

MoGE: 基于图上下文增强的多任务推荐算法

周 俊^{1,2}, 胡斌斌², 张志强², 陈超超¹

(1. 浙江大学计算机科学与技术学院, 浙江杭州 310027; 2. 蚂蚁科技集团股份有限公司机器智能部, 浙江杭州 310000)

摘要: 多任务学习 (Multi-Task Learning, MTL) 通过信息共享来共同处理多个任务, 已被广泛应用于大量推荐任务中。目前针对推荐的多任务学习方法, 主要集中在基于共享输入特征 (即描述用户-商品交互信息的特征工程) 的多门控混合专家网络 (Multi-gate Mixture-of-Experts, MMoE), 以此来学习不同任务间的关联。最近的一些工作表明, 图神经网络 (Graph Neural Network, GNN) 作为表征深度交互上下文的强大工具, 被应用于推荐任务中, 可极大地缓解在线个性化推荐服务中的数据稀疏问题。因此, 我们通过设计混合图增强专家网络 (Mixture of Graph enhanced Expert networks, MoGE), 首次探索了用于多任务推荐的图神经网络结构。具体地说, 我们提出了一种新的多通道图神经网络, 利用用户-商品二部图, 以及衍生的用户和商品的协同相似图来联合建模用户-商品的高阶交互信息。在学习到的深层次交互上下文的基础上, 引入了一组图增强的专家网络, 以协作的方式实现多任务推荐。在 3 个真实数据集上的实验结果表明, MoGE 在所有目标任务上都持续且显著地优于最优的基线。

关键词: 多任务学习; 图学习; 推荐算法; 注意力机制; 深度学习

基金项目: 国家自然科学基金 (No.72192823, No.62172362)

中图分类号: TP183

文献标识码: A

文章编号: 0372-2112(2023)11-3377-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220964

MoGE: Graph Context Enhanced Multi-Task Recommendation Method

ZHOU Jun^{1,2}, HU Bin-bin², ZHANG Zhi-qiang², CHEN Chao-chao¹

(1. College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang 310027, China;

2. Department of Machine Intelligence, Ant Group, Hangzhou, Zhejiang 310000, China)

Abstract: Multi-task learning (MTL), which jointly tackles multiple tasks through information sharing, has been extensively used in a variety of recommendation applications. Recently, current efforts targeted for recommendation focus on learning task relationships based on the multi-gate mixture-of-experts (MMoE) architecture with shared input features (i.e., subtle feature engineering for user-item interaction). Recent evidences suggest the graph neural network (GNN) as a powerful component in characterizing deep interaction context for recommendation, greatly contributing to easing the data sparseness issue in online advertising services. Hence, we make the first attempt to explore the GNN towards multi-task recommendation, by designing mixture of graph enhanced expert networks (MoGE). Specifically, we propose a novel multi-channel graph neural network to jointly model high-order information with the user-item bipartite graph as well as derived collaborative similarity graphs for users and items. A group of graph augmented expert networks are introduced on top of the learnt deep interaction context for cooperatively contributing to the multi-task recommendation. Experimental results on three real-world datasets show that MoGE consistently and significantly outperforms state-of-the-art baselines across all target tasks.

Key words: multi-task learning; graph learning; recommender system; attention mechanism; deep learning

Foundation Item(s): National Natural Science Foundation of China (No.72192823, No.62172362)

1 引言

近年来, 基于多任务学习的神经网络架构^[1]被广泛应用于各种工业应用中, 从实体检测^[2]到在线广告和推

荐^[3-9]。事实上, 最近将多任务学习整合到推荐系统中的一些工作已经证明了通过联合优化多个目标行为 (如点击和购买), 可以更好地捕捉用户的固有

偏好^[3,10-14].

因此,最近的一些研究提出了通过不同任务之间知识的有效迁移来学习多个目标^[3,10,11]. 其中最具有代表性的工作就是多门控的混合专家网络(Multi-gate Mixture-of-Experts, MMoE)结构^[15]以及它的变体^[4,16]. 具体地说,它由多个具有相应门控网络的专家网络组成,通过共享的输入特征对任务关系进行显式建模,并取得了相当优异的性能提升^[12-14]. 令人遗憾的是,它的成功非常依赖细致的特征工程和大量的标注数据,而这些在真实的工业系统中很难轻易获

得^[17]. 如图 1 所示,在支付宝这样的真实的流量分发平台上,大部分商品的点击/转化频次都较低,从而产生了令人不满意的数据稀疏性问题. 虽然目前的多任务推荐方法可以有效捕获多任务之间的知识迁移(即点击和转化之间),但在实际的工业环境下,获取足够的训练数据和长尾商品的特征仍然是很难实现的. 幸运的是,随着越来越多的用户-商品交互历史记录可以被平台收集利用,描述基于用户-商品的高阶连通性和基于用户(商品)的协同相似性有助于缓解数据稀疏问题,并促进多任务推荐.

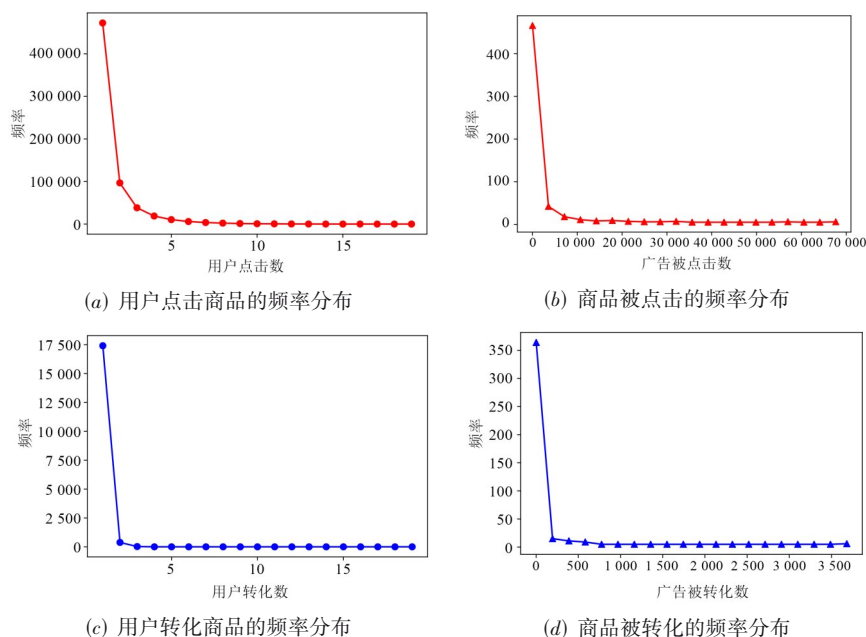


图1 用户点击/转化商品的频率分布及商品被点击/转化的频率分布

虽然从多任务推荐的交互历史中提取图上下文,即高阶连通性和协同相似性的想法很有吸引力,但解决方案并不简单,存在两个挑战:

(1)如何灵活地提取最相关/最重要的信息来表征用户和商品之间的交互? 用户-商品之间的高阶连通性和协同相似性是复杂的. 因此,通过自动探索和利用有用的图结构信息,并设计一个统一的模块来产生高质量的用户和商品的表征是势在必行的.

(2)如何有效地将学习到的深层次的交互上下文用于多任务推荐? 学习到的深层次的交互上下文由基于不同视图(即连通性和相似性)的表征组成,这促使我们开发一种新的体系结构,将这些不同视图信息以合作的方式融入多任务推荐.

最近一些研究表明,图神经网络^[18-20]在图结构特征学习方面具有卓越的能力. 因此,相应的方法被广泛应用于推荐系统中,用于表征交互上下文^[21,22]和对丰富的辅助信息^[23,24]进行建模. 在这些工作的启发下,我

们通过设计混合图增强专家网络(Mixture of Graph enhanced Expert networks, MoGE),首次尝试研究图神经网络对多任务推荐的有效性. 为了同时捕获基于用户-商品的高阶连通性和基于用户(物品)的协同相似性,我们提出了一种新的多通道图神经网络(Multi-channel GNN, MGNN)模块,可以自适应地学习用户-商品二部图以及衍生的用户(商品)协同相似性图最相关/最重要的信息. 在学习到的深层次的交互上下文的基础上,我们引入了基于MMoE架构的图增强专家网络,该网络可以灵活地融合这些上下文,来促进多任务推荐. 为了更好地指导MoGE的学习过程,我们采用自适应的方式来衡量每个任务的损失,并在训练过程平衡多个并行专家网络对预测目标的贡献.

综上所述,本文主要贡献如下:

(1)首次研究了图神经网络对多任务推荐的有效性,从模型和实验的角度验证图神经网络对于刻画用户和商品之间的深层次交互上下文的潜力.

(2)提出了一种新型的基于图上下文的多任务推荐框架—MoGE,该框架配备了多通道图神经网络和图增强专家网络,可以灵活地利用图结构来促进多任务推荐.

(3)分别在基准数据集和工业数据集上进行了大量的实验.实验结果表明,我们提出的 MoGE 在所有数据集上都一致且显著地优于一系列最先进的多任务推荐框架.

2 相关工作

多任务学习^[1,25]可以在多个任务中学习其中的关联性和差异性.在推荐平台和在线广告服务中,用户对某一商品通常具有多种目标行为(如点击和购买),因此多任务学习被普遍用于刻画这种复杂的交互场景.为了解决在线服务中的数据稀疏性问题,一些研究^[3,10,11]将用户的多种行为考虑成多任务问题,并采用特征迁移学习策略进行多任务推荐.另外,为了显式地从数据中建模多任务间的关系,多门控的混合专家网络(MMoE)框架^[15]及其变体^[4,16]已经在推荐任务中产生了很有竞争力的性能.然而,这些多任务推荐方法在实际的工业系统中仍然不足以达到令人满意的性能,因为忽略了从用户-商品的历史交互记录的高阶结构信息.

近年来,图神经网络在图数据的结构学习和特征传播方面显示出了巨大的潜力^[18-20].由于推荐系统中基于用户-商品的交互记录可以看作典型的二部图,因此图神经网络通过有效捕获用户和商品之间的高阶交互而被广泛应用于推荐任务中^[21,22].其中一类主要工作是将不同的精心设计的图神经网络框架直接应用到二部图上进行高阶结构学习,包括基于随机游走的图神经网络^[22]和基于多层交互的图神经网络^[21].随着我们可以在网络服务中获取的辅助信息越来越多,有很多采用图神经网络将这些上下文信息融入推荐系统的工作出现,包括社交关系^[23,26-28]和知识图谱^[24].然而,关于图神经网络对多任务推荐的有效性始终还未被充分探索.

3 预备知识

在本文中,我们考虑了多任务学习框架中的推荐任务.我们假设一组用户-商品的用户交互记录,记为 $\mathcal{H}=\{u, i, y_{u,i} \mid u \in \mathcal{U}, i \in \mathcal{I}, y_{u,i} \in \{0, 1\}^K\}$. 这里, \mathcal{U} 和 \mathcal{I} 分别表示用户和商品集合,而且每个 $\langle u, i \rangle$ 会有 K 个相关的标签(如大多数电商场景中的点击和转化标签),记为 $y_{u,i}$. 除此之外,每个用户和商品都会关联对应的特征向量,即 $x_u \in R^{d_u}$ 和 $x_i \in R^{d_i}$.

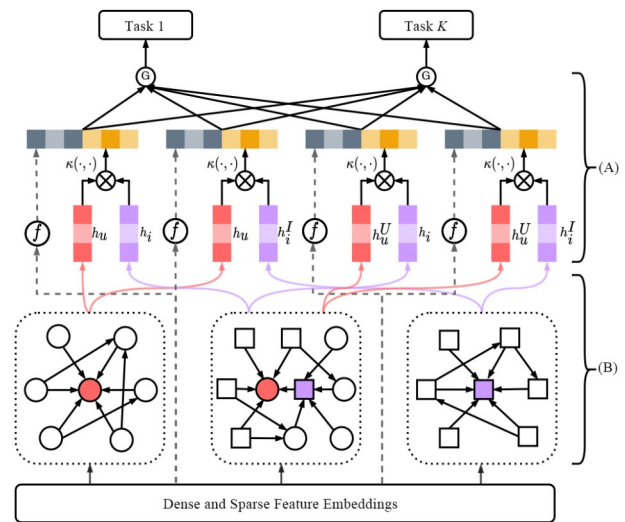
根据收集到的交互记录,我们定义用户-商品二部

图 $\mathcal{G}=\{\mathcal{U}, \mathcal{I}, \mathcal{E}\}$, 其中, \mathcal{U} 和 \mathcal{I} 分别表示用户和商品集合. $\mathcal{E} \subset \mathcal{U} \times \mathcal{I}$ 表示集合 \mathcal{U} 和 \mathcal{I} 之间的边集合,这里, $e_{u,i} = e_{i,u} = 1$ 当且仅当 $\sum_{k=1}^K y_{u,i}^k \geq 1$.

为了有效利用深层次的交互上下文进行推荐,我们进一步定义了具有相似点击(被点击)和购买(被购买)历史的用户(商品)的协同相似性.具体地,我们将基于用户的协同相似图定义为 $\mathcal{G}^u=\{\mathcal{U}, \mathcal{E}^u\}$, 其中 $\mathcal{E}^u \subset \mathcal{U} \times \mathcal{U}$ 是边集合,且 $e_{u,v} = e_{v,u} = \text{sim}(u, v)$, 该相似度由用户的共同交互记录求得^[18]. 类似地,我们也将基于商品的协同相似图表示为 \mathcal{G}^i .

4 MoGE 模型

在本节中,我们将介绍 MoGE,这是一种新颖的多任务推荐方法,通过图学习来利用深层次的交互上下文. MoGE 的整体架构如图 2 所示.



注:图中 MoGE 由多通道图神经网络模块(A)和图增强专家网络(B)组成

图 2 MoGE 的总体架构

4.1 深层次的交互上下文利用

为了充分利用 \mathcal{G} 的高阶连通性以及 \mathcal{G}^u (基于用户的协同相似图) 和 \mathcal{G}^i (基于商品的协同相似图) 的协同相似性,我们提出了一种多通道图神经网络(MGNN),来有效地聚合每个通道(即图)中信息量最大的邻居.

为了避免过度参数化的问题,我们提出的 MGNN 在每个通道上应用图机器学习时,用户和商品会具有(在各个通道上)共享的初始表示.为此,我们设置了一个共享的变换层,将原始特征空间映射为低维的稠密向量.这个向量会在接下来二部图和协同相似图的传播和聚合过程中共享.特别地,我们的变换层对应两个参数矩阵 $P \in R^{d_u \times d}$ 和 $Q \in R^{d_i \times d}$, 其中 d 为表征的维度.变换算子实现方式如下:

$$\begin{aligned} p_u &= x_u \cdot P, \\ q_i &= x_i \cdot Q. \end{aligned}$$

总的来说,目标用户 u 与商品 i 交互的潜在原因是相当复杂的,可能不仅与历史交互记录(即 \mathcal{G})有关,还可能与用户(即 \mathcal{G}^U)与商品(即 \mathcal{G}^I)之间的深层协同相似性有关.形式化地,我们令 $h_u^0 = h_u^{\mathcal{L}0} = p_u$, $h_i^0 = h_i^{\mathcal{L}0} = q_i$ 为 MGNN 模块的第一层的输入.然后, MGNN 层在每个通道内,通过邻域聚合的方式递归整合不同图的连通性或协同相似性.在这里,我们将进一步地介绍 MGNN 的整个工作流程.

MGNN 捕获 \mathcal{G} 中的用户-商品的连通性.给定第 l 层的用户和商品的输入 h_u^l 和 h_i^l ,我们通过如下的方式在 \mathcal{G} 中来更新用户和商品在第 $l+1$ 层的表征(即 h_u^{l+1} 和 h_i^{l+1}):

$$\begin{aligned} h_u^{l+1} &= f\left(h_u^l, \sigma\left(W \sum_{i \in \mathcal{N}(u)} \alpha_{u \leftarrow i} h_i^l\right)\right), \\ h_i^{l+1} &= f\left(h_i^l, \sigma\left(W \sum_{u \in \mathcal{N}(i)} \alpha_{i \leftarrow u} h_u^l\right)\right). \end{aligned}$$

其中, $\sigma(\cdot)$ 是 LeakyReLU 激活函数; $W \in R^{d \times d}$ 为 \mathcal{G} 的权重矩阵; $\mathcal{N}(u)$ 和 $\mathcal{N}(i)$ 表示用户 u 和商品 i 在 \mathcal{G} 中的邻居集合; $\alpha_{u \leftarrow i}$ 和 $\alpha_{i \leftarrow u}$ 是由新兴的注意力机制实现^[29],分别指示着通过连接 $u \leftarrow i$ 和 $i \leftarrow u$ 传播的信息量; $f(\cdot)$ 旨在聚合从对应邻居传播的信息,通常被实现为类似 LSTM 的算子^[30].

MGNN 捕获 $\mathcal{G}^U/\mathcal{G}^I$ 中的基于用户(商品)的协同相似性.对于任意一个用户 u (商品 i),给定它在第 l 层的表征 $h_u^{\mathcal{L}l}/h_i^{\mathcal{L}l}$,更新后的表征 $h_u^{\mathcal{L}l+1}/h_i^{\mathcal{L}l+1}$ 可由如下的方式计算得到:

$$\begin{aligned} h_u^{\mathcal{L}l+1} &= f\left(h_u^{\mathcal{L}l}, \sigma\left(W^U \sum_{u \in \mathcal{N}^U(u)} \alpha_{u \leftarrow u} h_u^{\mathcal{L}l}\right)\right), \\ h_i^{\mathcal{L}l+1} &= f\left(h_i^{\mathcal{L}l}, \sigma\left(W^I \sum_{i \in \mathcal{N}^I(i)} \alpha_{i \leftarrow i} h_i^{\mathcal{L}l}\right)\right). \end{aligned}$$

其中, W^U 和 $W^I \in R^{d \times d}$ 分别为 \mathcal{G}^U 和 \mathcal{G}^I 的权值矩阵, $\mathcal{N}^U(u)$ 和 $\mathcal{N}^I(i)$ 分别为用户 u 和商品 i 在 \mathcal{G}^U 和 \mathcal{G}^I 中的邻居集合, $\alpha_{u \leftarrow u}$ 和 $\alpha_{i \leftarrow i}$ 控制着 \mathcal{G}^U 和 \mathcal{G}^I 上的信息传播.

通过如上过程的传播过程, MGNN 将第 l 层不同图上的相关信息聚合在一起,优化了每个通道的图表征(即 h_u^{l+1} , h_i^{l+1} , $h_u^{\mathcal{L}l+1}$, $h_i^{\mathcal{L}l+1}$).

4.2 图增强的专家网络

本节研究如何将学习到的图表征集成到多任务推荐中.如之前的研究工作^[12,15,31]所证明的,专家网络的设计保证了在建模任务联系和冲突时的软参数共享.因此,我们在多门控混合专家网络(MMoE)架构的基础

上,结合图学习来促进多任务推荐.

直观上来说,用户和商品之间的交互不仅受用户-商品的二部图影响,也受相应的协同相似图的作用.给定一个用户-商品对 $\langle u, i \rangle$,经过 L 层的传播,我们可以得到最终的图表征列表 $\{h_u^L, h_i^L, h_u^{\mathcal{L}L}, h_i^{\mathcal{L}L}\}$,它们在不同的图中共同捕获用户的偏好和商品的属性.为了充分利用来自不同图的信息,我们提出在用户 u 与商品 i 之间以成对的方式应用图表征的交互:

$$\mathcal{Z} = \left\{ \kappa(h_u^L, h_i^L), \kappa(h_u^L, h_i^{\mathcal{L}L}), \kappa(h_u^{\mathcal{L}L}, h_i^L), \kappa(h_u^{\mathcal{L}L}, h_i^{\mathcal{L}L}) \right\}$$

其中, $\kappa(\cdot, \cdot)$ 表示交互函数,可以设置为按元素乘,拼接或者多层感知机.

在本文中,为了方便起见,令 $|\mathcal{Z}| = N$,这也是本文模型中专家网络的数目.将专家网络表示为 $\{f^n(\cdot)\}_{n=1}^N$,并且每个任务 k 都会对应一个独立的门控网络 $g^k(\cdot)$,因此,任务 k 的输出可以表示为

$$\begin{aligned} \hat{x}_{u,i} &= [\mathbf{x}_u \| \mathbf{x}_i], \\ o_{u,i}^k &= \sum_{n=1}^N g^k(\hat{x}_{u,i}^n) [f^n(\hat{x}_{u,i}^n) \| \mathcal{Z}^n], \\ \hat{y}_{u,i}^k &= v^T o_{u,i}^k. \end{aligned}$$

其中,“ $\|$ ”表示拼接操作, v 是输出的权重向量.遵循以往相关研究^[12,15,31]中常见的策略,第 n 个专家网络 $f^n(\cdot)$ 被实现为带有 LeakyReLU 激活函数的多层感知机,而第 k 个门控网络 $g^k(\cdot)$ 则被实现为具有 Softmax 层的简单线性变换.

4.3 模型学习

最后,本文通过联合优化多个相关任务来指导 MoGE 的学习过程.对于第 k 个任务,采用常用的交叉熵作为损失函数.因此,本文模型的主损失函数可以表示为

$$\mathcal{L}_{\text{main}} = \sum_{k=1}^K \mathcal{L}^k = \sum_{k=1}^K \sum_{\langle u,i \rangle \in \mathcal{H}} \mathcal{C}(\hat{y}_{u,i}^k, y_{u,i}^k).$$

其中 $\mathcal{C}(\cdot, \cdot)$ 为交叉熵函数.受同方差不确定性理论^[32]的启发,我们在模型训练过程中自适应地权衡每个任务的损失.具体来说,遵循文献^[33,34],根据依赖任务的不确定性为每个任务学习一个相对权值,因此将上述损失函数重写为

$$\begin{aligned} \mathcal{L}_{\text{main}}(\Theta, \Sigma) &= \sum_{k=1}^K \mathcal{L}^k(\theta^k, \sigma^k) \\ &\approx \sum_{k=1}^K \sum_{\langle u,i \rangle \in \mathcal{H}} \left(\frac{1}{\sigma^{k2}} \mathcal{C}(\hat{y}_{u,i}^k, y_{u,i}^k) + \log(\sigma^{k2}) \right). \end{aligned}$$

其中, $\Theta = \{\theta^k\}_{k=1}^K$, $\Sigma = \{\sigma^k\}_{k=1}^K$ 分别为模型的参数集合和任务不确定性集合.注意 $\log(\sigma^{k2})$ 可以视为防止模型过拟合的规则化项.上述任务依赖的不确定性损失函数的详细推导流程见附录.

如前所述, MoGE 包含多个并行的专家网络, 我们希望这些专家网络可以以合作的方式来助力最终的预测^[35,36]. 为此, 本文提出了一种辅助的协同损失来平衡每个并行专家网络的负载:

$$\mathcal{L}_{co} = - \sum_{\langle u,i \rangle \in \mathcal{H}} \sum_{k=1}^K \sum_{n=1}^N \sum_{m=n+1}^N g^k(\hat{x}_{u,i})_n g^k(\hat{x}_{u,i})_m.$$

将以上两种损失结合在一起, 本文模型的整体损失函数定义如下:

$$\mathcal{L} = \mathcal{L}_{main} + \beta \mathcal{L}_{co} + \lambda \|\Theta\|^2.$$

4.4 模型讨论

(1) MoGE 的灵活性. 本文提出的 MoGE 是一个灵活的框架, 利用图结构来促进多任务推荐. 一方面, 如果我们忽略 MGNN 模块学习到的深层交互上下文, MoGE 会退化为典型的 MME 模型^[15]. 另一方面, 当忽略了协同相似度和图增强专家网络时, MoGE 也可以自然地退化为神经图协同过滤模型 (Neural Graph Collaborative Filtering, NGCF)^[21]. 此外, MoGE 也是一个通用的框架, 可以很容易地进行扩展来包含更复杂的图结构 (如用户社交图谱和商品知识图谱).

(2) 在工业应用中的高效训练. 为了在 MGNN 模块中进行高效的数值计算, 我们遵循文献^[37]中的技巧在每个通道中执行注意力值的计算. 因此, 复杂度和存储都与每个通道对应的图的边数呈线性关系. 总的来说, MGNN 的复杂度上界为 $\mathcal{O}(|\mathcal{E}| + |\mathcal{E}^u| + |\mathcal{E}^i|)$. 此外, 遵循文献^[38]中类似的策略, 分批为每个 $\langle u, i \rangle$ 对随机收集多跳 (即 L) 邻居. 在 MoGE 的训练过程中, 我们给 MGNN 输入的是抽样的多跳邻居组织的子图, 而不是完整的图, 这赋予了 MoGE 在现实应用中的大规模数据上高效训练的能力. 具体的训练流程如算法 1 描述. 实际上, MoGE 已经部署在工业推荐系统中, 支持由数亿节点和边组成的用户-商品交互图.

5 实验分析

本节进行了一系列实验, 分别评估 MoGE 在工业和公开的基准数据集上的有效性.

5.1 数据集介绍

推荐数据集 (工业级场景). 我们在支付宝平台收集了一系列用户点击商品的流量日志, 提取了 Alipay A 和 Alipay B 两个子数据集做模型评估. 另外, 我们构建了一个大规模的统一图来表征用户和商家的深层交互上下文, 其中每个节点关联 93 个属性, 包括用户、商家和场景属性. 在这个数据集中, 有两个任务目标需要学习: 点击 (预测用户是否会点击目标商家) 和转化 (预测用户是否会成为目标商家的会员).

IJCAI CUP2015 数据集 (基准). 这个公共数据集由

算法 1 MoGE 的训练流程

输入: 用户商品二部图 $\mathcal{G} = \{\mathcal{U}, \mathcal{I}, \mathcal{E}\}$, 用户和商品的历史交互记录 $\mathcal{H} = \{u, i, y_{u,i} \mid u \in \mathcal{U}, i \in \mathcal{I}, y_{u,i} \in \{0, 1\}^K\}$, 用户特征 $\{x_u \mid u \in \mathcal{U}\}$, 商品特征 $\{x_i \mid i \in \mathcal{I}\}$

输出: MoGE 的模型参数 Θ, Σ

1. 根据协同相似性获取 \mathcal{G}^u 和 \mathcal{G}^i
2. 初始化变换层参数矩阵 $\mathbf{P} \in \mathbf{R}^{d_u \times d}$ 和 $\mathbf{Q} \in \mathbf{R}^{d_i \times d}$, MGNN 模块的全局参数以及图专家网络的相应参数
3. WHILE 模型未收敛 DO:
4. 从 \mathcal{H} 中采样一批交互记录 \mathcal{H}^B
5. /**MGNN 模块**/
6. 分别在 \mathcal{G}^u 和 \mathcal{G}^i 中为 \mathcal{H}^B 中的每个用户/商品采样 L 度邻居
7. 通过共享编码为 \mathcal{H}^B 中的每个用户/商品生成初始表征
8. FOR $l=1; l \leq L$ DO
9. 在每个通道 (即图) 上进行聚合和评估
10. END FOR
11. /**图增强的专家网络**/
12. 以成对的方式应用图表征的交互
13. FOR $k=1; k \leq K$ DO
14. 计算每个任务上的输出 $\hat{y}_{u,i}^k$
15. END FOR
16. 计算整体损失并反向传播更新参数
17. END WHILE
18. RETURN MoGE 的模型全部参数

匿名用户 “双 11” 之前 6 个月以及当天的购物日志组成. 具体地, 使用 07/09 到 10/08 的数据进行图表构建, 10/09 到 10/22 进行训练, 10/23 到 10/25 进行测试. 此外, 我们为这个数据集构造了两个任务: 收藏 (预测用户是否将目标商品添加到收藏列表中) 和购买 (预测用户是否将购买目标商品).

3 个数据集的详细数据统计汇总在表 1 中.

表 1 实验数据集的统计信息

	Alipay A	Alipay B	IJCAI CUP2015
#Train	2.67×10^6	2.18×10^6	3.81×10^6
#Test	1.45×10^5	1.72×10^5	9.69×10^5
#Nodes	4.01×10^7		1.07×10^6
#Edges	4.31×10^8		3.94×10^6
#Attributes	93		—
Tasks	点击/转化		喜欢/购买

5.2 对比方法

我们主要考虑 2 种具有代表性的预测方法及其变体.

5.2.1 单任务模型

MLP: 该方法基于共享特征学习用户和商品之间的交互函数.

NGCF^[21]: 该方法在用户-商品交互图上采用了多

层神经网络,旨在通过多跳的邻居信息来优化用户和商品的表示.

5.2.2 多任务模型

Shared-Bottom^[39]:该方法是最基本和最常用的多任务学习结构,其中的参数在不同的任务之间直接共享.

Cross-Stitch^[2]:该方法提出学习静态线性组合来融合不同任务的表示.

ESMM^[3]:该方法采用概率迁移的模式设计,通过对整个空间的训练来解决非端到端点击后转化率预估中样本选择偏差问题.

MMoE^[15]:该方法集成了多个专家网络,然后通过每个任务对应的门控网络来整合多个专家的知识.

SNR-trans^[16]:该方法将共享的底层的隐藏层模块化为多层子网络,并控制具有可学习潜在变量的子网络连接,实现了更加灵活的参数共享.

CGC^[4]:该方法使用递进分层抽取的模式将任务共享专家和任务特定专家明确地分离出来,当前在多任务推荐场景取得了最优的性能.

变体:为了证明 MGNN 模块学习到的图表示的有效性,我们准备了一系列上述多任务方法的变体,它们只是将 MGNN 模块的输出作为输入.

5.3 评价指标

我们采用了目前广泛使用的度量标准 GAUC (Group weighted Area Under Curve)^[3]来衡量所有推荐方法的性能,它是由经典的度量标准 AUC 衍生而来的. GAUC 的计算方法如下:

$$\text{GAUC} = \frac{\sum_g w_g \times \text{AUC}_g}{\sum_g w_g}.$$

其中, w_g 表示分组 g 的权重 (即分组 g 的样本数), AUC_g 是分组 g 的 AUC. 在实际推荐系统中, GAUC 是一个更细粒度的指标,被证明与线上性能更相关. 值得注意的是, 0.1% 的 GAUC 的提升在工业级场景中往往能取得很高的经济收益^[40-42].

5.4 实现细节

由于本文采用的数据集规模相当大,我们在基于参数服务器的分布式学习系统^[43]上实现了所有的模型. 为了公平比较,本文设置学习率为 10^{-4} , 正则化项为 10^{-3} , 批大小为 256, 表征维度为 64, 并选择 Adam 作为所有模型的优化器. 此外,对于 MLP 和 ESMM,我们设置它们的架构为 [512, 256, 128]. 对于 NGCF, 本文采用了原始论文中描述的最优架构,采用了 2 层图神经网络. 对于 Shared-Bottom, Cross-stitch, MMoE, SNR-trans 和 CGC, 本文设置专家网络的数量为 4, 设置每个专家网络的架构为 [32, 32], 每个门控网络为 [16, 16]. 对于

MoGE, 本文遵循与 MMoE 相同的专家网络和门控网络结构,并设置 $L=2$, $\beta=10^{-6}$, $\lambda=10^{-3}$.

5.5 性能对比

表 2 展示了所有方法在工业数据集和基准数据集上的总体性能. 其中,我们主要有以下几点发现:

(1) 在所有情况下,我们提出的 MoGE 模型一致且显著地优于所有基线. 这一结果验证了 MoGE 的有效性,其主要的增益来源于将深层次的交互上下文结合图学习用于多任务推荐.

(2) 在大多数情况下,多任务方法的变体 (即 +Graph) 比基础结构的性能要好得多,这归功于 MGNN 模块学习到的有用的图结构信息. 尽管如此,本文的 MoGE 提供了一个更合理的机制来结合图学习和多任务推荐,因此仍然优于这些变体.

(3) 在不考虑图结构的情况下,多任务方法在大多数情况下优于单任务方法,这说明学习任务关系在模型训练中是有用的. 这种现象在本文的工业数据集中相当明显,因为点击和转化之间的关系可能比收藏和购买更强.

5.6 消融实验 1: 不同交互函数的影响

本文考察交互函数 $\kappa(\cdot, \cdot)$ 的影响. 具体地,本文为 MoGE 模型实现了不同的交互函数,即按元素乘 (EP)、拼接 (CON)、按元素加 (ADD) 和多层感知器 (MLP), 并对比了相应的性能变化.

如图 3(a) 和图 3(b) 所示,当使用相对简单的交互函数 (即按元素乘或者按元素加) 时, MoGE 的性能相对较好. 一个可能的原因是,这样的交互似乎足以捕获用户和商品之间的联系,而且还增强了模型表征能力. 这一观察结果与之前研究^[21]的发现一致.

5.7 消融实验 2: 不同损失函数的影响

如前所述,为了有效学习多任务间的关联和差异,本文通过在原始多任务损失函数中添加任务依赖的不确定性损失函数和辅助的协同损失函数进行对比.

如图 4(a) 和图 4(b) 所示,完整的 MoGE (即 “i+ii+iii”) 与其他变体相比,在大多数情况下取得了最好的性能. 这个实验结果验证了本文损失函数设计的合理性和有效性,为多个任务间的知识迁移提供了一种更合理的学习方式.

5.8 消融实验 3: 图增强专家网络的作用

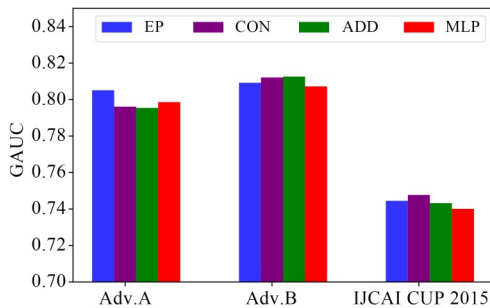
本文通过逐步将图增强的专家网络融入所提出的 MoGE 模型,分析他们对推荐性能的影响.

如图 5(a) 和图 5(b) 所示,加入图增强的专家网络 (如图 2(b) 所示,从左到右添加专家网络) 后, MoGE 的性能整体会得到提高. 在大多数情况下,如果没有图增强的专家网络,性能会下降很多. 这些观察结果证明了

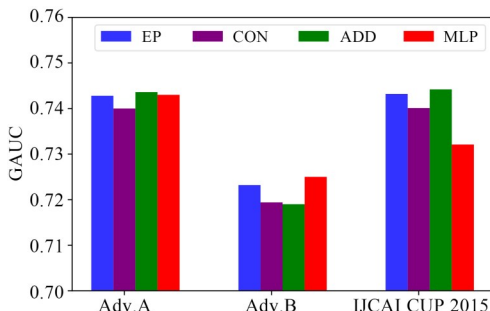
表 2 3 个数据集上整体性能比较

Dataset		Alipay A		Alipay B		IJCUP CUP2015	
Task		点击	转化	点击	转化	喜欢	购买
MLP		0.734 0	0.781 8	0.713 8	0.801 8	0.604 9	0.744 9
NGCF		0.738 3	0.781 1	0.720 7	0.802 7	0.743 1	0.746 6
Shared-Bottom	Base	0.728 8	0.785 7	0.705 1	0.801 5	0.604 4	0.746 2
	+Graph	0.733 3	0.787 2	0.718 9	0.807 9	0.743 9	0.735 1
Cross-Stitch	Base	0.731 4	0.785 7	0.710 8	0.806 3	0.604 9	0.744 5
	+Graph	0.730 9	0.786 6	0.717 5	0.802 7	0.737 8	0.745 7
ESMM	Base	0.726 1	0.786 2	0.700 9	0.797 0	0.605 5	0.744 8
	+Graph	0.736 3	0.787 9	0.717 7	0.806 7	0.741 7	0.743 8
MMoE	Base	0.728 8	0.786 0	0.709 2	0.803 3	0.601 3	0.736 9
	+Graph	0.737 0	0.788 0	0.714 0	0.803 5	0.731 4	0.732 9
SNR-trans	Base	0.732 7	0.788 9	0.715 9	0.805 9	0.604 0	0.741 6
	+Graph	0.735 2	0.786 3	0.718 5	0.802 1	0.736 6	0.738 5
CGC	Base	0.730 7	0.786 2	0.708 4	0.801 2	0.603 0	0.740 2
	+Graph	0.734 2	0.786 2	0.714 7	0.804 6	0.737 3	0.731 4
MoGE		0.742 8	0.805 1	0.723 2	0.809 2	0.744 2	0.747 7
RelaImpr v.s. Best		1.89%	5.61%	1.13%	0.42%	0.41%	0.44%

注:使用“**”(“或者*”)来表明 MoGE 相较最优基线提升的显著性水平为 0.01(或者 0.05);RelaImpr 为 relative improvement 的缩写。



(a) 点击/喜欢



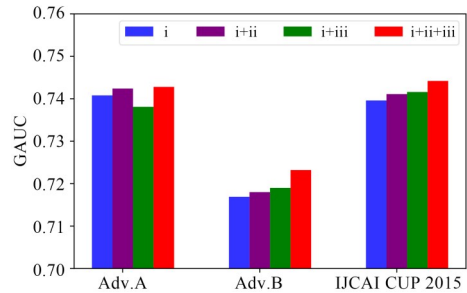
(b) 转化/购买

图 3 不同交互函数的影响

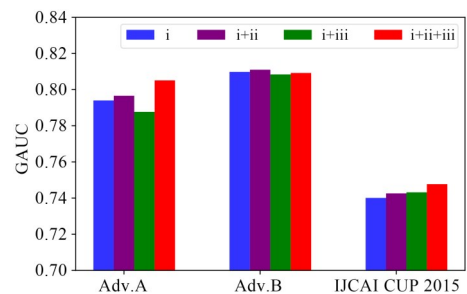
图学习对于缓解数据稀疏性问题的有效性。

6 系统部署

本节介绍 MoGE 模型在支付宝推荐平台的部署流程,如图 6 所示. 整个部署流程包含 2 个部分:线下训练和线上服务. 在线下训练过程中,用户的一系列点击/



(a) 点击/喜欢



(b) 转化/购买

注:i 为原始损失函数;ii 为不确定性损失函数;iii 为辅助的协作损失函数

图 4 不同损失函数的影响

转化行为都会以日志的形式存储到数据库中,作为训练样本生成和图构造的底层数据. 为了更好地使 MGNN 模块抽取深层次的上下文信息,我们会依次构造用户-商品二部图、用户协同相似图和商品协同相似图. 同时,我们会通过负采样策略从数据库中生成一系列训练实例. 然后,我们将这些训练实例和构造好的图

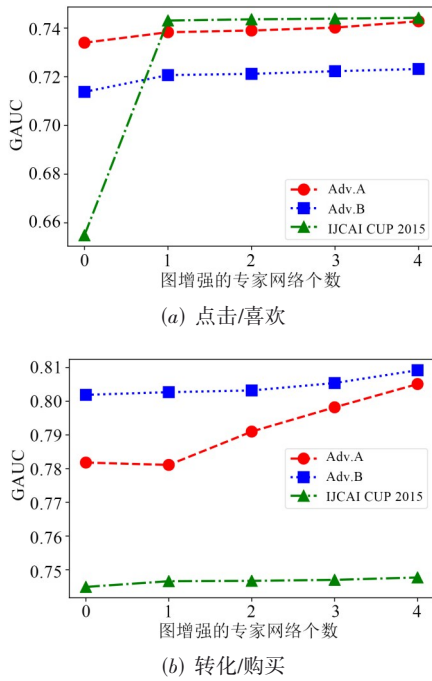


图5 图增强的专家网络的作用

数据送入 MoGE 进行训练^[44],并将最优模型上传到模型服务器.对于在线预测,给定一批新的用户,模型服务器会根据这些用户所具备的图特征进行预测.值得注意的是,模型训练和服务的图数据每天都会更新一次,以保证用户的属性和图结构保持最新.

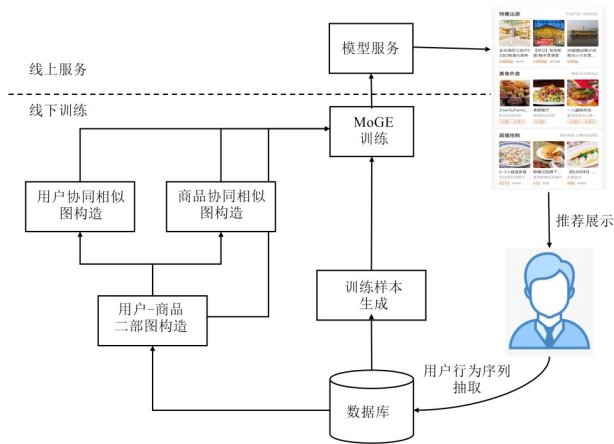


图6 MoGE 训练/部署流程

与已经部署的基线模型相比,本文提出的 MoGE 模型在核心指标 GAUC 上会有 1%~5% 的提升.这一表现进一步验证了本文 MoGE 模型的有效性.

7 结论与展望

本文提出了一种新的图增强的多门控专家混合推荐框架即 MoGE,精心开发了一个 MGNN 模块来联合建模高阶连通性和协同相似性.在学习到的结构信息基础上,引入一组图增强的专家网络,用于端到端训练过程中的多任务推荐.大量的实验结果证明了该模型的优越性.在未来的工作中,我们将研究将更复杂的图结构(如知识图谱和异构图)融入本文的框架.

附录 任务依赖的不确定性损失函数

为了简化表达式,令 $\hat{x}_{u,i}$ 为模型的输入,同时令 $f^\theta(\hat{x}_{u,i}) = \hat{y}_{u,i}$ 为模型的分数输出,其中 θ 为参数.接下来,模型在每个任务上的分类可能性可以表示为

$$p(y_{u,i} = c | f^\theta(\hat{x}_{u,i}), \sigma) = \frac{\exp\left(\frac{1}{\sigma^2} f_c^\theta(\hat{x}_{u,i})\right)}{\sum_{c'} \exp\left(\frac{1}{\sigma^2} f_{c'}^\theta(\hat{x}_{u,i})\right)}$$

其中, $y_{u,i}$ 为样本 $\hat{x}_{u,i}$ 的真实标签,正标量 σ^2 也称为温度变量,它可以调节输出的熵.如公式所示,如果 $\sigma^2 \rightarrow \infty$,输出概率趋于一致,表示最大的不确定性.反之如果 $\sigma^2 \rightarrow 0$,概率退化到质点,也就是说不确定性最小.这个输出函数的对数似然形式可以写成:

$$\begin{aligned} \log p(y_{u,i} = c | f^\theta(\hat{x}_{u,i}), \sigma) &= \frac{1}{\sigma^2} f_c^\theta(\hat{x}_{u,i}) - \log \sum_{c'} \exp\left(\frac{1}{\sigma^2} f_{c'}^\theta(\hat{x}_{u,i})\right). \end{aligned}$$

然后,假设模型的多个输出是离散变量,可以建模成 Softmax 的最大似然形式.因此,可以得到不确定性目标函数:

$$\begin{aligned} \mathcal{L}_{\text{main}}(\theta, \Sigma) &= -\log p(y_{u,i}^1 = c_1, \dots, y_{u,i}^K = c_K | f^\theta(\hat{x}_{u,i}), \sigma^1, \dots, \sigma^K) \\ &= -\log \left(\prod_{k=1}^K p(y_{u,i}^k = c_k | f^\theta(\hat{x}_{u,i}), \sigma^k) \right) \\ &= -\sum_{k=1}^K \left(\frac{1}{\sigma^{k2}} f_{c_k}^\theta(\hat{x}_{u,i}) - \log \sum_{c'_k} \exp\left(\frac{1}{\sigma^{k2}} f_{c'_k}^\theta(\hat{x}_{u,i})\right) \right) \\ &= -\sum_{k=1}^K \left(\frac{1}{\sigma^{k2}} f_{c_k}^\theta(\hat{x}_{u,i}) + \frac{1}{\sigma^{k2}} \log \sum_{c'_k} \exp\left(f_{c'_k}^\theta(\hat{x}_{u,i})\right) \right) \\ &\quad - \log \sum_{c'_k} \exp\left(\frac{1}{\sigma^{k2}} f_{c'_k}^\theta(\hat{x}_{u,i})\right) - \frac{1}{\sigma^{k2}} \log \sum_{c'_k} \exp\left(f_{c'_k}^\theta(\hat{x}_{u,i})\right) \end{aligned}$$

$$\begin{aligned}
&= -\sum_{k=1}^K \left(\frac{1}{\sigma^{k2}} \log p \left(y_{u,i}^k = c_k f^{\theta}(\hat{x}_{u,i}) \right) \right) \\
&\quad + \sum_{k=1}^K \log \frac{\sum_{c'_k} \exp \left(\frac{1}{\sigma^{k2}} f_{c'_k}^{\theta}(\hat{x}_{u,i}) \right)}{\left(\sum_{c'_k} \exp \left(f_{c'_k}^{\theta}(\hat{x}_{u,i}) \right) \right)^{\frac{1}{\sigma^{k2}}}} \\
&\approx \sum_{k=1}^K \left(\frac{1}{\sigma^{k2}} \sum_{\langle u,i \rangle \in \mathcal{H}} \mathcal{C}(\hat{y}_{u,i}^k, y_{u,i}^k) + \log \sigma^{k2} \right) \\
&= \sum_{k=1}^K \sum_{\langle u,i \rangle \in \mathcal{H}} \left(\frac{1}{\sigma^{k2}} \mathcal{C}(\hat{y}_{u,i}^k, y_{u,i}^k) + \log \sigma^{k2} \right).
\end{aligned}$$

其中, $\mathcal{C}(\cdot, \cdot)$ 为交叉熵损失. 在倒数第二个变换中, 采用和文献 [33] 相同的假设, 即

$$\frac{1}{\sigma^{k2}} \sum_{c'_k} \exp \left(\frac{1}{\sigma^{k2}} f_{c'_k}^{\theta}(\hat{x}_{u,i}) \right) \approx \left(\sum_{c'_k} \exp \left(f_{c'_k}^{\theta}(\hat{x}_{u,i}) \right) \right)^{\frac{1}{\sigma^{k2}}}.$$

参考文献

- [1] ZHANG Y, YANG Q. A survey on multi-task learning[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(12): 5586-5609.
- [2] MISRA I, SHRIVASTAVA A, GUPTA A, et al. Cross-stitch networks for multi-task learning[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 3994-4003.
- [3] MA X, ZHAO L Q, HUANG G, et al. Entire space multi-task model: An effective approach for estimating post-click conversion rate[C]//SIGIR' 18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. New York: ACM, 2018: 1137-1140.
- [4] TANG H Y, LIU J N, ZHAO M, et al. Progressive layered extraction (PLE): A novel multi-task learning (MTL) model for personalized recommendations[C]//Proceedings of the 14th ACM Conference on Recommender Systems. New York: ACM, 2020: 269-278.
- [5] 郭景峰, 朱晓松, 李爽. 直播电视推荐系统的评分预测算法研究[J]. 电子学报, 2020, 48(9): 1735-1740.
GUO J F, ZHU X S, LI S. Research on rating prediction algorithm of live TV recommender systems[J]. Acta Electronica Sinica, 2020, 48(9): 1735-1740. (in Chinese)
- [6] 李琳, 唐守廉. 基于多层注意力表示的音乐推荐模型[J]. 电子学报, 2020, 48(9): 1672-1679.
LI L, TANG S L. Hierarchical attention representation model for music recommendation[J]. Acta Electronica Sinica, 2020, 48(9): 1672-1679. (in Chinese)
- [7] 方晨, 张恒巍, 王娜, 等. 基于随机游走和多样性图排序的个性化服务推荐方法[J]. 电子学报, 2018, 46(11): 2773-2780.
- [8] FANG C, ZHANG H W, WANG N, et al. Personalized service recommendation method based on random walk and diversified graph ranking[J]. Acta Electronica Sinica, 2018, 46(11): 2773-2780. (in Chinese)
- [9] CHEN L, WU L, HONG R C, et al. Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(1): 27-34.
- [10] WU L, SUN P J, FU Y J, et al. A neural influence diffusion model for social recommendation[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 235-244.
- [11] WEN H, ZHANG J, WANG Y, et al. Entire space multi-task modeling via post-click behavior decomposition for conversion rate prediction[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 2377-2386.
- [12] GAO C, HE X N, GAN D H, et al. Neural multi-task recommendation from multi-behavior data[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). Piscataway: IEEE, 2019: 1554-1557.
- [13] LI D C, LI X, WANG J, et al. Video recommendation with multi-gate mixture of experts soft actor critic[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 1553-1556.
- [14] KIM S, AHN N, SOHN K A. Restoring spatially-heterogeneous distortions using mixture of experts network[C]//Asian Conference on Computer Vision. Cham: Springer, 2021: 185-201.
- [15] WANG X, YU F, DUNLAP L, et al. Deep mixture of experts via shallow embedding[EB/OL]. (2018) [2022]. <https://arxiv.org/abs/1806.01531.pdf>.
- [16] MA J Q, ZHAO Z, YI X Y, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 1930-1939.
- [17] MA J Q, ZHAO Z, CHEN J L, et al. SNR: Sub-network routing for flexible parameter sharing in multi-task learning[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 216-223.

- [17] HUAN Z X, ZHANG G D, ZHANG X L, et al. An industrial framework for cold-start recommendation in zero-shot scenarios[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 3403-3407.
- [18] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4-24.
- [19] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016)[2022]. <https://arxiv.org/abs/1609.02907.pdf>.
- [20] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. (2017)[2022]. <https://arxiv.org/abs/1710.10903.pdf>.
- [21] WANG X, HE X N, WANG M, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 165-174.
- [22] YING R, HE R N, CHEN K F, et al. Graph convolutional neural networks for web-scale recommender systems[C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2018: 974-983.
- [23] FAN W Q, MA Y, LI Q, et al. Graph neural networks for social recommendation[C]//The World Wide Web Conference. New York: ACM, 2019: 417-426.
- [24] WANG X, HE X N, CAO Y X, et al. KGAT: Knowledge graph attention network for recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2019: 950-958.
- [25] VANDENHENDE S, GEORGOULIS S, VAN GANSBEKE W, et al. Multi-task learning for dense prediction tasks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(7): 3614-3633.
- [26] 俞春花, 刘学军, 李斌, 等. 基于上下文相似度和社交网络的移动服务推荐方法[J]. 电子学报, 2017, 45(6): 1530-1536.
- YU C H, LIU X J, LI B, et al. Mobile service recommendation based on context similarity and social network[J]. Acta Electronica Sinica, 2017, 45(6): 1530-1536. (in Chinese)
- [27] 陈杰, 刘学军, 李斌, 等. 一种基于用户动态兴趣和社交网络的微博推荐方法[J]. 电子学报, 2017, 45(4): 898-905.
- CHEN J, LIU X J, LI B, et al. Personalized microblogging recommendation based on dynamic interests and social networking of users[J]. Acta Electronica Sinica, 2017, 45(4): 898-905. (in Chinese)
- [28] 肖云鹏, 刘瀚松, 刘宴兵. 一种基于二部图和节点角色划分的社交网络推荐方案[J]. 电子学报, 2017, 45(10): 2425-2433.
- XIAO Y P, LIU H S, LIU Y B. A social network recommendation scheme based on bipartite graph and node role division[J]. Acta Electronica Sinica, 2017, 45(10): 2425-2433. (in Chinese)
- [29] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.
- [30] XU K, LI C T, TIAN Y L, et al. Representation learning on graphs with jumping knowledge networks[EB/OL]. (2018)[2022]. <https://arxiv.org/abs/1806.03536.pdf>.
- [31] QIN Z, CHENG Y C, ZHAO Z, et al. Multitask mixture of sequential experts for user activity streams[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 3083-3091.
- [32] KENDALL A, GAL Y. What uncertainties do we need in Bayesian deep learning for computer vision? [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 5580-5590.
- [33] CIPOLLA R, GAL Y, KENDALL A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7482-7491.
- [34] WANG M H, LIN Y J, LIN G L, et al. M2GRL: A multi-task multi-view graph representation learning framework for web-scale recommender systems[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 2349-2358.
- [35] HSU W T, LIN C K, LEE M Y, et al. A unified model for extractive and abstractive summarization using inconsistency loss[EB/OL]. (2018)[2022]. <https://arxiv.org/abs/1805.06266.pdf>.
- [36] NIU X C, LI B F, LI C L, et al. Heterogeneous graph aug-

mented multi-scenario sharing recommendation with tree-guided expert networks[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 1038-1046.

- [37] LIU Z Q, CHEN C C, LI L F, et al. GeniePath: Graph neural networks with adaptive receptive paths[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 4424-4431.
- [38] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 1025-1035.
- [39] CARUANA R. Multitask learning[J]. Machine Learning, 1997, 28(1): 41-75.
- [40] LI S H, YANG D K, ZHANG B F. MRIF: Multi-resolution interest fusion for recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 1765-1768.
- [41] TIAN Y, CHANG J X, NIU Y N, et al. When multi-level meets multi-interest: A multi-grained neural model for sequential recommendation[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 1632-1641.
- [42] LI P C, LI R Z, DA Q, et al. Improving multi-scenario learning to rank in E-commerce by exploiting task relationships in the label space[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM, 2020: 2605-2612.
- [43] ZHOU J, LI X L, ZHAO P L, et al. KunPeng: Parameter server based distributed learning systems and its applications in alibaba and ant financial[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2017: 1693-1702.
- [44] ZHANG D L, HUANG X, LIU Z Q, et al. AGL: A scalable system for industrial-purpose graph machine learning [J]. Proceedings of the VLDB Endowment, 2020, 13(12): 3125-3137.

作者简介



周 俊 男,1986年出生,湖南人.浙江大学博士.蚂蚁科技集团股份有限公司机器智能部负责人.主要研究方向为机器学习、神经网络等.中国电子学会会员编号:E190028162S.
E-mail: jun.zhoujun@antfin.com



陈超超(通讯作者) 男,1988年出生,河南人.浙江大学特聘研究员.主要研究方向为图神经网络、推荐系统等.中国电子学会会员编号:E190025053M.
E-mail: zjuccc@zju.edu.cn