以观察到，在这两个数据集上，我们的 CoMeta 在所有方法中表现最佳。这是因为我们生成了更优的元嵌入，其中包含大量有用的协同信息。

消融实验。为了探究 CoMeta 中不同组件的影响，我们使用 DeepFM 在两个数据集上进行了消融实验：1) DeepFM：骨干网络。2) CoMeta（不含 B-EG）：不含 B-EG 的 CoMeta。3) CoMeta

表3。消融实验的结果。

数据集	方法	冷阶段 曲线下面积 对数损失	Warm-a阶段 曲线下面积 对数损失	Warm-b阶段 曲线下面积 对数损失	Warm-c阶段 曲线下面积 对数损失
MovieLens-1M	DeepFM	0.7043 0.6660	0.7311 0.6370	0.7520 0.6126	0.7679 0.5920
	CoMeta（不含 B-EG）	**0.7118** 0.6475	0.7378 0.6227	0.7576 0.6007	0.7723 0.5824
	CoMeta（不含 S-EG）	0.7045 0.6636	0.7743 0.5884	0.7823 0.5747	0.7887 0.5631
	CoMeta	0.7106 **0.6381**	**0.7789 0.5706**	**0.7861 0.5607**	**0.7917 0.5526**
淘宝广告	DeepFM	0.5740 0.2270	0.5910 0.2244	0.6025 0.2229	0.6124 0.2218
	CoMeta（不含 B-EG）	**0.5781** 0.2261	0.5945 0.2238	0.6051 0.2225	0.6145 0.2215
	CoMeta（不含 S-EG）	0.5779 0.2259	0.5948 0.2239	0.6049 0.2226	0.6140 0.2216
	CoMeta	**0.5781 0.2258**	**0.5982 0.2234**	**0.6078 0.2223**	**0.6165 0.2213**

（不含 S-EG）：不含 S-EG 的 CoMeta。4) CoMeta：整体框架。如表3所示，一个有趣的发现是，在 MovieLens-1M 上，B-EG 的贡献大于 S-EG；而在淘宝广告上，两者的贡献相近。通常，对于新物品而言，其旧的相似物品所包含的有用信息，比其属性特征和交互用户所包含的信息更多。但与 MovieLens-1M 相比，淘宝广告中的噪声交互更多，而且我们无法准确计算淘宝广告中新物品的基础嵌入。整体而言，CoMeta 的表现最佳，这表明每个组件对 CoMeta 都很重要。

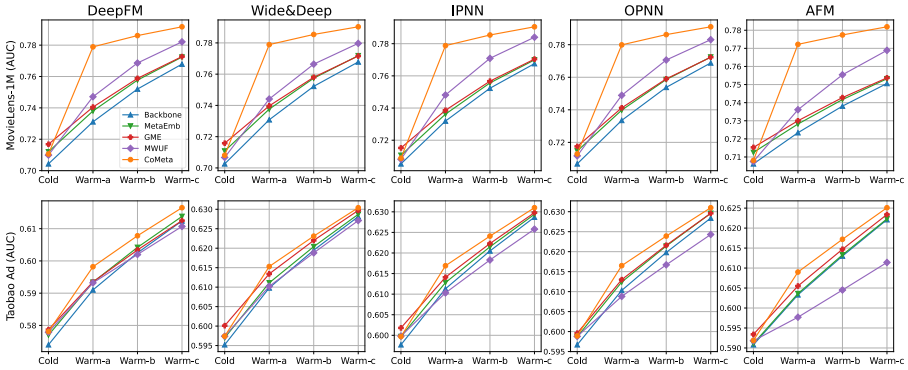

图3. 五种常用骨干模型在两个数据集上的曲线下面积。

方法兼容性。由于 CoMeta 与模型无关，我们在更多骨干模型上进行了实验。图 3 展示了五种具有代表性的骨干模型（DeepFM、Wide&Deep、IPNN、OPNN、AFM）在两个数据集上的结果。我们观察到，使用不同的骨干模型时，CoMeta 都能显著提升新物品的推荐性能。总体而言，AFM 的表现不如其他骨干模型，因为 AFM 无法捕捉高阶交互。

5 结论

本文着重解决物品冷启动问题。我们提出了 CoMeta，利用协同信息为新的物品 ID 生成理想的元嵌入。在两个数据集上的实验结果表明，我们提出的 CoMeta 优于最先进水平的冷启动方法，并具有良好的兼容性。未来，我们希望将这项工作拓展到更多场景（例如，用户冷启动问题）。

参考文献

1. Breese, J.S., Heckerman, D., Kadie, C.：协同过滤预测算法的实证分析。arXiv 预印本 arXiv:1301.7363 (2013) 2. Cheng, H.T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., An- derson, G., Corrado, G., Chai, W., Ispir, M., 等：面向推荐系统的宽度与深度学习。载于：第一届推荐系统深度学习研讨会论文集。第 7–10 页 (2016) 3. Dai, S., Lin, H., Zhao, Z., Lin, J., Wu, H., Wang, Z., Yang, S., Liu, J.：Poso：面向大规模推荐系统的个性化冷启动模块。arXiv 预印本 arXiv:2108.04690 (2021) 4. Du, X., Wang, X., He, X., Li, Z., Tang, J., Chua, T.S.：如何学习冷启动多媒体推荐中的物品表示？载于：第 28 届 ACM 多媒体国际会议论文集。第 3469–3477 页 (2020) 5. Finn, C., Abbeel, P., Levine, S.：用于快速适应深度网络的模型无关元学习。载于：国际机器学习会议。第 1126–1135 页。PMLR (2017) 6. Guo, H., Tang, R., Ye, Y., Li, Z., He, X.：DeepFM：一种基于因子分解机的点击率预测神经网络。arXiv 预印本 arXiv:1703.04247 (2017) 7. Kingma, D.P., Welling, M.：自编码变分贝叶斯。arXiv 预印本 arXiv:1312.6114 (2013) 8. Lee, H., Im, J., Jang, S., Cho, H., Chung, S.：MeLU：用于冷启动推荐的元学习用户偏好估计器。载于：第 25 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集。第 1073–1082 页 (2019) 9. Linden, G., Smith, B., York, J.：Amazon.com 推荐：物品到物品协同过滤。IEEE Internet Computing **7**(1), 76–80 (2003) 10. Ouyang, W., Zhang, X., Ren, S., Li, L., Zhang, K., Luo, J., Liu, Z., Du, Y.：学习图元嵌入，用于点击率预测中的冷启动广告。载于：第 44 届 ACM SIGIR 信息检索研究与发展国际会议论文集。第 1157–1166 页 (2021) 11. Pan, F., Li, S., Ao, X., Tang, P., He, Q.：为冷启动广告预热：通过学习学习 ID 嵌入来提升点击率预测。载于：第 42 届 ACM SIGIR 信息检索研究与发展国际会议论文集。第 695–704 页 (2019) 12. Qu, Y., Cai, H., Ren, K., Zhang, W., Yu, Y., Wen, Y., Wang, J.：用于用户响应预测的基于产品的神经网络。载于：2016 年 IEEE 第 16 届数据挖掘国际会议 (ICDM)。第 1149–1154 页。IEEE (2016) 13. Rendle, S.：因子分解机。载于：2010 年 IEEE 数据挖掘国际会议。第 995–1000 页。IEEE (2010)

14. Rendle, S., Freudenthaler, C., Gantner, Z., Schmidt-Thieme, L.：BPR：基于隐式反馈的贝叶斯个性化排序。arXiv 预印本 arXiv:1205.2618 (2012) 15. Rong, D., He, Q., Chen, J.：联邦学习场景下基于深度学习的推荐模型投毒。载于：Raedt, L.D.（编），第三十一届人工智能国际联合会议论文集，IJCAI 2022，维也纳，奥地利，2022 年 7 月 23–29 日。第 2204–2210 页。ijcai.org (2022)。https://doi.org/10. 24963 306 16 2022 306, 10 24963 2022 /ijcai. / https://doi.org/ . /ijcai. / 。Rong, D., Ye, S., Zhao, R., Yuen, H.N., Chen, J., He, Q.：FedRecAttack：针对联邦推荐的模型投毒攻击。载于：2022 年 IEEE 第 38 届数据工程国际会议 (ICDE)。第 2643–2655 页。IEEE (2022) 16. Rong, D., Ye, S., Zhao, R., Yuen, H.N., Chen, J., He, Q.：FedRecAttack：针对联邦推荐的模型投毒攻击。载于：2022 年 IEEE 第 38 届数据工程国际会议 (ICDE)。第 2643–2655 页。IEEE (2022) 17. Schein, A.I., Popescul, A., Ungar, L.H., Pennock, D.M.：冷启动推荐的方法与指标。载于：第 25 届 ACM SIGIR 信息检索研究与发展年度国际会议论文集。第 253–260 页 (2002) 18. Shi, S., Zhang, M., Yu, X., Zhang, Y., Hao, B., Liu, Y., Ma, S.：针对内容特征值缺失的推荐自适应特征采样。载于：第 28 届 ACM 信息与知识管理国际会议论文集。第 1451–1460 页 (2019) 19. Volkovs, M., Yu, G., Poutanen, T.：DropoutNet：解决推荐系统中的冷启动问题。神经信息处理系统进展 **30** (2017) 20. Wang, L., Jin, B., Huang, Z., Zhao, H., Lian, D., Liu, Q., Chen, E.：用于冷启动推荐的偏好自适应元学习。载于：IJCAI。第 1607–1614 页 (2021) 21. Wang, R., Fu, B., Fu, G., Wang, M.：用于广告点击预测的深度交叉网络。载于：ADKDD'17 论文集，第 1–7 页 (2017) 22. Xiao, J., Ye, H., He, X., Zhang, H., Wu, F., Chua, T.S.：注意力因子分解机：通过注意力网络学习特征交互的权重。arXiv 预印本 arXiv:1708.04617 (2017) 23. Zhang, Q., Li, J., Jia, Q., Wang, C., Zhu, J., Wang, Z., He, X.：UNBERT：用于新闻推荐的用户—新闻匹配 BERT。载于：IJCAI。第 3356–3362 页 (2021) 24. Zhao, K., Zheng, Y., Zhuang, T., Li, X., Zeng, X.：通过统一的图神经网络联合学习电子商务搜索与推荐。载于：第十五届 ACM 网络搜索与数据挖掘国际会议论文集。第 1461–1469 页 (2022) 25. Zhao, X., Ren, Y., Du, Y., Zhang, S., Wang, N.：通过模型无关条件变分自编码器改进物品冷启动推荐。arXiv 预印本 arXiv:2205.13795 (2022) 26. Zhou, G., Mou, N., Fan, Y., Pi, Q., Bian, W., Zhou, C., Zhu, X., Gai, K.：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 5941–5948 页 (2019) 27. Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., Yan, Y., Jin, J., Li, H., Gai, K.：用于点击率预测的深度兴趣网络。载于：第 24 届 ACM SIGKDD 知识发现与数据挖掘国际会议论文集。第 1059–1068 页 (2018) 28. Zhu, Y., Xie, R., Zhuang, F., Ge, K., Sun, Y., Zhang, X., Lin, L., Cao, J.：通过元偏移和缩放网络学习预热冷物品嵌入，以实现冷启动推荐。载于：第 44 届 ACM SIGIR 信息检索研究与发展国际会议论文集。第 1167–1176 页 (2021)