doi:10.3969/j.issn.1673-9833.2025.06.003

融合加权辅助任务感知的多任务推荐算法

黄海洋，黄贤明，翁成康

（湖南工业大学 计算机与人工智能学院，湖南 株洲 412007）

摘　要：在线推荐系统在个性化学习路径设计中起着关键作用，但传统的推荐算法往往采用固定的权重设置，无法灵活适应用户的兴趣变化。为解决这一问题，首先，提出了一种融合加权辅助任务感知的多任务推荐算法 WAA-TA，动态调整任务权重，并利用任务集合代表用户在不同生命周期阶段的学习需求，以提高推荐系统的准确性和个性化程度。其次，在 Edx 和 MOOCCubeX 两个数据集上与 6 个基线算法进行对比实验，实验结果表明，本算法在各项评估指标上表现更优越，尤其在提高用户满意度和推荐准确性方面有显著效果。

关键词：推荐系统；多任务推荐；动态权重；特征分解

中图分类号：TP312　　**文献标志码**：A　　**文章编号**：1673-9833(2025)06-0015-08

引文格式：黄海洋，黄贤明，翁成康 . 融合加权辅助任务感知的多任务推荐算法 [J]. 湖南工业大学学报，2025，39(6)：15-22.

Multi-Task Recommendation Algorithm with Weighted Auxiliary Task Perception Integrated

HUANG Haiyang，HUANG Xianming，WENG Chengkang

（School of Computer Science and Artificial Intelligence，Hunan University of Technology，Zhuzhou Hunan 412007，China）

Abstract：Despite the fact that online course recommendation system plays a key role in personalized learning path design, traditional recommendation algorithms, which often adopt fixed weight settings, fail to flexibly adapt to the changes of users' interests. In order to address this issue, firstly, a multi-task recommendation algorithm WAA-TA has been proposed with weighted auxiliary task perception integrated. The algorithm dynamically adjusts task weights and uses task sets to represent users' learning needs at different stages of their lifecycle, thus improving the accuracy and personalization of the recommendation system. Secondly, based on comparative experiments conducted with six baseline algorithms on the Edx and MOOCCubeX datasets, the experimental results show that this algorithm is characterized by an excellent performance in various evaluation indicators, especially in improving user satisfaction and recommendation accuracy.

Keywords：recommendation system；multi-task recommendation；dynamic weight；feature decomposition

0　引言

随着数字化时代的快速发展，个性化推荐系统已成为连接用户与海量信息资源的关键纽带。然而，信息爆炸式增长导致用户面临资源过载、信息迷航等问题，加上内容与需求匹配度低及广告侵扰等挑战 [1]，

收稿日期：2024-09-15

基金项目：教育部中国高校产学研创新基金资助项目（2020ITA05043，2023DT002）；湖南省教育厅科学研究基金资助项目（21C0409）

作者简介：黄海洋，男，湖南工业大学硕士生，主要研究方向为推荐系统，深度学习，E-mail：haiyangh@163.com

通信作者：黄贤明，男，湖南工业大学副教授，主要研究方向为推荐系统，深度学习，E-mail：841619296@qq.com

推荐系统的准确性与个性化水平亟待提升。而传统推荐系统多采用固定权重策略，难以适应用户兴趣的动态变化，导致推荐效果受限。如基于内容的协同过滤方法虽然考虑了用户历史行为，但未能充分应对用户兴趣的动态演变[2]。余学兵等[3]通过知识图谱挖掘用户语义信息，提升了推荐的语义辅助能力，但仍受限于静态权重的调整，未真正考虑用户兴趣随时间的动态变化。有些学者尝试通过深度学习等方法进行特征融合，如刘震等[4]结合序列相似度匹配（sequence similarity matching，SSM）算法和 Doc2Vec 模型，优化了序列相似度匹配过程，但未充分考虑权重的动态适应性。还有学者通过在模型中引入多个任务，来捕捉用户在不同情境下的兴趣变化。但现有的多任务学习方法大多对任务之间的关系缺乏细分和深入研究。如方宝等[5]利用多模态信息特征融合的隐语义模型（FLFM）融合多模态信息、多任务学习预测推荐评分。T. Standley 等[6]明确讨论了多任务学习中任务关系的研究挑战，并提出了一个框架来解决任务之间的合作与竞争问题。然而，该论文在理论深度和泛化能力方面仍有不足，未能完全解决 MTL 中任务关系研究的根本问题。

因此，本研究拟针对静态权重适应性不足、动态特征融合不充分、多任务关系研究薄弱等问题，提出一种融合加权辅助任务感知的多任务推荐算法（multi-task recommendation algorithm for weighted auxiliary task awareness，WAA-TA）。首先，利用多任务学习中任务的特征分解，把具体任务分离成具体的指标，增强推荐的准确性，从而减少负迁移带来的影响。其次，在辅助任务阶段，引入动态平衡权重，针对共享参数不同的目标自适应平衡，相比任务损失固定的权重更加灵活。最后，在两个数据集上进行实验验证，实验结果表明，WAA-TA 算法较其它推荐算法的效果显著提升。本研究成果不仅丰富了推荐系统的理论体系，也为实际应用提供了有效的技术支持，具有广泛的应用前景和社会价值。

1 学习算法介绍

1.1 多任务学习算法

WAA-TA 算法基于多任务学习框架，旨在通过动态平衡辅助任务的梯度幅度。在任务方面创建共享表征，通过多层提取的方式细化表征，以提高目标任务的测试准确性，其算法整体架构见图 1。

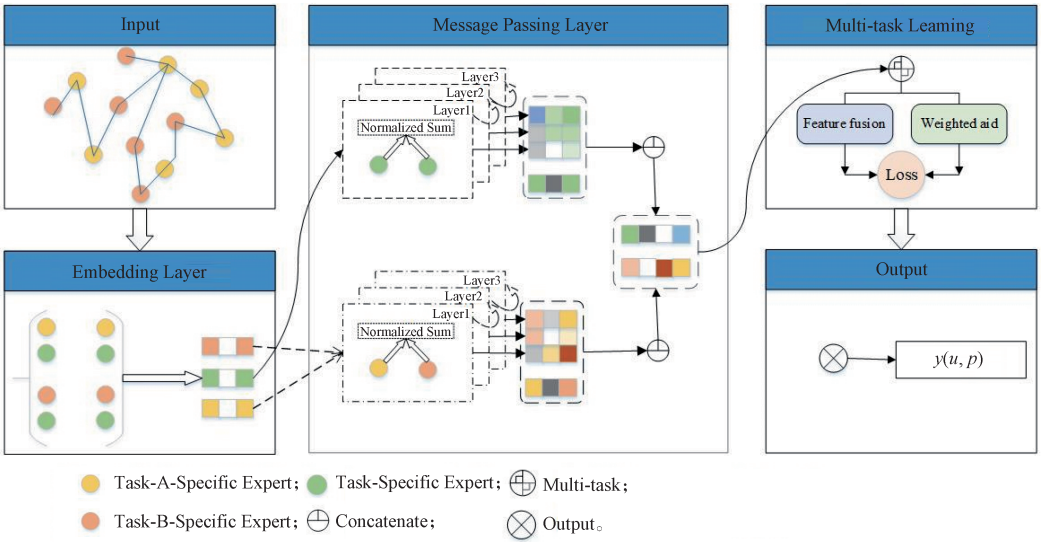


图 1 WAA-TA 的整体架构

Fig. 1 Overall architecture of WAA-TA

该算法的伪代码如下：

WAA-TA 框架

1：给定一个预测任务集 $D=\{user_n, item_n, (y_n, 1, y_n, 2, \cdots, y_n, k)\}_{n=1}^{N}$

2：初始化实体和关系向量

3：**for** 每个三元组 (h, r, t) in 预测任务集 **do**

4：　将头部实体 h 和尾部实体 t 映射到关系空间

5：　计算投影向量 $\boldsymbol{v}_{h_\perp}=\boldsymbol{M}_{rh}\boldsymbol{v}_h$，$\boldsymbol{v}_{t_\perp}=\boldsymbol{M}_{rt}\boldsymbol{v}_t$

6：**end for**

7：**for** 每个与预测任务集相关实体 e **do**

8：　计算关系注意系数 $\hat{\pi}_r(e,\ t)=(\boldsymbol{v}_{t_\perp})\tanh(\boldsymbol{v}_{h_\perp}+\boldsymbol{v}_r)$

9：　归一化关系注意系数 $\pi_r(e,\ t)=\exp\left(\hat{\pi}_r(e,\ t)\right)/\sum_{(r',\ t')\in N(e)}\exp\left(\hat{\pi}_{r'}(e,\ t')\right)$

10：　聚合邻域信息 $f_{\text{Aggregator}}=Leaky\ ReLu\left(\boldsymbol{W}_1\left(\boldsymbol{v}_e+\boldsymbol{v}_{N(e)}\right)\right)+Leaky\ ReLu\left(\boldsymbol{W}_2\left(\boldsymbol{v}_e\odot\boldsymbol{v}_{N(e)}\right)\right)$

11：　更新实体表示 $\boldsymbol{v}_e^{(l)}=f\left(\boldsymbol{v}_e^{(l-1)},\ \boldsymbol{v}_{N(e)}^{(l-1)}\right)$

12：**end for**

13：计算用户实体表示向量 $\boldsymbol{v}_u^*$ 与每个候选配方实体表示向量 $\boldsymbol{v}_{rp}^*$ 的内积，以生成匹配分数 $y(u,\ p)=\boldsymbol{v}_u^*\cdot\boldsymbol{v}_{rp}^*$

1.1.1　输入层

本实验给定一个由 k 个任务组成的预测任务集 $D=\left\{user_n,\ item_n,\ (y_n,\ 1,\ y_n,\ 2,\ \cdots,\ y_n,\ k)\right\}_{n=1}^{N}$，由 n 条交互记录组成，其中 $user_n$ 和 $item_n$ 表示第 n 个用户项。每条数据记录有 k 个 0~1 的二进制标签 y_1、y_2、…、y_k。

1.1.2　嵌入层

嵌入层是将实体和边转换成向量的过程。本算法基于预测任务集的数据，识别并提取出所有的实体。首先初始化实体和关系的向量，然后利用知识图嵌入对向量表示进行更新。本文采用 TransD[7] 作为知识图嵌入方法，将 CRKG（contextual relationship knowledge graph）中的实体和边嵌到连续向量空间中，同时保留其结构信息。给定一个三元组（h，r，t），他们的向量包含 $\boldsymbol{v}_h$、$\boldsymbol{v}_{hp}$、$\boldsymbol{v}_r$、$\boldsymbol{v}_{rp}$、$\boldsymbol{v}_t$、$\boldsymbol{v}_{tp}$，下标有 p 的表示投影向量。其中 $\boldsymbol{v}_h$、$\boldsymbol{v}_{hp}$、$\boldsymbol{v}_t$、$\boldsymbol{v}_{tp}\in\mathbf{R}^d$，而 $\boldsymbol{v}_r$、$\boldsymbol{v}_{rp}\in\mathbf{R}^z$。TransD 将头部实体 h 和尾部实体 t 映射到由实体和三元组中的关系构成的公共空间中，使用两个映射矩阵，并通过构造投影矩阵将头部实体 h 和尾部实体 t 映射到关系空间，如公式（1）所示。而 $\boldsymbol{M}_{rh}$ 和 $\boldsymbol{M}_{rt}$ 由图 2 所得。

$$\boldsymbol{v}_{h_\perp}=\boldsymbol{M}_{rh}\boldsymbol{v}_h,\ \boldsymbol{v}_{t_\perp}=\boldsymbol{M}_{rt}\boldsymbol{v}_t\text{。}\tag{1}$$

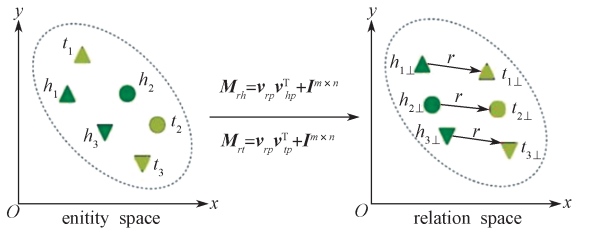


图 2　TransD 图解

Fig. 2　TransD diagram

在式（1）中，$\boldsymbol{v}_{h_\perp}$ 和 $\boldsymbol{v}_{t_\perp}$ 分别为关系空间中的头部实体 h 和尾部实体 t 的表示。三元组的似然性得分 $g(h,\ r,\ t)$ 由式（2）得出。其中较高的分数指示嵌入更合理，而较低的分数指示嵌入更不合理。

$$g(h,\ r,\ t)=-\boldsymbol{v}_{h\perp}+\boldsymbol{v}_r-\boldsymbol{v}_{t\perp}{}_2^2\text{。}\tag{2}$$

1.1.3　消息传递层

该层构建了一个知识感知的注意力卷积神经网络图，该网络通过消息传递，更新实体表示在消息传递层[8]。CRKG 中的每个实体将该网络通过广播传递给与其直接相连的一阶邻域实体，并从其邻域实体接收信息[9]。实体的一阶邻域是指与其直接相连的实体，高阶邻域是指与其间接相连的实体。在一层消息传递中，CRKG 中的所有实体将接收到的邻域信息集成以生成新的表示。通过叠加多层这样的操作，每个实体都可以从高阶邻域接收信息，从而捕获实体之间的高阶相似性。

消息传递的第一阶段是接收来自相邻节点的消息。CRKG 上的实体 e 传播邻域信息的计算方法定义为式（3）。

$$\boldsymbol{v}_{N(e)}=\sum_{(r,\ t)\in N(e)}\pi_r(e,\ t)\boldsymbol{v}_t\text{，}\tag{3}$$

式中：$\boldsymbol{v}_{N(e)}$ 为实体 e 的邻域信息；$N(e)$ 为实体 e 的所有邻域实体集合；$\boldsymbol{v}_t$ 为相邻实体 t 的表示；$\pi_r(e,\ t)$ 为实体 e 和实体 t 在关系下的归一化关系关注系数，为实体 e 的所有相邻实体的表示乘以归一化的关系注意系数与求和得到实体 e 的邻域信息。

关系注意系数的定义如式（4）所示，其反映了实体接收到相邻实体的信息时，相邻实体的重要程度。

$$\hat{\pi}_r(e,\ t)=\left(\boldsymbol{v}_{t_\perp}\right)^{\mathrm{T}}\tanh\left(\boldsymbol{v}_{e_\perp}+\boldsymbol{v}_r\right)\text{。}\tag{4}$$

使用 softmax 函数对与实体 e 连接的所有相邻实体的关系注意系数进行规范化，如式（5）所示。

$$\pi_r(e,\ t)=\exp\left(\hat{\pi}_r(e,\ t)\right)\div\sum_{(r',\ t')\in N(e)}\exp\left(\hat{\pi}_{r'}(e,\ t')\right)\text{。}\tag{5}$$

在获得实体 e 的邻域的所有信息后，对信息进行聚合。将实体 e 的邻域信息与实体 e 自身的表示进行聚合，更新实体 e 的表示。本文采用双向交互的方式对邻域信息进行聚合，可以全面捕捉和处理实体之间的交互信息。具体聚合公式如式（6）所示。

$$f_{\text{Aggregator}}=ReLu\left(\boldsymbol{W}_1\left(\boldsymbol{v}_e+\boldsymbol{v}_{N(e)}\right)\right)+ReLu\left(\boldsymbol{W}_2\left(\boldsymbol{v}_e\odot\boldsymbol{v}_{N(e)}\right)\right),\tag{6}$$

式中：$\boldsymbol{W}_1$、$\boldsymbol{W}_2\in\mathbf{R}^d$ 为可训练的参数矩阵；$\odot$ 算子表示向量之间的元素乘法；*ReLU* 为激活函数。经过一层消息传递后，更新实体表示，抽象为式（7）。

$$\boldsymbol{v}_e^{(l)}=f\left(\boldsymbol{v}_e^{(l-1)},\ \boldsymbol{v}_{N(e)}^{(l-1)}\right)。\tag{7}$$

算法通过计算用户实体表示向量 $\boldsymbol{v}_u^*$ 和配方实体表示向量 $\boldsymbol{v}_{rp}^*$ 之间的内积为输出预测的匹配分数。内积越大，算法预测的用户实体表示向量与配方实体表示向量的匹配程度越高。

$$y(u,p)=\boldsymbol{v}_u^*\cdot\boldsymbol{v}_{rp}^*。\tag{8}$$

1.1.4 多任务学习层

假设 $f^{\text{s}}(X)$ 为所有任务的共享信息表征，B 为底层信息表征，l 为多层的数量，底层的多层感知器结构由权重矩阵 $\boldsymbol{W}_b$ 和激活函数 f 组成，通过底层累计可以得到：

$$f^{\text{s}}(X)=\sum_{i=1}^{l}B_i\ ,\tag{9}$$

$$B_i=f(\boldsymbol{W}_b\times i)。\tag{10}$$

$f_{\text{t}}^k(X)$ 为第 k 任务独有的信息表征，

$$f_{\text{t}}^k(X)=\sum_{i=1}^{l}B_i^k\ 。\tag{11}$$

为了融合共享信息和任务独有信息，引入一个 Gate 权重 W_{g}^k，用于动态平衡两者的重要性。

$$G_k(X)=\sigma\left(W_{\text{g}}^k\times\left[f^{\text{s}}(X);\ f_{\text{t}}^k(X)\right]+\boldsymbol{b}_{\text{g}}^k\right),\tag{12}$$

式中：σ 为 Sigmoid 激活函数；$\left[f^{\text{s}}(X);f_{\text{t}}^k(X)\right]$ 将共享信息和任务独有信息在特征维度上拼接；$\boldsymbol{b}_{\text{g}}^k$ 为算法中的任务 k 的门控偏置向量。最终，第 k 个任务的输入 X^k 可以通过 Gate 权重对共享和任务独有信息进行加权组合：

$$X^k=G_k(X)\times f^{\text{s}}(X)+\left(1-G_k(X)\right)\times f_{\text{t}}^k(X)。\tag{13}$$

1.2 算法约束

1.2.1 特征表征正交约束

正交约束 L_{o} 的目的是使每个 task 的共享和独有两种特征表征尽可能差异化，从而提高算法对各个 task 之间共性和差异性的捕捉能力。

$$L_{\text{o}}=\sum_{k=1}^{K}\sum_{m=1}^{M}\left\|\left(f_m^{\text{s}}\right)^{\text{T}}f_m^p\right\|_{\text{F}}^2\ ,\tag{14}$$

式中：K 为 task 的数量；M 为 task 中 k 的特征分解对数量；f_m^{s} 为 task k 的第 m 个 task 共享的特征表征；f_m^p 为 task k 的第 m 个 task 独有的特征表征；$\|\cdot\|_{\text{F}}^2$ 为 Frobenius 范数。

另一种约束为辅助 loss 约束。其中 task k 的辅助 loss 函数为 $L_{k,m}$；y_k 为 task k 的真实标签，而 $\sigma\left(f_m^{k,p}(X)\right)$ 为 task k 中第 m 个独有特征表征的 task 预估，辅助 loss 的引入，可以使正交的子空间方向和 task 的目标保持一致。辅助任务 L_{aux} 表达式为

$$L_{\text{aux}}=\sum_{k=1}^{K}\sum_{m=1}^{M}L_{k,m}\left(\sigma\left(f_m^{k,p}(X)\right),y_k\right)。\tag{15}$$

1.2.2 辅助任务约束

辅助任务的目标是通过在多任务网络上训练辅助任务和目标任务来提高目标任务的测试准确率。设 θ 是共享参数的子集，共享底部网络中多层感知器的权重矩阵，通过参与共同最小化目标任务损失 L_{tar}、辅助任务损失 L_{aux} 和特征表正交约束 L_{o} 来学习 L_{total}。

$$L_{\text{total}}=L_{\text{tar}}+L_{\text{aux}}+L_{\text{o}}。\tag{16}$$

通过学习率 a 的梯度下降来更新 θ^t。

$$\theta^{t+1}=\theta^t-a\times G_{\text{total}}^t。\tag{17}$$

式中：t 为小批次的第 t 次训练迭代 (t=1, 2, ⋯)，G_{total}^t 是 L_{total}^t 的梯度，其中 G_{total} 为目标和辅助损失的每个梯度之和。

$$G_{\text{total}}^t=\nabla_\theta L_{\text{total}}^t=L_{\text{tar}}^t+\nabla_\theta L_{\text{aux}}^t+L_{\text{o}}。\tag{18}$$

2 实验设计

2.1 数据集

为了保证算法评估的公平性和有效性，本实验采用 MOOCCubeX 和 Edx 数据集进行实验。MOOCCubeX 数据集是一个大规模的在线教育数据集，它收集了来自多个 MOOC 平台的学生学习行为数据，包含超过 2.96 亿条关于 3 330 294 名学生的原始行为数据。Edx 数据集来源于 Kaggle 平台，是 Kaggle 上关于 2012—2016 年 Harvard 与 MIT 在 Edx 平台开设的在线课程的详尽数据报告，包含了 290 门在线课程，吸引了超过 450 万名学员参与学习。数据集的基本信息见表 1。

表 1 数据集统计信息

Table 1 Dataset statistics

数据集	users	courses	iteraction	sparsity/%
MOOCCubeX	3 330 294	4 216	23 615 324	99.95
Edx	4 500 000	756	7 588 365	99.89

本文随机选取每个数据集中 60% 的交互作为训练数据集，20% 作为验证数据集，另外 20% 作为测试数据集。

2.2 评估指标

本研究采用基于 Top-k 推荐的评测方法。在评估

算法性能时，引入了召回率（Recall@k）和归一化折损累计增益（NDCG@k）作为主要评测指标。召回率计算式为

$$\text{Recall}@k=\sum_{u\in U}\left|R(u)\cap T(u)\right|\Big/\sum_{u\in U}\left|T(u)\right|,\quad(19)$$

式中：$R(u)$ 为根据用户在训练集上的行为得到的 Top-k 推荐列表；$T(u)$ 为系统向用户推荐后，用户实际选择的集合。

归一化折损累计增益（NDCG@k）则考虑了用户对推荐物品的相关性分数和排列顺序，其计算式为

$$\text{NDCG}@k=\sum_{i}^{k}\frac{rel(i)}{lb(i+1)}\div\sum_{i}^{|REL|}\frac{rel(i)}{lb(i+1)}。\quad(20)$$

式中：NDCG@k 的取值范围为 [0,1]，越大效果越好；$rel(i)$ 为 $item(i)$ 的相关性得分；$|REL|$ 表示结果按真实相关性从大到小排序，返回由前 k 个结果组成的推荐列表。

2.3　对比算法

为了评估所提算法的性能，本实验选取以下对比算法：

1）MMoE（multi-gate mixture-of-experts）[10]，其首次提出拆分信息表征的方法，实现了 task 输入的差异化。

2）PLE（progressive layered extraction）[11]，对信息表征做了更细致和更具区分性的拆分。

3）FDN（Feature Decomposition Network）[12]，其在 CGC（customized gate control）思路的基础上，对信息表征的拆分进一步细化和差异化，并对信息表征的学习引入了更强的约束，以提高共享和独有的信息表征在空间上的正交性。

4）ESMM（entire space multi-task model）[13]，其利用转化行为对点击行为的依赖性，对点击率（click through rate，CTR）和转化率（conversion rate，CVR）任务进行级联，隐式地将 CVR 任务的建模空间变成曝光空间，解决了 CVR 任务在训练和推理空间不一致的问题。

5）ESM2（entire space counterfactual multi-task model for post-click conversion rate estimation）[14]：ESMM 的升级版，其根据行为的依赖性建模，对行为依赖路径进行了更细化的拆分。

6）GS-RS（a generative approach for alleviating cold start and filter bubbles in recommender systems）[15]，其提出生成自约束推荐系统，利用孪生条件生成对抗网络和虚拟邻居偏好，解决冷启动和过滤气泡问题。

2.4　实验设置

本实验在 Intel(R) Core(R) i5 12490FCPU @ 3.60 GHz，Ubuntu 22.04，512 GB，RTX3090 的环境下进行。本实验使用 Adam 优化器进行优化，具体配置详见表 2。

表 2　优化器的具体配置

Table 2　Configuration details

名称	数值	名称	数值
学习率	0.000 2	训练批次大小	1 024
专家神经元个数	32	训练轮次	20
任务数	2	迭代次数	400

3　实验结果和分析

3.1　基线对比

在对比了多种基线算法后，WAA-TA 在 MOOCCubeX 和 Edx 数据集上的 Recall 和 NDCG 性能指标均优于其它算法，包括 MMoE、PLE、FDN、ESMM、ESM2、GS-RS 等，其结果如表 3 所示。

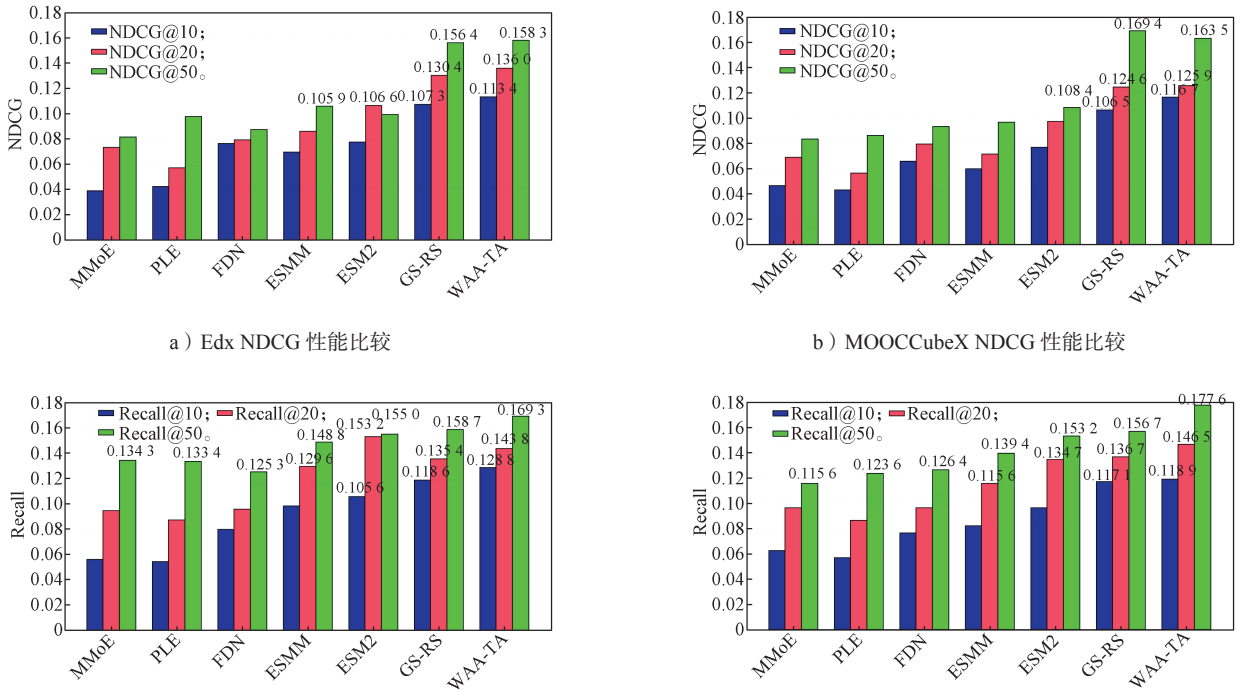
表 3　算法性能比较

Table 3　Model performance comparison

dataset	Metric	MMoE	PLE	FDN	ESMM	ESM2	GS-RS	WAA-TA	Improve/%
MOOCCubeX	Recall@10	0.062 3	0.056 7	0.076 5	0.082 1	0.096 4	0.117 1	0.118 9	+1.53
	NDCG@10	0.046 5	0.043 2	0.065 9	0.059 8	0.076 8	0.106 5	0.116 7	+9.57
	Recall@20	0.096 4	0.086 4	0.096 4	0.115 6	0.134 7	0.136 7	0.146 5	+6.51
	NDCG@20	0.068 9	0.056 4	0.079 5	0.071 6	0.097 5	0.124 6	0.125 9	+1.04
	Recall@50	0.115 6	0.123 6	0.126 4	0.139 4	0.153 2	0.156 7	0.177 6	+13.30
	NDCG@50	0.083 6	0.086 4	0.093 5	0.096 7	0.108 4	0.169 4	0.163 5	−3.48
Edx	Recall@10	0.056 0	0.054 3	0.079 9	0.098 2	0.105 6	0.118 6	0.128 8	+8.60
	NDCG@10	0.038 9	0.042 2	0.076 5	0.069 6	0.077 6	0.107 3	0.113 4	+5.68
	Recall@20	0.094 6	0.087 2	0.095 6	0.129 6	0.153 2	0.135 4	0.143 8	−6.13
	NDCG@20	0.073 1	0.057 1	0.079 1	0.086 1	0.106 6	0.130 4	0.136 0	+4.29
	Recall@50	0.134 3	0.133 4	0.125 3	0.148 8	0.155 2	0.158 7	0.169 3	+6.67
	NDCG@50	0.081 6	0.097 7	0.087 5	0.105 9	0.099 4	0.156 4	0.158 3	+1.21

由表 3 可知，MOOCCubeX 平均提升了 4.75%，其中提升最高的为 Recall@50 和 NDCG@10，提升效果分别为 13.30% 和 9.57%。但在 NDCG@50 上，GS-RS 表现非常突出，达到了 0.169 4，而 WAA-TA 为 0.163 5，其性能相比下降了 3.48%。此结果表明 WAA-TA 能够更全面地涵盖潜在物品，但以 Top50 的排序准确性为代价。而 Edx 数据集的平均提升为 3.39%，从实验结果上来看，在 Recall@10 和 Recall@50 上，WAA-TA 表现较好，分别提升了 8.60% 和 6.67%。在排序相关结果方面，ESM2 和 WAA-TA 较好。而在 NDCG@k 上，WAA-TA 算法稳步提升。总体而言，WAA-TA 算法比基线算法更有竞争力且效果更为优秀。根据表 3 可以观察出，在 MOOCCubeX 数据集中，WAA-TA 在 Recall@k 中全部优于其他基线算法。而在 NDCG@k 中的 WAA-TA 与 GS-RS 大致相当，都优于其他基线算法。且随着 k 值增大为 50 时，WAA-TA 的算法能力稍逊于 GS-RS 的算法能力。在 Edx 中，WAA-TA 在 NDCG@k 中同样领先于其他基线算法。而在 Recall@20 中，ESM2 稍领先于 WAA-TA 算法，但对比其他算法发现，ESM2 的算法上升阶段规律性不强。

数据集性能指标如图 3 所示，WAA-TA 比其他基线算法均有所提升。WAA-TA 算法的出色表现主要归因于其动态调整任务权重的能力，以及利用任务集合来代表用户在不同生命周期阶段的学习需求。这种动态权重机制允许算法根据用户兴趣的实时变化进行调整，增强了推荐的个性化和准确性。此外，算法中融合的加权辅助任务感知机制，使得算法能够更有效地处理多任务学习中的信息共享与任务特异性，从而提高了推荐系统的性能。



a）Edx NDCG 性能比较

b）MOOCCubeX NDCG 性能比较

c）Edx Recall 性能比较

d）MOOCCubeX Recall 性能比较

图 3　数据集性能指标对比

Fig. 3　Comparison of dataset performance metrics

3.2　消融实验

为了验证加权辅助任务、task 特征分解和特征表正交约束对本文算法 WAA-TA 性能的影响，本文使用 WAA-TA 的 3 个变体，在 Edx 和 MOOCCubeX 两个数据集上做了消融实验研究。

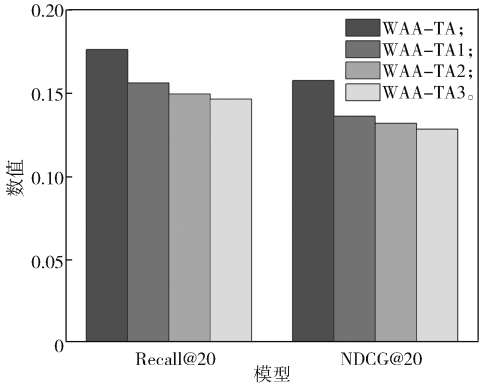
1）WAA-TA1：在这个变体中，去掉加权辅助任务的影响。

2）WAA-TA2：在这个变体中，去掉任务特征分解的影响。

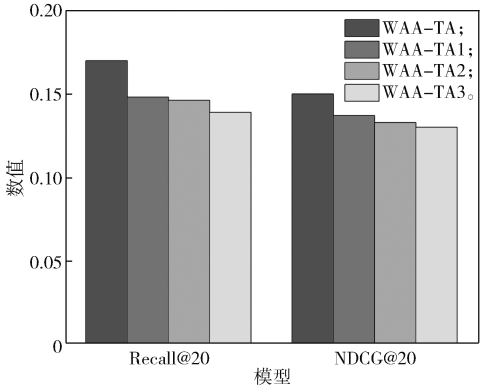
3）WAA-TA3：在这个变体中，去掉特征表正交约束的影响。

图 4 呈现了各变体与原算法在数据稀疏性较高的 Edx 数据集和 MOOCCubeX 数据集上的实验结果。观察图 4 可知，消融算法 WAA-TA1 和 WAA-TA2 相较于 WAA-TA3 有一定的提升。然而相较于原算法 WAA-TA，两者在性能上都存在不足之处。在 Edx 和 MOOCCubeX 数据集上，去掉加权辅助任务或任务特征分解的影响（WAA-TA1 和 WAA-TA2）导致算法性

能下降，而去除特征表正交约束（WAA-TA3）对性能影响相对较小。这表明加权辅助任务和任务特征分解是 WAA-TA 算法中关键的组成部分，它们协同工作以提升推荐质量。本文验证了 WAA-TA 算法中作为推荐任务的各个模块的设计合理性和有效性，进一步强调了这些模块相互协作的不可或缺性。



a）MOOCCubeX 数据集



b）Edx 数据集

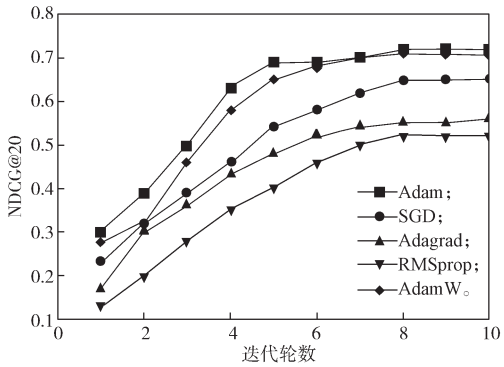
图 4　数据集消融实验数据

Fig. 4　Dataset from ablation experiments

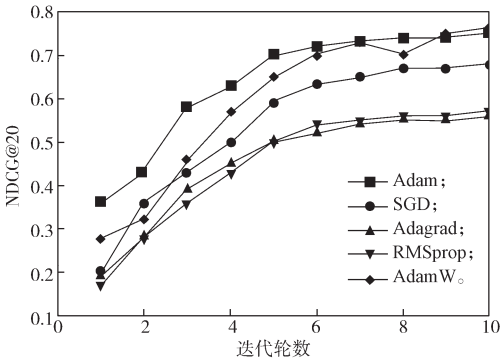
3.3　优化器对比实验

为了分析不同优化器对性能的影响，本研究选取 SGD、Adagrad、RMSprop、Adam 和 AdamW 优化器训练 WAA-TA 算法，并在 AEdx 和 MOOCCubeX 两个数据集上进行实验，所得结果如图 5 所示，其中横坐标是迭代轮数，纵坐标是 NDCG@20 评估指标下的值。由图 5 可以推断出 Adam 和 AdamW 具有较快的收敛速度，又因两者结合了 Adagrad 和 RMSprop 的优点，所以领先于其他优化器。SGD 在收敛过程中容易陷入局部极小值，收敛速度相对较慢。而 Adagrad 与 RMSprop 收敛速度则比较缓慢。

Adam 和 AdamW 优化器在 Edx 和 MOOCCubeX 数据集上显示出了更快的收敛速度。这是因为 Adam 和 AdamW 结合了 Adagrad 和 RMSprop 的优点，能够自动调整学习率，有效避免了 SGD 等优化器容易陷入局部极小值的问题。



a）Edx 数据集优化器对比



b）MOOCCubeX 数据集优化器对比

图 5　不同优化器的对比实验结果

Fig. 5　Comparison experiment results with different optimizers

4　结论与展望

本研究提出了一种多任务推荐算法 WAA-TA，旨在解决个性化推荐系统中的挑战。通过引入特定任务的特征分解、特征表征正交约束和辅助任务约束，WAA-TA 能够更好地学习用户和物品之间的关系，提高推荐准确性。在 Edx 和 MOOCCubeX 两个数据集上的广泛实验验证了 WAA-TA 算法在不同场景下的优越性。在未来将考虑从以下几个方面进行研究：1）将所提出的 WAA-TA 算法拓展到其他领域，如电子商务、社交媒体等，验证其在不同应用场景中的泛化能力；2）研究如何融合异构信息，包括图结构、文本信息等，这要求对算法进行扩展，以便更好地理解和利用这些额外信息，从而增强推荐的多样性和准确性；3）以进一步提升 WAA-TA 算法对用户兴趣的建模能力为目标。

同时本文也存在一些改进方法作为下一步的研究：1）动态权重机制优化，尽管 WAA-TA 算法已经实现了动态权重调整，但进一步研究如何更精确地捕捉用户兴趣的微小变化，以及如何在不同任务之间实现更智能的权重分配，将是提高推荐质量的关键，未来使用强化学习与此方法相结合以便达到更精确

的推荐效果。2）构建一个能够及时收集和分析用户反馈的机制，用以持续优化推荐算法。这不仅能够提高用户满意度，还能帮助系统快速适应市场和用户偏好的变化。

参考文献：

[1] 成亚玲，谭爱平 . 如何帮助学习者走出学习资源迷航：基于学习者画像的个性化学习资源推荐 [J]. 当代职业教育，2023(2)：103−112.
CHENG Yaling，TAN Aiping. How to Help Learners Get out of Learning Resource Maze：Personalized Learning Resource Recommendations Based on Learner Persona[J]. Contemporary Vocational Education，2023(2)：103−112.

[2] BALABANOVI M，SHOHAM Y. Fab：Content-Based，Collaborative Recommendation[J].Communication of the ACM，1997，40(3)：66−72.

[3] 佘学兵，蔡爱平，刘承启 . 基于知识图谱的网络用户长短期偏好推荐算法 [J]. 计算机仿真，2021，38(8)：425−429.
SHE Xuebing，CAI Aiping，LIU Chengqi. Long-Term and Short-Term Preference Recommendation Algorithm for Network Users Based on Knowledge Graph[J]. Computer Simulation，2021，38(8)：425−429.

[4] 刘 震，赵 嵩，杨 涛，等 . 基于深度学习的施工安全隐患整改智能推荐系统 [J]. 大数据，2023，9(6)：124−136.
LIU Zhen，ZHAO Song，YANG Tao，et al. Intelligent Recommendation System for Rectification of Construction Safety Hazards Based on Deep Learning[J]. Big Data Research，2023，9(6)：124−136.

[5] 方 宝，王 亮 . 基于多模态特征融合的多任务学习推荐 [J]. 物联网技术，2023，13(11)：78−80.
FANG Bao，WANG Liang. Multi-Task Learning Recommendation Based on Multi-Modal Feature Fusion[J]. Internet of Things Technologies，2023，13(11)：78−80.

[6] STANDLEY T，ZAMIR A，CHEN D，et al. Which Tasks Should Be Learned Together in Multi-Task Learning[C]//International Conference on Machine Learning. Shenzhen：PMLR，2020：9120−9132.

[7] JI G L，HE S Z，XU L H，et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1：Long Papers). Beijing：ACL，2015：687−696.

[8] KIPF T N，WELLING M. Semi-Supervised Classification with Graph Convolutional Networks[EB/OL]. [2024−09−15]. https://arxiv.org/abs/1609.02907v4.

[9] 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. London：ACM，2018：974−983.

[10] 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. London： ACM，2018：1930−1939.

[11] 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]//Fourteenth ACM Conference on Recommender Systems. Rio de Janeiro：ACM，2020：269−278.

[12] ZHOU J，YU Q，LUO C，et al. Feature Decomposition for Reducing Negative Transfer：A Novel Multi-Task Learning Method for Recommender System (Student Abstract)[J]. Proceedings of the AAAI Conference on Artificial Intelligence，2023，37(13)：16390−16391.

[13] MA X，ZHAO L Q，HUANG G，et al. Entire Space Multi-Task Model：An Effective Approach for Estimating Post-Click Conversion Rate[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. Ann Arbor：ACM，2018. http://arxiv.org/pdf/1804.07931.

[14] DUNNE J P，JOHN J G，ADCROFT A J，et al. GFDL’S ESM2 Global Coupled Climate-Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics[J]. Journal of Climate，2012，25(19)：6646−6665.

[15] XU Y B，WANG E，YANG Y J，et al. GS-RS：A Generative Approach for Alleviating Cold Start and Filter Bubbles in Recommender Systems[J]. IEEE Transactions on Knowledge and Data Engineering，2024，36(2)：668−681.

（责任编辑：姜利民）