融合多模态特征的在线课程资源推荐方法

冯 勇， 王 鸿， 申锦涛， 徐红艳
（辽宁大学 信息学院，沈阳 110036）

摘 要：当前在线课程资源繁杂，现有多模态推荐方法难以深度挖掘各模态间的内在关联，普遍存在推荐精度不高、大规模数据运算效率偏低等缺陷。为此，提出一种融合多模态特征的在线课程资源推荐方法。首先，进行课程向量构建，使用 Word2Vec 和 VGG-16 网络对属性级短文本特征、图像特征及长文本特征3种模态进行嵌入，从而得到3种模态特征，3种模态间利用注意力机制进行交互融合，得到课程多模态特征向量。其次，进行用户向量构建，利用预先训练好的嵌入矩阵对用户属性特征进行嵌入，再结合用户历史交互课程向量，得到用户兴趣画像。最后，对两向量计算生成候选集，由点击预测模块计算预测得分，进而得到 Top-N 课程进行推荐。实验结果表明：与对比方法中的最优值相比，所提方法在 MOOC 数据集上的 P 提高了 15.4%，R 提高了 7.5%，$F_{1\text{-score}}$ 提高了 11.6%，α_N 提高了 1.2%。

关键词：个性化推荐；注意力机制；多模态；特征融合；在线课程
DOI：10.15938/j.jhust.2026.03.001
中图分类号：TP391 **文献标志码**：A **文章编号**：1007-2683(2026)03-0001-12

Online Course Resource Recommendation Method Fusing Multimodal Features

FENG Yong， WANG Hong， SHEN Jintao， XU Hongyan
(College of Information, Liaoning University, Shenyang 110036, China)

Abstract: Given the current overload of online course resources, existing multimodal recommendation methods fail to deeply exploit correlation information between different modalities. This limitation results in low recommendation accuracy and computational efficiency when handling large-scale multimodal data. This paper proposes a method for recommending online course resources by integrating multimodal features. Firstly, a course vector was constructed using Word2Vec and VGG-16 network to embed attribute-level short text features, image features, and long text features, yielding three modal features. An attention mechanism was then employed to interact and fuse these modalities into a multimodal feature vector for the course. Secondly, a user vector was constructed by embedding user attribute features using a pre-trained matrix; the user interest profile was obtained by combining the user's historical interaction with the course vector. Subsequently, candidate sets were generated through vector calculations, prediction scores were computed via the click prediction module, and Top-N courses were recommended accordingly. Experimental results demonstrate that compared to optimal values from comparative methods on MOOC dataset precision@10 increases by 15.4%, 7.5%, 11.6%, and 1.2% respectively; furthermore, there is an improvement of 3.8% on MovieLens-100K dataset Precision@10.

Keywords: personalized recommendation; attention mechanism; multimodal; feature fusion; online courses

收稿日期：2024-07-10
基金项目：辽宁省教育厅科学研究基金(LJKMZ20020447)；辽宁省本科教学改革项目(2022-21)；辽宁省属本科高校基本科研业务费专项资金(LJKLJ202414).
作者简介：冯 勇(1973—)，男，博士，教授；
王 鸿(1999—)，女，硕士研究生.
通信作者：徐红艳(1972—)，女，硕士，教授，E-mail：xuhongyan@lnu.edu.cn.

0 引 言

随着在线学习方式被大众普遍接受,MOOC[1]、网易云课堂[2]、荔枝微课[3]和学堂在线[4]等学习平台不断充实课程资源,以满足广大学习者的需求。然而,在众多课程资源中,学习者常常难以发现符合自身兴趣的学习资料与课程。因此,课程推荐系统被广泛用于过滤无用信息、推荐适宜课程资源,以增强学习者的学习成效。

在课程资源推荐方面具有代表性的研究成果有:WANG 等[5]通过改进循环神经网络获取学生的信息(如历史点击率等)来分析学生的个性化兴趣,使课程推荐的准确性得以提高。马文蕴[6]利用全局视角下的课程顺序图来表示跨序列课程关系,并利用注意力机制进行建模,同时使用位置信息和序列编码器重新整合学习者偏好。但该方法处理大规模数据时效率低下,不适用于海量的在线课程推荐。SHU 等[7]利用卷积神经网络(CNN)提取学习资源中的文本特征,并用于基于内容的推荐系统,显著提升了推荐的精准度,但无法为学习者推荐多样化的学习内容。综上可知,现有单模态在线课程推荐方法的推荐准确率已经处于较高水平,但由于课程介绍包含文本和图片等多个模态,如果仅从单一模态出发进行推荐,会使推荐结果过于片面。因此,将多个模态特征融合后再进行推荐,可进一步提升推荐的准确率。

由于单模态的局限性使推荐结果过于片面,故已有学者开始考虑多模态融合的课程资源推荐。如 WU 等[8]提出的图注意力网络(GAT)中,通过计算图中节点间的 Attention 权重,实现对相邻节点信息的融合,从而获取更加精确的节点特征表示。但这些信息在聚合过程中可能会引入噪声,从而导致推荐效果不佳。XU 等[9]提出了基于深度学习模型的课程推荐方法,通过提取课程名称、音频和评论等多模态特征来改进基于内容的课程推荐方法,使推荐结果的准确性提高。潘华莉等[10]提出了 M2DR-RM 方法,通过构建多模态的表示网络,使用注意力实现模态间的交互融合,并引入 Actor-Critic 强化学习方法,根据用户反馈实时改进推荐策略,实现个性化推荐。但该方法在处理多模态信息并实时调整推荐策略时资源消耗较大,需要更多的计算资源和时间,对硬件设备和算法工程师提出了较高的要求。尤其是在处理大规模数据和高并发场景时,可能需要更强大的计算能力和优化算法。综上所述,虽然已有的基于多模态的推荐方法已经取得了不错的效果,但却未将课程的属性特征(如课程名、类别和开课单位等)、文本特征(如课程介绍)和图像特征(如课程封面)充分结合起来,故基于多模态的在线课程资源推荐方法尚存在提升空间。

为进一步提高在线课程资源推荐的准确度,本文提出了融合多模态特征的在线课程资源推荐方法(online course resource recommendation method fusing multi-modal features,CRFMF)。首先,进行课程向量构建,使用 Word2Vec 词向量和 VGG-16 网络对属性级短文本特征、图像特征及长文本特征等 3 种模态进行嵌入表征,从而得到 3 种模态特征。其一,由于课程标题具有对科目全覆盖、简洁的概括(如专业、最核心内容等),因此,课程标题是对 1 门课程最准确的表示,所以将其与课程介绍和课程封面进行两两交互。其二,为了从课程封面中提取出与课程文本介绍贴近的特征,进行课程文本介绍与课程封面融合操作来提取出更准确的课程信息。因此,将 3 种模态利用注意力机制进行模态间交互融合,用一种模态对另一种模态加权,筛选出与指导模态更贴近的信息,得到课程的多模态特征向量。同时,通过用户特征向量和历史交互课程向量两部分来构建用户兴趣向量。接下来,使用 Word2Vec 和嵌入矩阵,以向量形式拟合课程特征向量和用户兴趣向量。然后,对两向量计算生成候选集,由点击预测模块计算出候选课程的预测得分,进而得到 Top-N 课程进行推荐。最后,在学堂在线采集到的 MOOC 数据集及 MovieLens-100K 公开数据集上进行对比实验,实验结果表明本文所提方法相较于现有主流方法具有更好的性能。

本文的主要贡献如下:

1)提出多模态课程向量构建方法,利用标题、文本、图像 3 种单模态特征交互建模。采用注意力机制进行模态间交互融合,并使用残差连接思想减少过拟合,较准确地刻画了课程内容画像;

2)提出新颖的推荐方法 CRFMF,该方法通过多模态课程向量与用户向量计算得到候选集后,加入点击预测模块,进一步缩小推荐范围,提高了推荐准确率,并提高了推荐列表与理想推荐顺序的接近

程度;

3)在 MOOC 数据集和公开数据集 MovieLens-100K 上进行了充分实验,验证了方法有效性。结果表明本文方法不仅在在线课程推荐领域取得了更好的性能,对具有相似数据形式的电影推荐性能也有所提升。

1　相关概念

1.1　注意力机制

注意力机制是专注于序列的权重系数,同时与序列中各个位置密切相关的机制[11]。能够聚焦于输入序列的关键信息,并忽略其中不相关的信息,从而提高模型的整体性能。因此,越来越多的研究人员尝试利用注意力机制探索模态内部和不同模态之间的交互作用。其核心思想是通过计算相似度概率分布函数来确定每个特征对输出的贡献程度,从而为重要的信息分配更多的注意力。针对给定的查询向量($\boldsymbol{Q}$),键向量($\boldsymbol{K}$)和值向量($\boldsymbol{V}$),注意力机制会根据 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 之间的相关性为 $\boldsymbol{V}$ 分配权重 s_i,d 为键向量维度,并应用 Softmax 函数进行归一化操作得到权重系数,以凸显核心元素的权重,这个权重越大,意味着该信息越重要,如式(1)~(2)所示。

$$s(\boldsymbol{K}_i,\boldsymbol{Q}) = \frac{\boldsymbol{K}_i^{\mathrm{T}}\boldsymbol{Q}}{\sqrt{d}} \tag{1}$$

$$\alpha_i = \mathrm{Softmax}(s_i) = \frac{\mathrm{e}^{s_i}}{\sum_{j=1}^{L_x}\mathrm{e}^{s_i}} \tag{2}$$

最后将权重系数 α_i 与对应的值向量 $\boldsymbol{V}$ 进行加权计算,获得最终的注意力值 $\boldsymbol{A}$,如下:

$$\boldsymbol{A} = \sum_{i=1}^{L_x}\alpha_i\boldsymbol{V}_i \tag{3}$$

注意力机制在本文中主要用于模态间的交互融合,用指导模态对被指导模态进行加权,从而筛选出与指导模态更贴近的信息。

1.2　多模态特征融合

特征融合是将抽取的不同特征向量进行优化组合[12]。以下是 2 种经典的特征融合方法:

1)concat:拼接特征融合。直接将 2 个特征进行拼接,若 2 个输入特征 x 和 y 的维数为 p 和 q,则输出特征 z 的维数为 $p+q$。

2)add:加和特征融合。将 2 个特征向量组合成复向量,对于输入特征 x 和 y,输出特征 $z = x + \mathrm{i}y$。其中,i 为虚数单位。

特征融合在多个领域中得到了广泛应用,包括但不限于语义视频分析、多视图图片分类及视觉问答系统等。然而,在多模态推荐领域,相关研究尚显不足。这主要是因为多模态推荐任务涵盖了诸如视频、图片、文本等多种模态的信息,使得其处理过程较为复杂。尽管如此,仍可以从文本 - 图片检索的相关研究中汲取经验,以推进多模态推荐技术的发展。例如,在 LI 等[13]提出的 W2VV + + 模型中,将词袋模型、word2vec 和 GRU 3 种不同的文本预训练模型的输出特征进行融合得到最终文本特征。而 ZHOU 等[14]提出的 SEA 模型,构建统一的文本特征空间,对这个空间内学习到的各个特征进行平均处理,从而得到最终特征。

本文对前述策略进行了精简与优化,将那些卓有成效的方法与探究予以保留。首先,进行课程向量构建,使用 Word2Vec 词向量和 VGG-16 网络对属性级短文本特征、图像特征及长文本特征等 3 种模态进行嵌入,从而得到 3 种模态特征。3 种模态间利用注意力机制进行交互融合,得到课程多模态特征向量。这样不仅保留了上述方法的优势,还有效解决了传统的基于多模态特征融合的推荐方法带来的资源消耗较大、需要更多的计算资源和时间及各模态考虑不充分、融合技术效率低等问题。并且在特征融合的最后阶段引入了注意力机制,以满足建模用户对各个模态的偏好。这一举措进一步提升了推荐的个性化程度,使得推荐方法能够更准确地捕捉用户的兴趣。

2　融合多模态特征的在线课程资源推荐方法

2.1　方法框架

为解决在线课程推荐准确率低下的问题,本文提出了 CRFMF 方法,其核心思想是将用户行为数据和课程多模态特征数据进行计算,以提供个性化的课程推荐。通过分析用户的行为数据,如学习记录、收藏记录和评分记录,了解用户的行为模式和关注点。通过引入多模态特征数据,获取更丰富的课程信息。例如,属性级短文本特征由课程的标题获得,长文本特征从课程文本介绍中提取,图像特征可通过课程封面提取。这样,在进行特征融合时,可以将不同模态特征进行交互融合,形成更全面的特征

表示。CRFMF 方法通过从用户行为数据中学习到的规律及与课程信息和多模态特征相关的知识,能够更加准确地推荐符合用户兴趣的课程资源。

方法的整体框架如图 1 所示,主要由 3 部分组成:①课程向量构建;②用户向量构建;③点击预测。CRFMF 相较于现有课程推荐方法,在课程向量构建中,使用预先训练好的 Word2Vec 词向量和在 ImageNet 上预训练的 VGG-16 网络对课程的标题属性、文本介绍和课程封面进行嵌入表征,从而得到 3 种单模态特征。3 种模态间利用注意力机制进行交互融合,最终得到课程多模态特征向量。此外,在生成推荐列表前,加入点击预测模块,用于计算用户学习所推荐课程的概率得分,根据得分进行排序从而生成 TOP-N 课程进行推荐,提高推荐准确率。

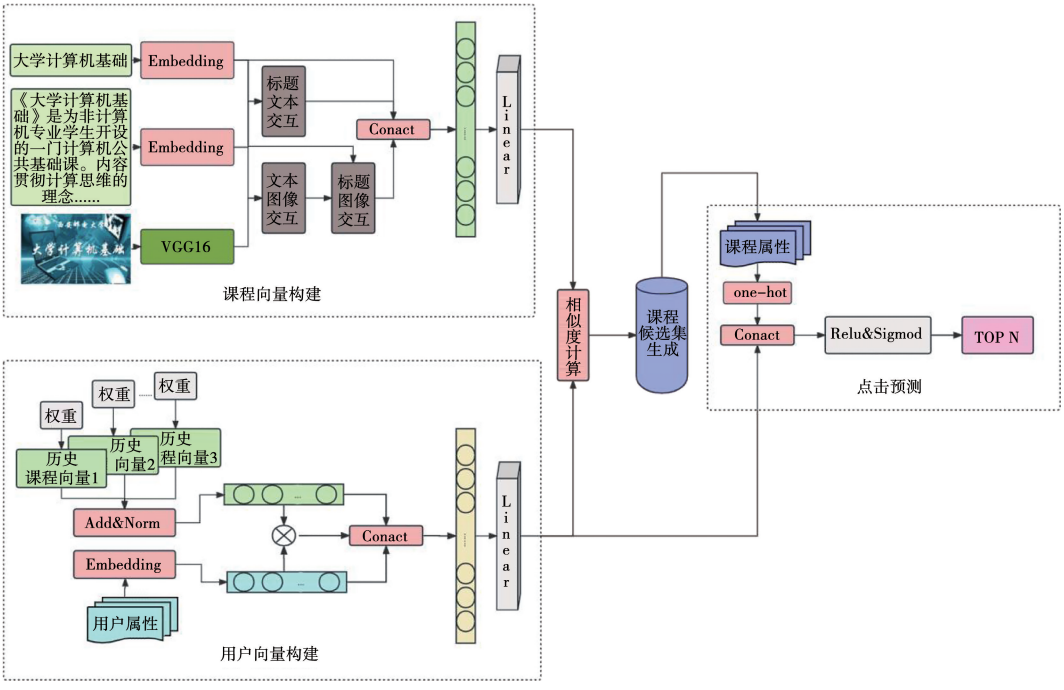


图 1 CRFMF 方法框架

Fig. 1 Method framework of CRFMF

2.2 课程向量构建

在线课程推荐领域与其他领域最大的不同在于,课程推荐与否取决于课程内容本身。用户在未观看课程视频之前,了解课程信息的主要途径有课程标题、课程封面和课程简介,其次才是课程的其他特征[15]。其中,课程简介可以全面概括课程内容,对课程向量构建起主要作用;课程标题可以高度概括课程内容;课程封面在一定程度上会吸引用户,同时也会包含一些课程信息。因此,课程向量是 1 种融合多模态特征的特征向量,课程向量构建包括单模态特征提取和多模态特征融合 2 个部分,如图 2 所示。

2.2.1 单模态特征提取

1)标题特征

将课程标题通过 Word2Vec 嵌入为课程标题向量 $\boldsymbol{T}$。在训练时,由于课程标题中每个分词对课程来说都具有代表性,故对每个分词赋予相同重要性。

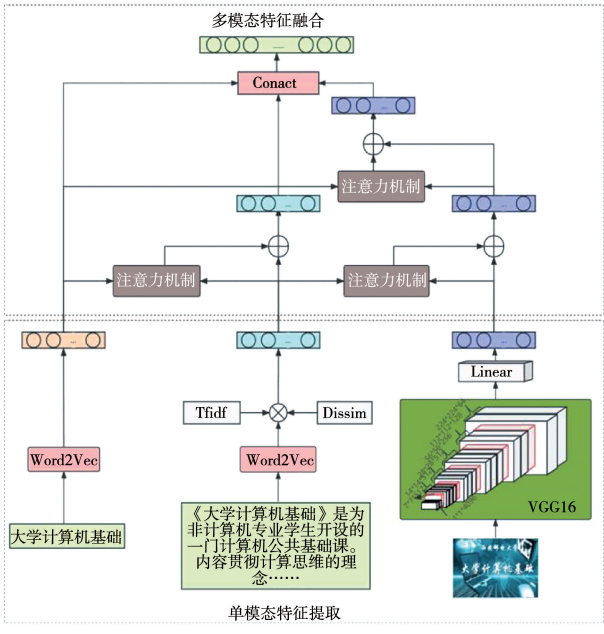


图 2 课程向量构建

Fig. 2 Course vector construct

2)文本特征

对课程多模态数据集来说,课程介绍由大量文字组成,详细阐述了课程内容,故在对其做分词及去除停用词处理后,其中仍然包含大量与课程内容相关性小的分词。为此,采用 1 种新的针对文本特征的向量构建方法,如式(4)所示:

$$\boldsymbol{D}=\frac{\sum_{n=1}^{k}\boldsymbol{d}_n\alpha_{\mathrm{T}}\alpha}{\sum_{n=1}^{k}\alpha_{\mathrm{T}}\alpha} \tag{4}$$

式中:$\boldsymbol{D}$ 为课程介绍向量;k 为对课程介绍进行分词后的分词数量,对于其中的第 n 个分词,通过 Word2Vec 嵌入其对应的词向量 $\boldsymbol{d}_n$。此外,为了更精准地调整各个词语在课程介绍中的重要性,为每 1 个分词的词向量赋予了 2 个权重:分词的 Tf-idf 值 α_{T} 和逆相似度权重 α[16]。Tf-idf 值在关键词抽取中扮演着重要角色,可有效降低通用高频词的权重,同时增强了课程介绍中独特且具代表性的词的权重,降低了无意义词的影响。而引入逆相似度权重则进一步优化了课程介绍的向量表示,其计算方法如下:

$$\alpha=\begin{cases}0, \boldsymbol{d}_n\in\boldsymbol{T}\\ 1-s(\boldsymbol{d}_n,\boldsymbol{T}), \boldsymbol{d}_n\notin\boldsymbol{T}\end{cases} \tag{5}$$

逆相似度权重的取值取决于分词是否出现在标题分词集中。如果分词与标题分词重复,则权重设为 0,这样可以降低与标题高度相关的分词在课程介绍中的重要性。这种处理方式符合人类的直觉,因为标题作为课程内容的高度概括,通常包含了课程的核心信息,而课程介绍则是对标题的进一步扩展和细化。因此,在构建课程向量时,不需要过分强调那些与标题重复的分词。

另外,对于不在标题分词集中的分词,需要计算其与标题向量的余弦相似度,并对其做逆运算获得权重,突出那些与标题不相似或差异较大的分词。这些分词往往代表了课程介绍中独有的、区别于其他课程的信息,是课程向量的重要组成部分。通过降低与标题相似度高的分词的权重,可以更好地凸显这些独特的信息,从而提高课程向量的区分度。

3)图像特征

本文在对课程封面图像 image 进行特征提取时,采用了在 ImageNet 上完成预训练的 VGG-16 网络[17]。首先,将输入图像的大小处理为 $224\times224\times3$,符合网络结构的处理需求。随后,图像会经过一系列卷积层,这些卷积层的卷积核大小固定为 3×3,而通道数则随着网络的深入而逐渐增加。在卷积操作之后,通过最大池化层对特征进行降维。最大池化层使用 2×2 大小的滤波器,能够在保留重要特征的同时,减少数据的维度,从而加速后续的计算过程。随着网络的深入,卷积层的通道数进一步增加,以提取更高级别的特征。这些特征不仅包含了图像的局部信息,还融入了更多的上下文信息,能够更好地理解图像的内容。在卷积和池化操作之后,依次经过节点数分别为 4096、4096 的全连接层将特征整合起来。这些全连接层使用了 ReLU 激活函数,能够增加网络的非线性表达能力,使得网络能够学习更复杂的特征映射关系。

本文在得到整合后的 4096 维特征向量后,便将其映射到与标题特征和文本特征相同的向量空间作为最终的图像特征 $\boldsymbol{V}$,如下:

$$\boldsymbol{V}=\mathrm{VGG}(\mathrm{image}) \tag{6}$$

式中:VGG 为 VGG-16 网络的计算。

2.2.2　多模态特征融合

在课程多模态数据中,不同模态对于课程的贡献程度不同,一般来说文本模态贡献最大,图像模态贡献相对较小,并且不同模态间存在相关性[18]。因此,将得到的 3 种单模态特征向量进行有效融合,实现特征互补,融合后的特征作为课程多模态特征向量。

对课程数据集来说,课程标题与课程封面中的内容有很强的相关性,课程介绍与课程封面之间也有一定的内在联系。为捕获这种联系,首先利用文本-图像注意力网络将文本特征与图像特征进行交互,如式(7)所示。同时,为了保留原始图像特征的完整性,采用残差连接的思想确保在交互过程中不会丢失关键信息,如式(8)所示:

$$\alpha_{\mathrm{dv}}=\frac{\boldsymbol{V}\boldsymbol{D}^{\mathrm{T}}}{\sqrt{d_D}} \tag{7}$$

$$\boldsymbol{V}_1=\alpha_{\mathrm{dv}}\boldsymbol{V}+\boldsymbol{V} \tag{8}$$

式中:α_{dv} 为文本特征与图像特征之间权重;d_D 为 $\boldsymbol{D}$ 的维度;$\boldsymbol{V}_1$ 为文本指导下的图像特征。

然后,将标题特征与文本特征、文本指导下的图像特征分别交互,均采用注意力机制且使用残差连接的思想,如式(9)和(10)所示:

$$\alpha_{\mathrm{tv}}=\frac{\boldsymbol{V}\boldsymbol{T}^{\mathrm{T}}}{\sqrt{d_T}},\alpha_{\mathrm{tv}}=\frac{\boldsymbol{D}\boldsymbol{T}^{\mathrm{T}}}{\sqrt{d_T}} \tag{9}$$

$$\boldsymbol{V}_2=\alpha_{\mathrm{tv}}\boldsymbol{V}_1+\boldsymbol{V}_1,\boldsymbol{D}_1=\alpha_{\mathrm{td}}\boldsymbol{D}+\boldsymbol{D} \tag{10}$$

式中:α_{tv},α_{td} 分别为标题-文本、标题-图像之间的权重;d_T 为 $\boldsymbol{T}$ 的维度。最后将交互后的 3 种特征进行拼接,得到融合后的课程多模态特征向量 $\boldsymbol{C}$,如式(11)所示。

$$\boldsymbol{C}=\mathrm{Linear}(\mathrm{Conact}(\boldsymbol{T},\boldsymbol{D}_1,\boldsymbol{V}_2)) \tag{11}$$

式中:$\boldsymbol{D}_1$ 为标题指导下的文本特征;$\boldsymbol{V}_2$ 为标题与文本共同指导下的图像特征。

2.3 用户向量构建

对于在线课程推荐领域而言,用户向量的构建是至关重要的环节,除构建用户画像外还要准确表征用户近期兴趣。首先,用户的基本信息,如用户名、学校、院系、专业、学历和性别等,是构成用户特征向量的基础。这些信息能够反映用户的背景和教育程度,对于推荐系统来说,它们有助于理解用户的潜在学习需求和兴趣点。此外,用户的近期兴趣,特别是与用户有过交互的课程,是推荐方法更为关注的部分。用户产生收藏、关注、学习行为,都意味着他们对这些课程内容产生了兴趣。将这些历史交互课程纳入用户向量的构建中,能够更准确地反映用户的当前学习状态和偏好。因此,将用户特征向量和历史交互课程向量相结合,可以形成更加全面和精确的用户向量,如图3所示。这样的用户向量不仅能够反映用户的长期学习需求,还能够捕捉用户的短期兴趣变化,从而提高推荐的准确性和个性化程度。

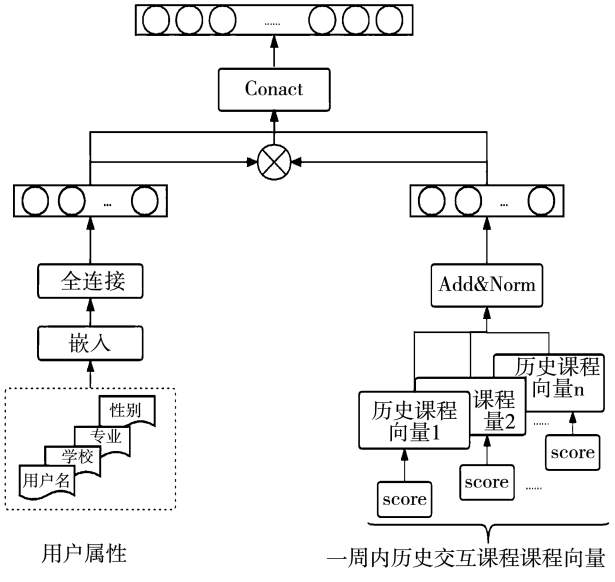


图3 用户向量构建

Fig. 3 User vector construction

1)用户属性特征向量

知识学习领域中,用户对某一知识点感兴趣的前提是,用户了解该知识点的背景知识,或者该知识点与用户所学专业领域的相似度很高。因此,用户的属性信息对用户向量的构建至关重要。首先,将用户属性信息(用户名、学校、院系、专业、学历、性别)通过特定的嵌入矩阵转换为向量表达。随后,将这些向量送入全连接层,进一步提取和凝练用户属性的特征向量,如式(12)所示,以用户名向量 $\boldsymbol{u}_{\mathrm{n}}$ 为例:

$$\boldsymbol{u}_{\mathrm{n}}=\mathrm{ReLU}\left(\boldsymbol{W}_{\mathrm{ne}}^{\mathrm{T}}\boldsymbol{u}_{\mathrm{ne}}+b\right),\boldsymbol{u}_{\mathrm{ne}}=\boldsymbol{E}_{\mathrm{ne}}\boldsymbol{u}_{\mathrm{name}} \tag{12}$$

式中:$\boldsymbol{W}_{\mathrm{ne}}$ 转置后为权重矩阵;$\boldsymbol{b}$ 为偏置;$\boldsymbol{u}_{\mathrm{ne}}$ 为用户名的嵌入向量;$\boldsymbol{E}_{\mathrm{ne}}$ 为嵌入矩阵;$\boldsymbol{u}_{\mathrm{name}}$ 为用户名的高维稀疏矩阵。再以同样的方法得到学校、院系、专业、学历、性别的向量表示,最后拼接为用户属性特征向量 $\boldsymbol{u}_A$,如式(13)所示:

$$\boldsymbol{u}_A=\mathrm{concat}\left(\boldsymbol{u}_{\mathrm{n}},\boldsymbol{u}_{\mathrm{s}},\boldsymbol{u}_{\mathrm{d}},\boldsymbol{u}_{\mathrm{m}},\boldsymbol{u}_{\mathrm{e}},\boldsymbol{u}_{\mathrm{g}}\right) \tag{13}$$

式中:$\boldsymbol{u}_{\mathrm{n}}$ 为用户名向量;$\boldsymbol{u}_{\mathrm{s}}$ 为学校向量;$\boldsymbol{u}_{\mathrm{d}}$ 为院系向量;$\boldsymbol{u}_{\mathrm{m}}$ 为专业向量;$\boldsymbol{u}_{\mathrm{e}}$ 为学历向量;$\boldsymbol{u}_{\mathrm{g}}$ 为性别向量。

2)用户历史交互课程向量

将用户近期交互历史中的课程多模态特征向量进行加权融合,如式(14)所示:

$$\boldsymbol{u}_{\mathrm{c}}=\frac{\sum_{i=1}^{j}\boldsymbol{C}_i S_{c_i}}{\sum_{i=1}^{j}S_{c_i}} \tag{14}$$

式中:$\boldsymbol{u}_c$ 为用户向量;c_i 为用户近期有交互记录的第 i 门课程;$\boldsymbol{C}_i$ 为课程向量;j 的取值是用户1周内所有交互过的课程数量;S_{c_i} 为用户对课程 c_i 的交互权重。

本文选取1周作为限制范围的依据主要来源于艾宾浩斯提出的遗忘曲线[19]。遗忘曲线揭示了人类遗忘的速率变化规律,即遗忘速度在初始阶段较快,随后逐渐减缓,表现为曲线的斜率从陡峭逐渐变得平缓。这一规律与用户对于内容的兴趣变化有着相似的趋势:用户在刚开始接触某内容时,兴趣通常最为浓厚,随着时间的推移,兴趣则逐渐衰减。在遗忘曲线到达第6天时,斜率已接近0,意味着遗忘速度已非常缓慢,之后的遗忘变化不再显著。因此,选择1周内的交互记录作为参考范围既能够捕捉到用户兴趣的高峰期,又能够避免因为过长时间的记录而引入过多的噪声和不必要的计算复杂度。

由于存在多种不同交互行为,所以为不同行为设置不同权重,如式(15)所示:

$$S_{c_i}=\begin{cases}S,\text{打分}\\ \dfrac{S_{\max}+S_{\min}}{2},\text{收藏}\\ S_{\max},\text{学习}\end{cases} \tag{15}$$

其中,用户对课程的打分 S 是重要参考依据,同时结合用户历史打分的上限 $S_{\max}$ 和下限 $S_{\min}$ 阈值,旨在消除不同用户因打分标准差异而带来的偏差。具体来说,若用户对课程进行了打分操作,则直接将其打分值设为权重;若用户执行了学习操作,这通常表明用户当时的兴趣度较高,将赋予其更高的权重;而收藏操作则意味着用户对课程有兴趣,但尚未立即开始学习,故将权重设定为历史打分上限和下限的

平均值,赋予适当的权重。

最后,将用户属性特征向量与用户历史交互课程向量进行交互建模,再与用户属性特征向量和交互课程向量进行拼接,得到最终的用户向量表示,如下:

$$\boldsymbol{u} = \text{Linear}(\text{concat}(\boldsymbol{u}_a, \boldsymbol{u}_s \otimes \boldsymbol{u}_c, \boldsymbol{u}_c)) \tag{16}$$

式中:⊗为叉乘运算。

2.4　点击预测模块

点击预测模块的核心是 1 个多层感知机模型,它具有 3 个隐藏层,如图 4 所示,这种结构赋予了推荐方法足够的深度来捕捉用户与课程之间的复杂关系。该模块的输入是待推荐用户的用户向量与课程的标签类特征的拼接结果。这种拼接方式能够同时考虑到用户的个性化特征和课程的属性特征,为后续的点击预测提供了丰富的信息。

在课程的标签类特征中,分类、讲师和开课单位等属性经过 One-hot 编码转换成数值向量。这种编码方式能够处理离散型的数据,将每个属性转换为 1 个稀疏的向量,便于模型进行运算。而统计类特征,如得分和学习人数,则经过归一化处理,将数值统一到[0,1]区间。归一化能够消除不同特征之间的量纲差异,使得模型更加稳定,并提高了预测的准确性。在模块的输出层使用 Sigmoid 函数作为激活函数,Sigmoid 函数能够将模型的输出映射到[0,1]区间,表示用户学习该课程的概率。这种概率化的输出方式使得模型能够更直观地表达预测结果,便于后续的排序和推荐。除了输出层,其他全连接层均使用 ReLU 函数作为激活函数,ReLU 函数能够解决梯度消失问题并增强方法的非线性表达能力。最后,通过该模块计算得到候选集内所有课程的预测得分,再根据得分对这些课程进行排序,并向用户推荐得分较高的课程。

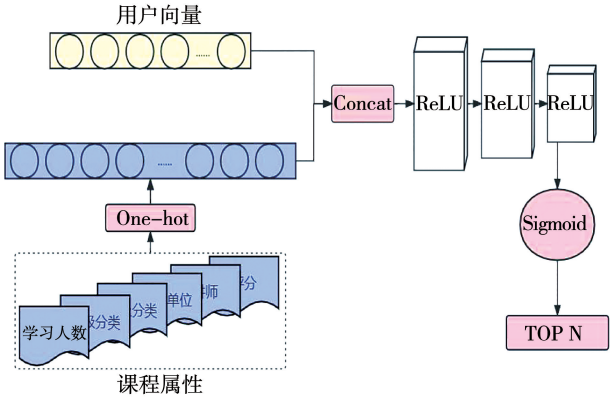


图 4　点击预测模块

Fig. 4　Click the prediction module

2.5　方法实现

CRFMF 方法的实现过程如算法 1 所示。其中 Course 为全部课程,History 为用户 1 周内产生交互的课程,CourseProfile 为课程的属性特征,UserProfile 为用户的属性特征。

算法 1:

输入:课程(Course)、交互历史(History)、课程特征(CourseProfile)、用户特征(UserProfile)

输出:推荐列表(TopN)

```
BEGIN
1 FOR i in Course Do:
2 CourseVec[i] = getCourseVec(i); //构建课程向量
3 ENDFOR
4 H = getHistory(History); //获取用户历史交互记录
5 IF H = NULL: //判断是否有交互记录
6 TopN = Hot(C); //热度推荐
7 Return TopN;
8 ELSE:
9 U = getUserVec(H, CourseVec, UserProfile); //构建用户向量
10 ENDIF
11 C = getSList(U, CourseVec); //生成课程推荐候选集
12 TopN = MLP(U, C, CourseProfile); //多层感知机预测得分
13 Return TopN; //推荐列表
END
```

3　实验与分析

3.1　实验设置

本文在实验中选用 2 个真实场景下的数据集为评估数据集。

由于本方法涉及的数据信息没有类似的公开数据集,故采用标准化的爬虫程序从中国最大的 MOOC 平台之一学堂在线采集数据[20]。对于采集到的数据,实施统一的数据清洗和预处理步骤,包括去除重复数据、填补空缺数据及数据格式的标准化,确保数据集的质量和一致性。处理后得到 23 个学科门类的 876 门课程的详细信息,50 666 名用户的 136 964 种学习 - 交互行为。该数据集中数据项包括:用户 ID、用户名、用户所在单位、所学专业、性别、用户最近一次学习课程时间、课程名称、课程类

别、课程类别 ID、课程学习人数、开课单位和授课老师。在划分训练集和测试集时,使用固定的随机种子来确保每次实验中数据集的划分都是一致的,确保了实验的可重复性和对比实验的公平性。

此外,还在 MOVIELENS-100K 数据集上进行了实验,验证该方法是否适用于与课程推荐具有相似数据形式的推荐任务。MOVIELENS-100K 数据集来源于互联网电影资料库(IMDB),是记录用户观看视频行为的数据集,包含对电影的评分行为,但缺失图像信息。因此,在 IMDB 网站上爬取每部电影的简介和海报,过滤信息不完整的数据后,保证每条数据均具备基本属性、文本介绍和图像信息[21]。数据集信息如表 1 所示,2 个数据集均使用随机种子选取 80% 作为训练集,另外 20% 作为测试集。

表 1 数据集统计数据

Tab. 1 Statistics for the dataset

数据集	用户数	项目数	交互数	类别数
MOOC	50 666	876	136 964	23
MovieLens-100K	943	1 682	100 000	18

本文选用推荐方法中常用的准确率 P、召回率 R、F1 - 分数($F_{1-\text{score}}$)和归一化折损累计收益 α_N 作为评估指标来衡量所提方法的推荐性能,如式(17)~(20)所示:

$$P=\frac{\sum_{u=1}^{N}|R(u)\cap T(u)|}{\sum_{u=1}^{N}|R(u)|} \tag{17}$$

$$R=\frac{\sum_{u=1}^{N}|R(u)\cap T(u)|}{\sum_{u=1}^{N}|T(u)|} \tag{18}$$

$$F_{1-\text{score}}=\frac{2PR}{P+R} \tag{19}$$

$$\alpha_N=\frac{\alpha_{DC}}{I_D} \tag{20}$$

式中:$R(u)$ 为推荐的课程列表,$|R(u)|=N$;$T(u)$ 为测试集中的课程学习列表;N 为推荐课程的个数;α_N 为归一化折损累计收益,评估推荐列表的排序优势,衡量长度为 N 的推荐列表与理想排名的接近程度,如式(21)所示:

$$\alpha_{DC}=\frac{1}{N}\sum_{u=1}^{N}\sum_{i=1}^{N}\frac{r_i-1}{\log(i+1)},r_i\in[0,1] \tag{21}$$

式中:α_{DC} 为第 i 门课的兴趣值;I_D 为最优排列的 α_{DC},当推荐结果是完美排序(所有高相关的课程都排在最前面)时,$\alpha_{DC}=I_D$,此时 $\alpha_N=1$;当推荐结果完全不相关时 α_{DC} 远小于 I_D,α_N 趋于 0。

3.2 对比方法

VBPR[22]:基于视觉信息和评分信息的多模态推荐方法。将视觉信息作为辅助信息融入预测用户偏好中,实现了更准确的个性化排名。本文将课程封面作为视觉信息,结合课程评分进行推荐。

AMR[23]:基于视觉信息和评分信息的多模态推荐方法。通过对图像增加噪声来训练鲁棒性更强的多模态推荐模型。本文将课程封面作为视觉信息,并对其进行扰动,再结合课程评分进行对抗式学习训练,从而获得更鲁棒的推荐方法。

CSGMC[6]:基于课程顺序图的图注意力网络课程推荐方法。先根据用户的学习顺序构建课程顺序图,再将不同顺序中的相同课程连接起来学习其潜在联系,并利用注意力机制进行建模,最后重新整合用户偏好。

M2DR-RM[10]:融合多模态特征的深度强化学习推荐方法。考虑了项目的属性、文本和图像信息,利用多模态融合技术重新构建深度强化学习 ACTOR-CRITIC 的状态表示网络,解决了传统推荐方法忽略用户兴趣漂移和长期奖励及数据稀疏性的问题。

3.3 结果分析

为研究不同推荐列表长度 N 对推荐性能的影响,本文在 MOOC 数据集上对 N 从 1 ~20 的不同取值做了实验,实验结果如表 2 和图 5 所示。随着推荐数量的增加,$F_{1-\text{score}}$ 总体呈上升趋势,且本文推荐方法的效果整体好于其他对比方法,并且在推荐数量为 10 时逐渐趋于稳定。

表 2 不同 N 值的 $F_{1-\text{score}}$ 对比实验结果

Tab. 2 Comparison of $F_{1-\text{score}}$ experimental results with different N values

方法	$F_{1-\text{score@1}}$	$F_{1-\text{score@5}}$	$F_{1-\text{score@10}}$	$F_{1-\text{score@15}}$	$F_{1-\text{score@20}}$
VBPR	0.072 8	0.175 2	0.211 0	0.272 8	0.243 1
AMR	0.081 1	0.160 0	0.300 1	0.326 3	0.320 6
CSGMC	0.102 8	0.364 1	0.482 2	0.510 6	0.573 7
M2DR-RM	0.084 9	0.403 8	0.496 8	0.537 7	0.511 9
CRFMF	0.085 6	0.331 9	0.613 0	0.632 8	0.658 3

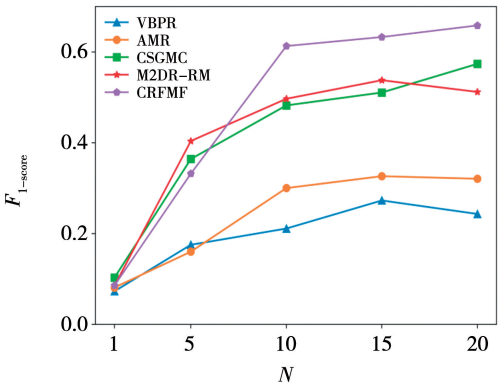


图 5　不同 N 值的 $F_{1-score}$ 对比实验结果图

Fig. 5　Comparison diagram of $F_{1-score}$ experimental results with different N values

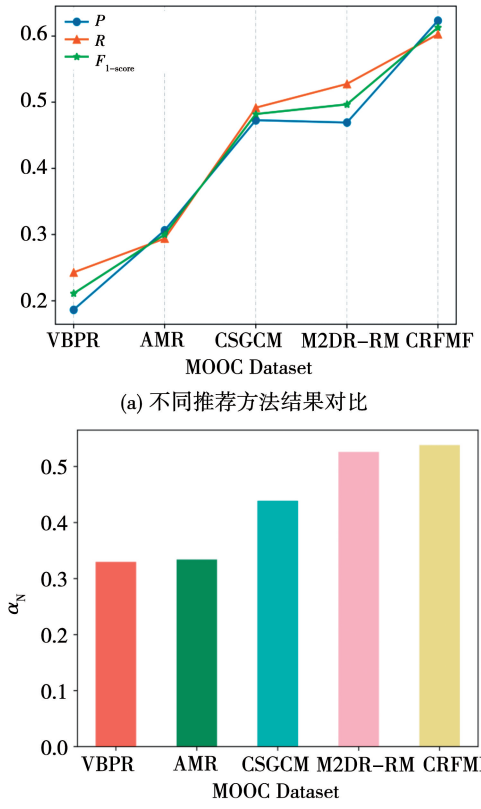
由实验结果可知,本文提出的推荐方法对比其他基线方法在所有数据集上的推荐性能均有提升。

表 3 为在 MOOC 数据集上的实验结果,图 6 为不同推荐方法在 MOOC 数据集上的 P、R、$F_{1-score}$ 和 α_N 结果对比。可知 CRFMF 与其他 4 种基线方法相比,在 4 个指标上的表现均为最优。与 VBPR 和 AMR 方法相比性能提升明显,是因为 CRFMF 并不仅仅依赖图片和评分构建课程向量,而是通过属性级标题特征、文本特征和图像特征相互融合构建多模态特征向量来表达课程信息,所以推荐性能有较大提升。与 CSGMC 方法相比,P 提高了 31.85%,R 提高了 22.57%,$F_{1-score}$ 提高了 27.13%。这是因为 CSGMC 只利用选课序列来学习课程和用户的表示,忽略了课程属性和用户属性等特征,而 CRFMF 则充分利用这些信息进行建模。与多模态推荐方法 M2DR-RM 相比,P 提高了 32.86%,R 提高了 14.23%,$F_{1-score}$ 提高了 23.39%。比较来说,本文方法在构建用户向量时,加入了用户的历史交互课程信息并限制了 1 周的范围,考虑了用户兴趣的衰减,实验结果表明:本文方法能够更好地建模用户兴趣,挖掘用户近期的交互信息。

表 3　MOOC 数据集对比实验结果($N=10$)

Tab. 3　Comparison of experimental results on the MOOC dataset($N=10$)

方法	P	R	$F_{1-score}$	α_N
VBPR	0.186 5	0.242 9	0.211 0	0.329 5
AMR	0.306 5	0.294 0	0.300 1	0.333 9
CSGMC	0.472 9	0.491 8	0.482 2	0.438 8
M2DR-RM	0.469 3	0.527 7	0.496 8	0.526 1
CRFMF	0.623 5	0.602 8	0.613 0	0.538 4



(a) 不同推荐方法结果对比

(b) 不同推荐方法的 α_N 对比

图 6　MOOC 数据集上对比实验结果图($N=10$)

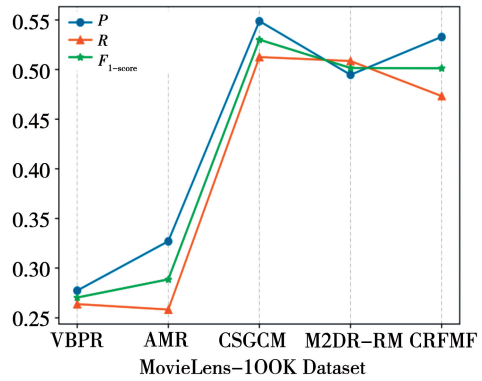
Fig. 6　Comparison diagram of experimental results on the MOOC dataset($N=10$)

表 4 为在 MOVIELENS-100K 数据集上的实验结果,图 7 为不同推荐方法在 MOVIELENS-100K 数据集上的 P、R、$F_{1-score}$ 和 α_N 结果对比。CRFMF 与 VBPR、AMR 方法相比,性能提升明显,与 M2DR-RM 方法的 $F_{1-score}$ 基本持平,CSGCM 表现最好。这是由于 CRFMF 方法依赖于丰富的用户和项目特征来构建准确的推荐模型,而在 MOVIELENS-100K 数据集中,缺失部分用户属性特征和电影属性特征,CRFMF 无法充分利用这些信息来表达用户兴趣,导致性能提升有限。

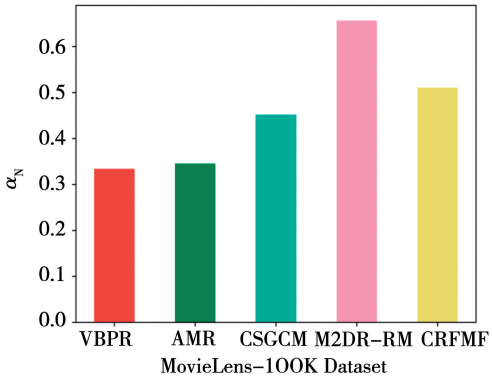
表 4　MovieLens-100K 数据集上对比实验结果($N=10$)

Tab. 4　Comparison of experimental results on the MovieLens-100K dataset($N=10$)

方法	P	R	$F_{1-score}$	α_N
VBPR	0.277 4	0.263 8	0.270 4	0.333 8
AMR	0.327 1	0.258 3	0.288 7	0.345 5
CSGMC	0.548 8	0.512 6	0.530 1	0.451 4
M2DR-RM	0.494 7	0.508 6	0.501 6	0.656 8
CRFMF	0.532 9	0.473 3	0.501 3	0.510 9



(a) 不同推荐方法结果对比



(b) 不同推荐方法的 α_N对比

图 7 MovieLens-100K 数据集上对实验结果图($N=10$)

Fig. 7 Comparison diagram of experimental results on the MovieLens-100K dataset($N=10$)

此外,CRFMF 在生成 TOP-N 推荐结果时加入了点击预测模块,这一策略在课程推荐中更为有效,因为用户对课程的选择可能更依赖于内容的质量和相关性。而在电影推荐中,用户的决策可能更多受到个人口味和娱乐需求的影响,这要求 CRFMF 模型需要更好地捕捉和预测这些复杂的用户行为模式,导致 CRFMF 在非课程推荐领域性能提升受限。

3.4 消融实验

本文所提方法充分考虑了课程的标题特征、文本特征和图像特征。为了验证融合多模态特征提高推荐性能的有效性,在 MOOC 数据集上进行了消融实验,结果如表 5 所示。

实验设置如下:

CRFMF-T:单模态输入。仅使用课程标题构建课程向量,删去多模态特征融合模块。

CRFMF-TD:双模态输入。使用课程的标题和文本介绍构建课程向量,以标题特征指导文本特征进行交互融合。

CRFMF-TDV:三模态输入。使用课程的标题、文本介绍和课程封面构建课程向量,3 种模态特征直接进行拼接融合。

CRFMF- RES:三模态输入。使用课程的标题、文本介绍和课程封面构建课程向量,模态间交互时采用注意力机制并结合残差连接思想进行融合。

表 5 消融实验

Tab. 5 Ablation experiment

方法	标题信息	文本信息	图像信息	模态融合	点击预测	$F_{1-score}$	α_N
CRFMF-T	√	×	×	×	×	0.435 1	0.337 2
CRFMF-TD	√	√	×	×	×	0.513 3	0.391 7
CRFMF-TDV	√	√	√	×	×	0.521 6	0.402 5
CRFMF-RES	√	√	√	√	×	0.562 9	0.456 1
CRFMF	√	√	√	√	√	0.613 0	0.538 4

注:“√”表示使用该模块,“×”表示不使用该模块。

CRFMF:本文模型。在多模态交互的基础上,在最终生成 TOP-N 推荐结果时加入点击预测模块,进一步提高推荐精确率。

实验结果如图 8 所示,3 种模态交互融合后 CRFMF 表现最好。在 CRFMF-T 中仅考虑了课程标题特征,此时由于存在数据稀疏性,F1 - 分数和归一化折损累计增益相对较低。CRFMF-TD 与 CRFMF-T 相比,加入了课程介绍文本特征,此时推荐性能会有所提升,这是因为文本特征作为辅助信息丰富了课程的表达,更准确地刻画了用户兴趣。CRFMF-TDV 与 CRFMF-TD 相比又加入了课程封面,但此时推荐性能提升不够明显,这是由于图像特征所表达的信息有限,且存在部分冗余信息,在直接拼接作为多模态特征时,噪声会在一定程度上影响推荐结果。CRFMF-RES 与 CRFMF-TDV 相比,虽然都加入了封面图像,但二者的融合方式不同,前者采用注意力机制进行交叉融合,后者则是直接拼接融合。结果显示:3 种模态特征交叉融合时,推荐性能提升显著,这是因为经文本特征指导后的图像特征再经标题特征指导时,更关注与课程相关的部分,经

标题特征指导的文本特征也更关注相关度高的文本内容。由此可见,课程封面图像信息的加入对推荐任务有一定的意义,但如何将该信息与文本信息进行融合更加重要。最后,加入点击预测模块后,推荐候选集会进一步缩小,推荐性能会进一步提高,且根据预测概率排序由高到低进行推荐,故归一化折损累计收益的提升较为明显。

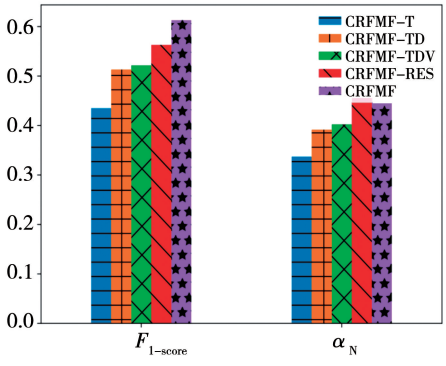


图 8　消融实验结果

Fig. 8　Ablation experiment results

4　结　语

本文针对目前多模态特征间存在特征融合准确率低及在线课程信息过载带来的课程信息获取效率低下的问题,提出了 CRFMF 方法。通过对比实验验证,本文提出的方法在 2 个数据集上均取得了不错的对比效果,相较于当下主流方法,性能都有所提升,能够有效解决当前因在线课程资源过载导致的在线课程推荐准确率低下等问题。尽管本文方法已经取得了一定阶段性成果,但是未能充分考虑对不同模态特征的权重分配,所以下一步将针对如何确定不同模态特征对推荐任务的重要性进行更深入的研究。

参 考 文 献:

[1] 中国大学 MOOC 平台[DB/OL]. (2024 - 07 - 10)[2024 - 04 - 11]. https://www.icourse163.org/.

[2] 网易云课堂平台[DB/OL]. (2024 - 07 - 10)[2024 - 04 - 11]. https://study.163.com/.

[3] 荔枝微课平台[DB/OL]. (2024 - 07 - 10)[2024 - 04 - 11]. https://www.lizhiweike.com/.

[4] 学堂在线平台[DB/OL]. (2024 - 07 - 10)[2024 - 04 - 11]. https://www.xuetangx.com/.

[5] WANG J, XIE H, WANG F L, et al. Top-N Personalized Recommendation with Graph Neural Networks in MOOCs[J]. Computers and Education: Artificial Intelligence, 2021, 2: 100010.

[6] 马文蕴. 基于慕课的在线开放课程推荐算法研究[D]. 济南: 山东师范大学, 2022.

[7] SHU J, SHEN X, LIU H, et al. A Content-based Recommendation Algorithm for Learning Resources[J]. Multimedia Systems, 2018, 24(2): 163.

[8] WU S, SUN F, ZHANG W, et al. Graph Neural Networks in Recommender Systems: A Survey[J]. ACM Computing Surveys, 2023, 55(5): 1.

[9] XU W, ZHOU Y. Course Video Recommendation with Multimodal Information in Online Learning Platforms: A Deep Learning Framework[J]. British Journal of Educational Technology, 2020, 51(5): 1734.

[10] 潘华莉, 谢珺, 高婧, 等. 融合多模态特征的深度强化学习推荐模型[J]. 数据分析与知识发现, 2023, 7(4): 114.
PAN Huali, XIE Jun, GAO Jing, et al. A Deep Reinforcement Learning Recommendation Model with Multimodal Features[J]. Data Analysis and Knowledge Discovery, 2023, 7(4): 114.

[11] YUAN Y, ZHOU X, PAN S, et al. A Relation-specific Attention Network for Joint Entity and Relation Extraction[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 2020: 4054.

[12] 卢敏, 胡娟, 张先超, 等. 基于用户多特征融合的个性化推荐模型[J]. 电信科学, 2023, 39(5): 101.
LU Min, HU Juan, ZHANG Xianchao, et al. Personalized Recommendation Model Based on Users Multi-features Fusion[J]. Telecommunications Science, 2023, 39(5): 101.

[13] LI X, XU C, YANG G, et al. W2vv + + fully Deep Learning for Ad-hoc Video Search[C]//Proceedings of the 27th ACM International Conference on Multimedia, 2019: 1786.

[14] LI X, ZHOU F, XU C, et al. SEA: Sentence Encoder Assembly for Video Retrieval by Textual Queries[J]. IEEE Transactions on Multimedia, 2021, 23: 4351.

[15] 王嵘冰,闫晓楠,冯勇,等. 基于向量构建和点击预测的在线视频课推荐模型[J]. 辽宁大学学报(自然科学版), 2022, 49(3): 214.
WANG Rongbing, YAN Xiaonan, FENG Yong, et al. Online Video Courses Recommendation Model Based on Vector Construction and Click Prediction[J]. Journal of Liaoning University(Natural Science Edition), 2022, 49(3): 214.

[16] 杨攀攀. 基于 Flink 的电影推荐系统研究与设计[D]. 西安: 陕西科技大学, 2023.

[17] 童占北, 钟建伟, 李祯维, 等. 基于 VGG16 图像特征提取和 SVM 的电能质量扰动分类[J]. 电工电气, 2023(7): 7.
TONG Zhanbei, ZHONG Jianwei, LI Zhenwei, et al. Power Quality Disturbance Classification Based on VGG16 Image Feature Extraction and SVM[J]. Electrotechnics Electric, 2023(7): 7.

[18] 闫晓楠. 基于向量构建与点击预测的课程推荐系统研究与实现[D]. 沈阳:辽宁大学, 2021.

[19] 金楠, 王瑞琴, 陆悦聪. 基于艾宾浩斯遗忘曲线和注意力机制的推荐算法[J]. 电信科学, 2022, 38(10): 89.
JIN Nan, WANG Ruiqin, LU Yuecong. Ebbinghaus Forgetting Curve and Attention Mechanism Based Recommendation Algorithm [J]. Telecommunications Science, 2022, 38(10): 89.

[20] 陈辉, 白骏, 殷传涛, 等. 基于学习行为的 MOOC 用户持续学习预测框架[J]. 北京航空航天大学学报, 2023, 49(1): 74.
CHEN Hui, BAI Jun, YIN Chuantao, et al. Behavior Based MOOC User Dropout Predication Framework[J]. Journal of Beijing University of Aeronautics and Astronautics, 2023, 49(1): 74.

[21] 杨本臣, 叶洪宇, 孟祥福. 融合偏好传播的多任务推荐模型研究[J]. 软件导刊, 2024, 23(6): 9.
YANG Benchen, YE Hongyu, MENG Xiangfu. Research on Multi-task Recommendation Model Fused with Preference Propagation[J]. Software Guide, 2024, 23(6): 9.

[22] HE R, MCAULEY J. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016: 144.

[23] TANG J, DU X, HE X, et al. Adversarial Training Towards Robust Multimedia Recommender System [J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 32(5): 855.

(编辑:温泽宇)