

# 基于 RoBERTa 和图增强 Transformer 的序列推荐方法

王明虎, 石智奎\*, 苏佳, 张新生

(西安建筑科技大学管理学院, 陕西 西安 710055)

**摘要:** 自推荐系统出现以来, 有限的信息就一直制约着推荐算法的进一步发展。为降低数据稀疏性的影响, 增强非评分数据的利用率, 基于神经网络的文本推荐模型相继被提出, 但主流的卷积或循环神经网络在文本语义理解和长距离关系捕捉方面存在明显劣势。为了更好地挖掘用户与商品之间的深层潜在特征, 进一步提高推荐质量, 提出一种基于 RoBERTa 和图增强 Transformer 的序列推荐(RGT)模型。引入评论文本数据, 首先利用预训练的 RoBERTa 模型捕获评论文本中的字词语义特征, 初步建模用户的个性化兴趣, 然后根据用户与商品的历史交互信息, 构建具有时序特性的商品关联图注意力机制网络模型, 通过图增强 Transformer 的方法将图模型学习到的各个商品的特征表示以序列的形式输入 Transformer 编码层, 最后将得到的输出向量与之前捕获的语义表征以及计算得到的商品关联图的全图表征输入全连接层, 以捕获用户全局的兴趣偏好, 实现用户对商品的预测评分。在 3 组真实亚马逊公开数据集上的实验结果表明, 与 DeepFM、ConvMF 等经典文本推荐模型相比, RGT 模型在均方根误差(RMSE)和平均绝对误差(MAE)2 种指标上有显著提升, 相较于最优对比模型最高分别提升 4.7% 和 5.3%。

**关键词:** 推荐算法; 评论文本; RoBERTa 模型; 图注意力机制; Transformer 机制

中图分类号: TP18

文献标志码: A

DOI: 10.19678/j.issn.1000-3428.0068307

## Sequence Recommendation Method Based on RoBERTa and Graph-Enhanced Transformer

WANG Minghu, SHI Zhikui\*, SU Jia, ZHANG Xinsheng

(College of Management, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China)

**【Abstract】** Since the emergence of recommendation systems, further development of recommendation algorithms has been constrained by limited data. To reduce the impact of data sparsity and enhance the utilization of nonrated data, text-recommendation models based on neural networks have been successively proposed. However, mainstream convolutional and recurrent neural networks have clear disadvantages as concerns text semantic understanding and capturing long-distance relationships. To better explore the deep latent features between users and items, and further improve the quality of recommendations, a sequence recommendation method based on RoBERTa and a Graph-enhanced Transformer (RGT) is proposed. This model incorporates textual comment data by first utilizing a pre-trained RoBERTa model to capture the semantic features of words in the comment text, thereby modeling the personalized interests of the user. Subsequently, based on historical interaction information between users and items, a graph attention mechanism network model with the temporal characteristics of item associations is constructed. Using the graph-enhanced Transformer method, the feature representations of various items learned by the graph model are sequentially input to the Transformer encoding layer. Finally, the obtained output vectors, along with the previously captured semantic and computed global representations of the item association graph, are input into a fully connected layer to capture the global interest preferences of the user and achieve prediction ratings for items. The experimental results on three groups of real Amazon public datasets demonstrate that the proposed recommendation model significantly improves the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) compared to several existing classical text recommendation models, such as DeepFM and ConvMF. Compared to the optimal comparison model, the highest increases are 4.7% and 5.3%, respectively.

**【Key words】** recommendation algorithm; review text; RoBERTa model; graph attention mechanism; Transformer mechanism

收稿日期: 2023-08-29 修回日期: 2023-11-30

基金项目: 陕西省科技厅重点产业创新链(群)-工业领域项目(2022ZDLGY16-04); 陕西省教育厅哲学社会科学重点项目(21JZ035); 西安建筑科技大学自然科学基金专项(1609720032)。

通信作者 E-mail: \*szk1357234291@xauat.edu.cn

## 0 引言

随着互联网的飞速普及和全球数字化工程的不断推进,互联网用户量呈指数级增长,导致网络信息的过载问题愈发突出。推荐系统作为缓解大数据时代用户获取信息压力的重要手段,已经在人们日常生活的各个领域得到广泛应用,如在多媒体平台(抖音、快手)、个性化广告(朋友圈广告)以及网络购物(亚马逊、淘宝)等场景中均发挥着重要作用。推荐算法作为推荐系统的核心组成部分,不仅直接影响用户的使用体验,还代表着互联网企业的重要竞争力,因此,提高推荐算法的推荐精度、增强系统的个性化推荐效率,对提高用户满意度和增加企业营收都具有十分重要的现实意义。

当前主流的推荐系统大多都是基于协同过滤的推荐算法,其主要思想是通过计算用户或者商品之间的相似度来预测用户对商品的喜好程度,根据相似度的高低依次进行推荐。这种推荐方法过度依赖用户对商品的历史评分,然而现实中用户对商品的评分数据是极度稀少的。为解决该问题,基于矩阵分解(MF)<sup>[1]</sup>的协同过滤推荐方法相继被提出,如文献[2]将改进的交替最小二乘法(ALS)与在线学习相结合提高推荐效率,文献[3]利用奇异值分解(SVD)技术对降维后的评分矩阵进行数据填充并完成聚类推荐。尽管这些方法能够有效增强模型对稀疏数据的处理能力,但是在实际应用中并不能达到理想的性能提升,数据信息的贫乏仍然在很大程度上限制着推荐模型的预测效果。

近年来,随着深度学习在自然语言处理领域的广泛应用,不少学者将文本信息引入推荐系统,通过挖掘文本数据获取用户与物品的隐藏特征,利用卷积神经网络(CNN)<sup>[4]</sup>和注意力机制<sup>[5]</sup>等方法处理属性信息或文本评论,增强模型学习效果。单从性能提升上来看,上述方法均有着良好表现,但也存在一定的局限性,无法捕获不同用户或物品之间复杂的相互关系。而图神经网络(GNN)<sup>[6]</sup>可以将用户或项目的相互关系表示成图,通过聚合邻居节点的不同信息提高模型的表示能力,然而 GNN 模型也存在无法处理用户行为序列的问题,忽略了不同项目在用户行为序列中的重要性。

针对上述问题,本文通过对比学习不同推荐方法,提出一种基于 RoBERTa 和图增强 Transformer 的序列推荐(RGT)方法。本文主要工作如下:

1)引入评论文本,采用预训练的 RoBERTa 模型深度挖掘文本信息的潜在特征,捕获语义特征,增

强模型的学习表示能力。

2)构建商品关联有向图,利用图注意力网络计算不同商品之间的权重信息,获取不同商品之间存在的关联表征。

3)设计图增强 Transformer 模型,从用户的行为信息出发,利用商品关联有向图构建用户对应的商品序列,并输入 Transformer 编码层获取用户的行为特征表示,最后融合语义特征和商品关联特征,优化模型推荐效率。

## 1 相关工作

### 1.1 基于主题建模的推荐方法

评论文本最初被应用于推荐领域时常用的建模方法大都基于主题模型,基于文本数据的主题建模方法通过提取文本数据中的主题信息,推断出隐藏的主题分布和文本分布,揭示潜在主题,从而实现主题推荐功能。隐含狄利克雷分布(LDA)<sup>[7]</sup>模型作为最具代表性的文本主题构建模型,不仅可以实现语义分析,还能利用推测的主题分布获取用户和项目的相关信息,完成隐含主题提取。近年来,随着 LDA 模型影响力的不断提高,基于 LDA 的主题模型相继被提出。文献[8]利用 LDA 模型获取微博用户的点赞文本主题,并基于该主题对微博新用户进行特征预测;文献[9]将 Word2Vec 技术引入 LDA 模型,利用 K-Means 算法对融合的主题分布特征和词向量特征进行主题聚类,有效挖掘微博文本的语义信息;文献[10]将 LDA 主题模型与机器学习算法相结合,通过 LDA 模型量化文本数据,然后基于支持向量机(SVM)构建商品评分与用户期望评分的匹配机制,实现商品的个性化推荐服务。

为更精确地挖掘潜在兴趣主题,部分学者将深度学习方法与 LDA 模型结合起来,如:文献[11]通过融合 LDA 模型与 CNN 网络分别处理项目评论数据,将获取到的不同层次的主题信息和文本信息经过特征融合输入基于概率的矩阵分解(PMF)推荐模型,以进行评分预测;文献[12]在 LDA 模型的基础上引入 HowNet 增强语义理解,并利用构建的注意力模型将实体和语义部分注入网络层,从而优化特征词的重要性同时提高推荐精度。

在上述研究工作中,大部分都是采用概率主题模型来挖掘评论文本信息以提取用户偏好和商品潜在属性,但是主题建模通常都是基于词袋模型来记录文本信息,这种方法往往仅能提取到全局水平的文本语义信息,无法有效保留评论文本的上下文词

序信息,从而破坏了数据的完整性。另外,从数据利用率的角度来看,主题模型不具备深度提取文本数据中非线性潜在因子的能力,不能最大化地提高文本数据的吸收效率。

## 1.2 基于深度学习的推荐方法

深度学习的快速发展降低了非结构化数据的处理难度,相比于主题建模,深度学习方法可以更好、更准确地构建文本推荐模型。常见的深度学习方法包括自编码器、神经网络、多层感知机等,近年来,越来越多的学者将它们应用到推荐系统领域。文献[13]利用长短时记忆(LSTM)网络对自编码器模型进行改进,在保留数据时序性的同时将其应用于课程推荐任务中,取得了良好的推荐效果;文献[14]提出一种混合推荐模型 DAAI,其利用降噪自编码器和 DNN、CNN 等方式分别提取评分矩阵和属性文本信息中的潜在特征,并通过多层感知机融合不同特征以计算预测评分;文献[15]提出的 DeepRec 推荐模型通过 CNN 网络分别学习用户和项目的属性文本集,并利用多层感知机在处理评分矩阵获取用户偏好的同时融合不同数据特征进行最终推荐。

为了进一步提高推荐效率,加强文本信息的提取质量,研究人员将注意力机制引入深度学习模型。文献[16]通过结合两组 CNN 网络和注意力机制,捕捉上下文语义特征和商品与用户之间的交互关系,计算用户对商品的动态偏好;文献[17]通过向卷积神经网络注入自注意力机制的方式提取新闻文本特征矩阵,然后对用户浏览数据添加时序预测,并利用多头自注意力挖掘用户兴趣特征。除此之外,一些基于深度学习模型的混合推荐方法也被提出,如文献[18]提出的 HRS-DC 模型和文献[19]提出的 DCFM 模型,两者均是通过深度学习技术与传统协同过滤或因子分解等算法相结合而实现的评分预测。另外,一些基于图模型<sup>[20-21]</sup>和知识图谱<sup>[22-23]</sup>的推荐方法近年来也被不断提出,将商品或用户作为模型中的实体或节点,构建异质信息网络用以表示用户或项目之间的关联特征,不仅能够提高推荐准确度,还可以增强系统可解释性。

上述研究中所应用的深度学习技术已经趋于成熟,且推荐效果也有显著提升,但是仍存在一定的局限性,如卷积神经网络无法对文本信息进行更细粒度的高阶特征融合、注意力机制无法捕捉位置信息等。不同于上述基于深度学习的推荐方法,本文从文本语义、商品关联关系和用户行为 3 个方面出发,对商品推荐场景进行建模优化。具体地,首先利用

RoBERTa 模型在使用更大 Batch size 的基础上增加训练数据,解决传统静态词向量存在的一词多义问题,同时增强词的语义表征能力;然后借助图模型的优良特性,将与用户产生过交互行为的商品构建成具有关联关系的有向图,计算商品关联图模型的全图表征;最后将图模型提取到的每个商品特征有序地输入 Transformer 编码层,增强用户的行为特征表达能力,通过多维度的特征融合进一步提升模型的推荐能力。

## 2 基于 RoBERTa 和图增强 Transformer 的序列推荐模型

### 2.1 模型结构

本文基于 RoBERTa 和图增强 Transformer 的序列推荐模型 RGT 结构如图 1 所示,该模型主要分为两部分:第 1 部分利用用户评论文本集,通过预训练的 RoBERTa 模型挖掘用户对商品的兴趣偏好,提取语义表征;第 2 部分利用用户的历史行为数据,首先构建具有时序特征和关联关系的商品有向图结构,通过图注意力网络捕获所有商品的图表征,然后将用户行为以序列的形式输入 Transformer 编码层,利用图增强 Transformer 的学习方式获取用户行为中的个性化兴趣,最后将计算得到的用户行为商品关联图的全图表征、语义表征以及用户行为序列特征统一输入全连接层,计算用户对商品的评分。

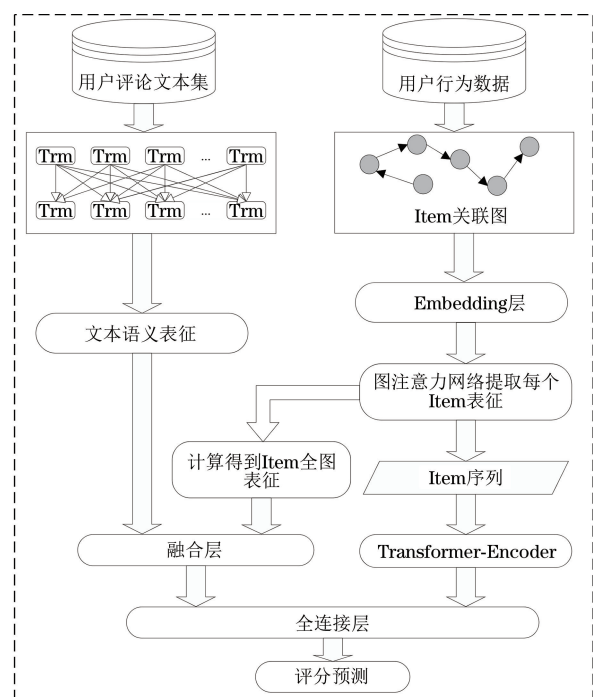


图 1 RGT 模型框架

Fig.1 RGT model framework

## 2.2 输入特征表示

### 2.2.1 基于预训练的 RoBERTa 评论文本语义特征表示

RoBERTa 是由 BERT 模型改进而来,改进方向包括以下 3 个方面:首先为了增加输入文本序列的长度而扩大了 BERT 模型的 Batch size 参数,在使用更丰富语料库的同时增加了文本数据的训练步长;其次削减了 NSP(Next Sentence Prediction)任务,降低了其对模型性能的影响;最后调整了 Masking 机制,采取动态 Masking 改进训练数据的生成方式,提高了模型的预训练效率。预训练主要有以下 2 个目的:一是为了让模型在庞大的复杂数据输入时逐渐适应不同的动态掩码策略,保障模型与数据的高度耦合;二是为了丰富模型的语义信息,使模型全方位地学习输入层的词嵌入向量。

本文通过预训练的 RoBERTa 模型将输入的用户评论语料进行嵌入,从文本数据中捕获用户的个性化兴趣。首先通过叠加用户评论文本  $T = [t_0, t_1, \dots, t_{n-1}]$  中的字符向量  $w_i$ 、分句向量  $c_i$  和位置向量  $P_{pos,i}$ ,得到文本  $T$  中的第  $i$  个字词表示,将其标记为嵌入向量  $S_{seq,i}$ ,即  $S_{seq,i} = w_i + c_i + P_{pos,i}$ ;然后将叠加向量  $S_{seq,i}$  输入 RoBERTa 模型,  $S_{seq,i}$  向量经多层 Transformer 编码器单元处理后,最后输出包含整个文本  $T$  中所有字词语义信息的动态词向量  $V_i$ ,其计算过程的数学表示为:

$$V_i = \text{RoBERTa}(w_i + c_i + P_{pos,i})$$

$$i \in [0, n-1] \quad (1)$$

其中:  $V_i$  为 RoBERTa 模型处理后输出的文本动态词向量表示,  $V_i = (V_0, V_1, V_2, \dots, V_{i-1})$ 。图 2 所示的结构即为一个简单的 RoBERTa 模型。

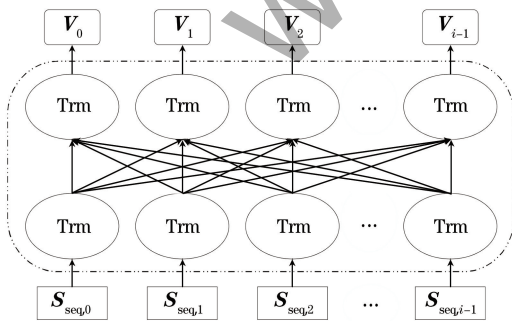


图 2 RoBERTa 模型结构

Fig.2 RoBERTa model structure

### 2.2.2 基于图增强 Transformer 的商品关联特征和用户行为特征表示

#### 1) 商品关联图构建。

用户在购买商品时通常会与同类商品进行多次比较,在比较浏览的过程中会留下交互信息,如

用户的点击行为。同一个用户通常会有次序地点击浏览多种商品,这些被点击浏览过的商品之间往往也会存在某些潜在的关联关系,如果仅利用商品的语义信息对其进行关联匹配,则无法较好地建模商品之间的关联关系。因此,为更好地捕获不同商品之间的关联关系,本文根据用户的历史行为,构建用户与商品的次序交互图以对商品数据进行嵌入表示,通过图模型学习并提取不同商品之间隐含的关联特征。

如图 3 所示,与  $user_2$  有过交互行为的所有商品(item)被构建为一个图结构,将图中所有 item 标记为图节点,将节点与节点之间通过潜在关联关系相互连接的线记为边。本文用  $G_i = \{x_i, \epsilon_i\}$  表示用户  $i$  对应的所有商品的关联图,其中,  $x_i$  表示商品节点,  $\epsilon_i$  表示节点通过边与邻接节点组成的三元关系集  $(x_h, r, x_t)$ ,  $r$  为连接节点  $x_h$  与节点  $x_t$  的边。

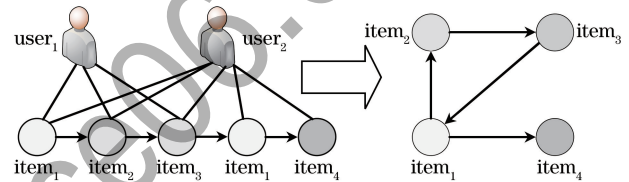


图 3 商品关联有向图

Fig.3 Commodity related directed graph

#### 2) 商品(节点)嵌入与特征表示。

在商品关联图构建完成后,把每个商品 ID 作为嵌入层输入,将其映射到统一的嵌入空间,得到每个 item 的初始嵌入表示。为获取图中的商品(节点)特征表示,本文通过图注意力网络对上述商品关联图进行节点聚合,图注意力网络不依赖图的结构,有效利用了图中商品的自身信息,不仅可以处理有向图,还可以为图中商品的每个邻居分配不同的权重。

假设当前处于商品关联图注意力网络的第  $l$  层,对于输入的商品关联图  $G_i$ ,商品用  $x_h$  表示,  $x_h$  的邻接商品用  $N_h = \{x_t | x_h, r, x_t\}$  (其中  $N_h$  包含  $x_h$ ) 表示。根据图注意力网络的计算规则,首先计算商品  $x_h$  的邻接商品  $x_t$  的权重:

$$\alpha^l(x_h, r, x_t) = \frac{e^{T^l(x_h, r, x_t)}}{\sum_{\tilde{x}_t \in N_h} e^{T^l(x_h, \tilde{r}, \tilde{x}_t)}} \quad (2)$$

$$T^l(x_h, r, x_t) = \sigma[(x_h^{l-1} W_1^l)(x_t^{l-1} W_1^l + e_r^{l-1} W_2^l)] \quad (3)$$

其中:  $\sigma$  为 LeakyReLU 激活函数;  $W_1^l$  和  $W_2^l$  分别为对应商品与连接边的转换权重矩阵;  $x_h^{l-1}$ 、 $x_t^{l-1}$ 、 $e_r^{l-1}$  分别为  $x_h$ 、 $x_t$ 、 $r$  在第  $l-1$  层的向量表示。

权重用以表示相邻节点的重要程度,在求得商品  $x_t$  对其邻接商品  $x_h$  的权重后,根据其权重计算第  $l$  层商品  $x_h$  的向量表示,计算方式如下:

$$\mathbf{x}_h^l = \text{Tanh} \left( \sum_{\tilde{x}_t \in N_h} \alpha^l(x_h, r, x_t) \mathbf{x}_t^{l-1} \mathbf{W}_1^l \right) \quad (4)$$

其中:  $\text{Tanh}$  为激活函数。

假设商品关联图注意力网络的堆叠层数为  $L$ , 可得到对应  $L$  个商品  $x_h$  的特征表示, 记为  $\{\mathbf{x}_h^1, \mathbf{x}_h^2, \dots, \mathbf{x}_h^L\}$ 。

3) 商品关联图的全图特征表示。

在计算完图  $G_i$  中所有商品节点在第  $l$  层的向量表征后, 需要聚合所有商品的向量表示从而得到整个商品关联图的全图表征, 用来建模用户的个性化表示, 此时, 根据用户  $u$  和项目  $i$  的信息, 利用图注意力机制优先计算所有商品的交互水平权重。假设对上述商品关联图注意力网络第  $l$  层的商品表征进行聚合, 则商品  $x_h$  的权重为:

$$\beta_h^l = \frac{e^{T^l(x_h, u, i)}}{\sum_{\tilde{x}_h \in x_u} e^{T^l(x_h, r, \tilde{x}_i)}} \quad (5)$$

$$T^l(x_h, u, i) = \sigma(\mathbf{x}_h^{l+1} (\text{Tanh}([\mathbf{e}_u \parallel \mathbf{e}_i] \mathbf{W}_3^l))^\top) \quad (6)$$

其中:  $\mathbf{W}_3^l$  为转换权重矩阵;  $\sigma$  表示激活函数; 符号  $\parallel$  代表向量拼接。

在已知商品关联图  $G_i$  中所有商品的向量表征和交互水平权重后, 通过权重和向量表征的加权相加计算出商品关联图  $G_i$  在第  $l$  层的输出表征  $\mathbf{g}_i^{l+1}$ , 计算方式如下:

$$\mathbf{g}_i^{l+1} = \sum_{x_h \in x_u} \beta_h^l \mathbf{x}_h^{l+1} \quad (7)$$

其中:  $x_u$  为用户行为序列, 包含所有交互商品。

最后, 对  $l$  层堆叠的商品关联图注意力网络中的每一层图的输出表征进行拼接, 得到商品关联图的全图最终表示  $\mathbf{G}_i$ :

$$\mathbf{G}_i = \mathbf{g}_i^1 \parallel \mathbf{g}_i^2 \parallel \dots \parallel \mathbf{g}_i^{l+1} \quad (8)$$

上述过程将用户的行为数据以商品关联有向图的形式进行表示, 当商品的初始嵌入向量输入图模型时, 图注意力机制通过学习并生成邻居节点对目标节点的不同注意力权重, 利用加权融合的思想实现目标节点的特征聚合, 在增强模型可解释性的同时使得用户行为数据的嵌入过程更加合理。

4) 基于图增强 Transformer 的用户行为特征表示。

在学习得到商品关联图的每个商品(节点)表征和商品关联图的全图表征后, 考虑到用户行为数据

中的时间因素会影响点击行为, 然而上述嵌入模式都忽略了时间的影响, 为解决时间序列问题, 本文在构建有向商品关联图模型时, 根据用户与商品的交互信息, 从历史数据中获取商品的点击序列, 并将上述图注意力网络提取到的所有的商品表征  $x_h$  以序列的形式嵌入到 Transformer 编码层, 构建基于图注意力网络增强的 Transformer 模型, 获取用户的时序行为特征表示。基于图注意力网络增强的 Transformer 模型结构如图 4 所示。

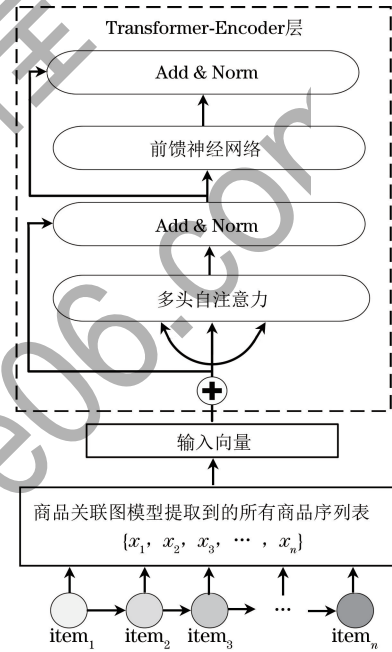


图 4 基于图注意力网络增强的 Transformer 模型

Fig.4 Enhanced Transformer model based on graph attention network

为了学习到更加丰富的节点表征, 本文使用具有多头自注意力机制的 Transformer-Encoder, 对输入序列中图模型提取到的所有商品序列表征  $x_i x_u$  进行编码。Transformer-Encoder 主要由多头自注意力机制和前馈神经网络组成, “Add & Norm”作为标准化层, 通常紧跟在多头自注意力机制和前馈神经网络层之后, 主要作用是简化输入向量在模型不同网络层之间的传播过程, 具体如下:

1) 多头自注意力机制的网络结构如图 5 所示, 该部分主要利用多个 Self-Attention 模块对输入向量进行并发操作, 对多个子空间设置不同的参数, 采用相同的计算方式, 通过多层并行计算, 使得输入向量中每个字符的隐状态都包含该向量中其他字符的所有信息, 从而捕捉到多个不同空间下的多种数据表征, 更丰富的特征信息可以有效避免过拟合。

对于输入的所有商品序列表征  $x_i$ , 多头自注意力机制通过线性变换, 首先分别计算查询矩阵  $\mathbf{Q}_i$ 、

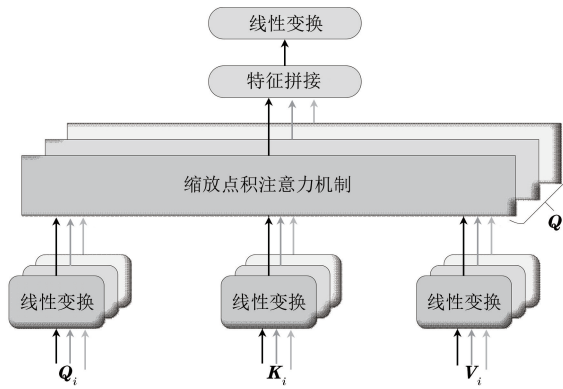


图 5 多头自注意力机制网络结构

Fig.5 Multi-head self-attention mechanism network structure

键矩阵  $K_i$  和值矩阵  $V_i$ ，将  $x_i$  映射到不同的子空间，计算公式为：

$$Q_i = x_i W_i^Q, K_i = x_i W_i^K, V_i = x_i W_i^V \quad (9)$$

其中： $W_i^Q$ 、 $W_i^K$ 、 $W_i^V$  分别为对应矩阵的权重矩阵。

然后根据  $Q_i$ 、 $K_i$ 、 $V_i$  矩阵计算多头注意力的输出并进行拼接操作：

$$H_{\text{Head},i} = \text{Attention}(Q_i, K_i, V_i) = \text{Softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \quad (10)$$

$$\text{Mul}(H_{\text{Head},i}) = \text{Concat}(H_{\text{Head},1}, H_{\text{Head},2}, \dots, H_{\text{Head},i}) W^0 \quad (11)$$

其中： $H_{\text{Head},i}$  表示第  $i$  个注意力头； $d_k$  表示输入商品序列的向量维度；Concat 表示拼接操作； $W^0$  为附加的权重矩阵。

2) Add & Norm 为 LayerNorm 和残差连接。LayerNorm 可以标准化每一层的向量输入，将其转化为归一化均值方差，残差部分可以在反向传播过程中将输入的多头注意力结果进行数据叠加，减少信息损失同时避免梯度消失，计算公式为：

$$O = \text{LayerNorm}(\text{Mul}(H_{\text{Head},i}) + H_{\text{Head},i}) \quad (12)$$

3) 前馈神经网络包含 2 个线性全连接网络层和 1 个非线性激活函数，用以接收上游模块传入的特征向量并计算生成对应的向量矩阵，具体表示为：

$$H = W_2(\text{ReLU}(W_1 O + c_1)) + c_2 \quad (13)$$

其中： $W_1$  和  $W_2$  分别表示 2 个全连接层的权值矩阵； $c_1$  与  $c_2$  分别对应 2 个全连接层的偏置项；ReLU 为激活函数。

值得注意的是，Transformer-Encoder 模型为了区分输入特征的位置信息，还为输入向量添加了位置编码，本文使用正弦函数对输入序列中每个商品的特征表示进行位置编码，进一步增强 RGT 推荐模型对时序信息的敏感性。所有商品的序列特征

$x_h$  经  $M$  层 Transformer-Encoder 后得到用户的最终行为特征表示  $F_i = H + P_i$ ，其中， $P_i$  为用户序列  $x_u$  中各个商品的表征  $x_i$  在序列中所处位置的编码。

### 2.3 特征融合与评分预测

本文使用全连接层对上述提取到的不同输入特征进行特征融合和评分预测，对文本语义表征  $V_i$ 、商品关联图的全图表征  $G_i$  以及用户的最终行为表征  $F_i$  进行拼接操作，得到向量  $Z = \text{Concat}(V_i, G_i, F_i)$ ，再将其输入预测层，最终得到用户  $u$  对商品  $i$  的预测评分  $\hat{Y}_{(u,i)}$ ：

$$\hat{Y}_{(u,i)} = ZW^T + b_u + b_i + \eta \quad (14)$$

其中： $b_u$ 、 $b_i$  和  $\eta$  分别对应用户偏倚项、商品偏倚项和全局偏倚项； $W^T$  代表全连接层的权重参数。

### 2.4 模型优化

在模型学习过程中，预测评分  $\hat{Y}_{(u,i)}$  与真实评分  $Y_{(u,i)}$  之间的误差无法避免，为加快模型收敛，使得  $Y_{(u,i)}$  与  $\hat{Y}_{(u,i)}$  之间的差值最小化，本文给出如下目标损失函数：

$$L_{\min} = \sum_{u,i \in B} (\hat{Y}_{(u,i)} - Y_{(u,i)})^2 + \lambda \|\varphi\|^2 \quad (15)$$

其中： $B$  表示训练集中的样本集合； $\lambda$  为正则化系数； $\varphi$  为整个模型的所有可训练参数。为优化目标损失函数，本文引入 Adam<sup>[24]</sup> 作为优化器来最小化  $L_{\min}$  的值。

### 2.5 算法流程

RGT 推荐算法的详细流程如算法 1 所示。

#### 算法 1 RGT 推荐算法

输入 用户评论文本集，用户历史行为数据，用户对商品的真实评分

输出 用户  $u$  对商品  $i$  的预测评分  $\hat{Y}_{(u,i)}$

1. 初始化所有参数
2. 设置模型训练迭代次数
3. FOR 用户  $u$ 、商品  $i$ 、 $u$  对  $i$  的真实评分  $Y_{(u,i)}$
4. 根据式(1)计算评论文本  $T$  的语义特征  $V_i$
5. 构建具有用户行为次序的商品关联图  $G_i$
6. 根据式(2)~式(13)计算商品关联特征  $G_i$  和用户行为特征  $F_i$
7. 拼接  $V_i$ 、 $G_i$  和  $F_i$  得到最终的特征向量  $Z$
8. 根据式(14)预测评分  $\hat{Y}_{(u,i)}$  并输出
9. 使用梯度下降法 Adam 训练 RGT 模型并进行参数更新
10. END FOR

### 2.6 模型时间复杂度

假设每个用户的评论文本序列长度为  $n$ ，将其映射到隐藏层维度为  $d$  的特征向量，则提取用户评论文本特征的 RoBERTa 模型时间复杂度为  $O(dn^2)$ 。设图注意力机制网络中的图节点数量为

$f$ , 每个节点的特征维度为  $e$ , 图堆叠层数为  $L$ , 则基于商品关联图注意力机制网络的时间复杂度为  $O(Lef^2)$ 。Transformer-Encoder 需要对由语义特征和图节点特征组合成的复合高阶特征进行处理, 假设每个 Transformer 编码器的隐藏层维度为  $c$ , 注意力机制的头数为  $Q$ , 那么每个 Transformer 编码器的时间复杂度为  $O(Qfc(d+e))$ ,  $M$  层 Transformer 编码器总时间复杂度为  $O(MQfc(d+e))$ , 输出维度为  $k$  的全连接层需要对输出的复合高阶特征进行处理, 其时间复杂度为  $O(fk(d+e))$ 。综上, RGT 模型的总时间复杂度为  $O(dn^2 + Lef^2 + MQfc(d+e) + fk(d+e))$ 。

### 3 实验结果与分析

#### 3.1 实验设置

##### 3.1.1 实验环境与数据来源

为验证本文推荐模型的有效性, 实验采用亚马逊评论公开数据集 Amazon 5-score, 该数据集含有多个子类数据集, 是推荐系统领域最常用的公开数据集, 具备一定的权威性。实验选取 Instant-Video(视频)、Digital-Music(音乐)以及 Toys and Games(玩具游戏) 3 组子数据集用来模拟现实生活中的常见推荐场景, 3 组数据集均包含用户 ID、商品 ID、用户对商品的评分(1~5 分)以及用户对商品的评论文本这 4 种特征, 均属于大型数据集, 具体统计信息如表 1 所示。

表 1 3 组数据集的基本信息

Table 1 Basic information of three datasets

| 数据集            | 用户数    | 商品数    | 样本数     | 稀疏度/<br>% | 每个用户平均<br>发布评论量 |
|----------------|--------|--------|---------|-----------|-----------------|
| Instant-Video  | 5 130  | 1 685  | 37 126  | 99.57     | 7.23            |
| Digital-Music  | 5 541  | 3 568  | 64 706  | 99.67     | 11.68           |
| Toys and Games | 19 412 | 11 924 | 167 597 | 99.93     | 8.63            |

虽然上述 3 组数据集都包含庞大的用户基数和商品样本, 但是数据稀疏度均达到了 99% 以上, 这就说明用户与商品之间产生过交互行为的数据只占极小的一部分, 3 组数据集都属于高度稀疏数据。本文主要利用评论文本数据缓解推荐中存在的稀疏问题, 3 组数据集中的评论文本数量均达到了上万条, 且每个用户发布的平均评论长度均在 7~11 个字词范围内, 由此可见评论文本中所携带的语义信息足以体现用户偏好和商品特征, 丰富的文本数据足够指导建模。另外, 本次实验在 Windows 系统下的 Pycharm 平台上通过 Python 编程完成, Python 版本和 PyTorch 版本分别为 3.8 和 1.12.1。

##### 3.1.2 实验衡量指标

本文将以上 3 组数据集按照 8:1:1 的比例划分为训练集、验证集和测试集, 并分别在验证集和测试集上进行超参数选取和模型性能评估。为评估本文所提模型的预测准确度, 实验均采用均方根误差 (RMSE) 和平均绝对误差 (MAE) 作为评价模型性能的衡量指标, RMSE 和 MAE 的取值越小, 代表模型的预测精度越高, 性能越强。假设用户  $u$  对商品  $i$  的预测评分和真实评分分别为  $\hat{Y}_{u,i}$  和  $Y_{u,i}$ , 定义 RMSE 和 MAE 计算方式如下:

$$R_{\text{RMSE}} = \sqrt{\frac{1}{N} \sum_{(u,i)} (\hat{Y}_{u,i} - Y_{u,i})^2} \quad (16)$$

$$M_{\text{MAE}} = \frac{1}{N} \sum_{(u,i)} |\hat{Y}_{u,i} - Y_{u,i}| \quad (17)$$

其中:  $(u, i) \in N$ ,  $N$  表示测试样本的总个数。

##### 3.1.3 实验超参数设置

在实验开始前, 为最大可能地保证模型结果的优良性, 需要对模型的相关参数进行预设置。由于 RGT 推荐模型涉及的超参数众多, 参数取值的不同会直接影响模型性能, 因此实验利用网格搜索算法寻找最优参数, 具体为: 学习率  $l_r$  和正则化系数  $\lambda$  分别在  $\{0.0005, 0.001, 0.005\}$  和  $\{0.01, 0.1, 0.5, 1\}$  范围内选取; 图注意力网络堆叠层数  $L$  和 Transformer 层数  $M$  分别在  $\{1, 2, 3, 4\}$  和  $\{1, 2, 3, 4, 5\}$  中选取。考虑到本次实验数据样本较大, 计算平台资源耗费较高, 设置 Batch size 值为 16, 多头自注意力头数为 8, 文本嵌入维度和图嵌入维度分别为 64 和 128, 预训练迭代次数 Epoch 为 10。为防止实验过程中可能出现的过拟合现象, 本文加入 Dropout 指标, 初始化为 0.5。

#### 3.2 性能比较与结果分析

本次实验选择若干模型进行测试与性能对比, 通过对比模型主要验证以下两点: 1) 利用评论文本数据是否可以降低评分数据稀疏性给推荐结果带来的影响; 2) 相比于其他利用文本数据进行推荐的模型, 融合不同类别的数据特征是否可以减小预测评分的误差。对比模型具体如下:

1) MF<sup>[1]</sup>: 标准的矩阵分解模型, 推荐系统中运用最广泛的协同过滤模型。

2) DeepFM<sup>[25]</sup>: 经典的深度神经网络模型, 共享输入特征, 利用因子分解机 (FM) 和神经网络分别捕捉低阶和高阶特征, 性能良好。

3) ConvMF<sup>[26]</sup>: 基于高斯分布, 利用 CNN 和 PMF 分别处理评论文本和评分数据, 能够保留文本词序同时预测性能也有一定提高。

4) DeepCoNN<sup>[27]</sup>: 深度协同神经网络模型, 利用 CNN 分别提取用户和商品评论文本集的语义特征, 性能优越, 是首个将用户评论和商品评论相结合来缓解数据稀疏性问题的深度学习模型。

5) NARRE<sup>[28]</sup>: 对 DeepCoNN 进行改进, 引入注意力机制消除无用评论干扰, 同时结合评分数据进一步提高预测评分的准确性。

6) IRIA<sup>[29]</sup>: 利用 CNN 提取文本特征, 采用注意力机制捕获用户与物品的交互信息, 增加门控层

合并特征向量并通过 FM 实现评分预测。

7) SAFMR<sup>[30]</sup>: 利用 CNN 提取评论特征, 将获取到的特征输入自注意力机制与因子分解机相结合, 神经网络以学习不同特征的动态权重, 降低了预测误差。

为保障对比实验的可靠性, 对比模型的实验数据划分与 RGT 模型保持一致, 相关参数根据对应论文进行初始化, 通过微调使其达到性能最佳, 所有实验同时在 3 组数据集上进行验证比较, 实验结果如表 2 所示, 最优结果加粗标注。

表 2 不同推荐模型的性能对比结果

Table 2 Performance comparison results of different recommendation models

| 模型       | RMSE          |               |                | MAE           |               |                |
|----------|---------------|---------------|----------------|---------------|---------------|----------------|
|          | Instant-Video | Digital-Music | Toys and Games | Instant-Video | Digital-Music | Toys and Games |
| MF       | 1.267         | 1.303         | 1.365          | 1.135         | 1.229         | 1.162          |
| DeepFM   | 1.199         | 1.084         | 1.109          | 0.742         | 0.698         | 0.715          |
| ConvMF   | 0.934         | 0.823         | 0.865          | 0.718         | 0.683         | 0.672          |
| DeepCoNN | 0.932         | 0.846         | 0.868          | 0.697         | 0.679         | 0.668          |
| NARRE    | 0.928         | 0.834         | 0.852          | 0.704         | 0.671         | 0.653          |
| IRIA     | 0.916         | 0.815         | 0.846          | 0.685         | 0.669         | 0.647          |
| SAFMR    | 0.907         | 0.803         | 0.837          | 0.679         | 0.658         | 0.641          |
| RGT      | <b>0.864</b>  | <b>0.791</b>  | <b>0.823</b>   | <b>0.676</b>  | <b>0.648</b>  | <b>0.607</b>   |

由表 2 可知, 本文 RGT 模型在 3 组不同数据集上的性能均优于其余几种对比模型, 分析原因如下:

1) 基于矩阵分解的模型 (MF、DeepFM) 主要价值在于解决矩阵的稀疏性问题, 同时实现了高阶特征提取, 但是该类模型将类别特征对应的稠密向量拼接作为输入, 然后对元素进行两两交叉, 过多的组合造成了不必要的冗余, 并未考虑用户行为数据中商品的交互顺序带来的影响。

2) ConvMF 模型同时使用了评论信息和评分数据, 而且还保留了词序信息, 但是该模型仅使用了项目的评论文本, 且 CNN 在捕捉文本相邻词之间的全局和非连续依赖关系时能力有限, 无法有效获取用户与商品在交互行为上的时间特征。

3) 基于语义的模型 (DeepCoNN) 采用 Word2Vec 作为初始嵌入, 对文本数据进行处理获取推荐结果, 作为首次将语义特征引入推荐算法的模型, 其取得了不错的效果, 但该模型无法保留评论文本的词序信息, 且忽略了商品之间可能存在的隐含关系。

4) 基于评论文本的推荐模型 (NARRE、IRIA、SAFMR) 与 DeepCoNN 模型的侧重点不同, 该类模型引入注意力机制, 能够有效识别无用评论, 动态捕捉不同特征的差异权重, 还可利用评分行为更精细

地捕捉辅助文本信息的语义特征, 但该类模型在提取文本特征时和 ConvMF 模型一样使用了 CNN, 该模型在复杂语言任务中无法深度理解词语的多义性, 特征提取效果较差, 且仍未考虑不同商品之间可能存在的关联关系。

5) 本文在充分学习已有模型的基础上, 结合数据特征实现了融合语义、物品关联关系以及用户历史行为的推荐架构, 将文本数据以语义的形式作为初始输入, 很好地对用户的兴趣偏好进行了个性化建模。此外, 为了消除商品语义信息相差过大而忽略商品关联特征的问题, 本文引入图神经网络模型, 通过将不同商品映射到图节点的方式构建商品间的关联关系, 同时考虑到用户行为对商品嵌入的时序影响, 构建基于图增强 Transformer 的表示学习方法, 把用户行为与商品序列相互对应, 将图模型中提取到的各个商品特征以序列的形式输入 Transformer 编码层, 对时间特征进行单独处理, 使模型充分挖掘到不同层次的向量表示, 相比其他模型在特征提取上更具优势, 模型评分预测的准确性更高。

### 3.3 实验验证

为了进一步验证上文实验所涉及的 3 个特征 (语义特征、商品关联特征、用户行为/商品时序特征) 对推荐任务的有效性, 本文设计消融实验以评估

以下 4 个 RGT 变种算法的推荐性能:

1) Graph+T 算法:从本文所提 RGT 算法中剔除语义特征,只考虑商品关联特征和商品时序特征。

2) RoBERTa+T 算法:从本文所提 RGT 算法中剔除基于图的商品关联特征,即只输入语义特征和商品时序特征。

3) Graph+FC 算法:从本文所提 RGT 算法中剔除语义特征和输入商品的时序特征,即不考虑特征融合,直接将图模型中提取到的商品关联特征输入全连接层进行评分预测。

4) RoBERTa+FC 算法:不做特征融合,从本文所提 RGT 算法中同时剔除商品关联特征和用户行为特征,直接利用 RoBERTa 模型中提取到的语义特征输入全连接层进行线性回归预测。

为保障对比实验的可靠性,上述 4 类变种算法仍然在 3 组数据集上进行实验,且对比算法的相关参数均根据本文的参数设置进行初始化,通过微调保证其达到最优性能,利用 RMSE 指标进行性能对比,实验结果如图 6 所示。

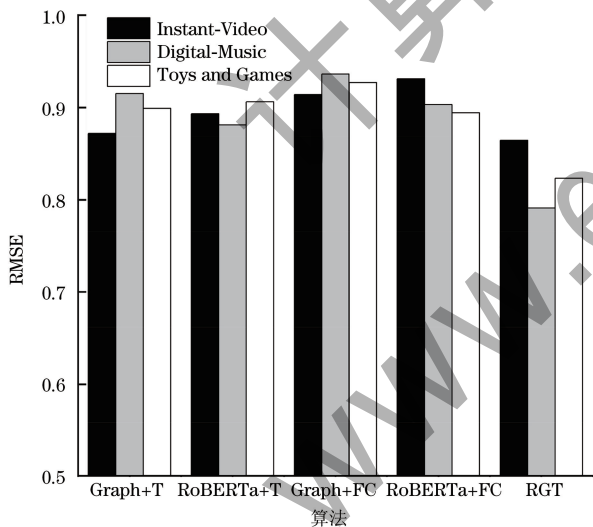


图 6 3 组数据集下的消融实验结果

Fig.6 Ablation experimental results under three datasets

从图 6 可以看出:本文 RGT 算法在 3 组数据集上的性能表现优于 4 类变种算法,说明在 RGT 算法中,利用评论文本提取到的语义特征和具有连接关系的商品关联特征对模型性能的提升均起到了积极作用。对于 4 类变种算法, RoBERTa+FC 的性能最差, Graph+FC 次之,验证了 RGT 算法从多角度和多方位提取并融合文本数据特征的有效性。

### 3.4 参数灵敏度分析

为探究超参数取值对模型性能的影响,本文设置 3 组参数进行实验,3 个超参数分别为图注意力

网络堆叠层数  $L$ 、Transformer 层数  $M$  以及注意力头数  $Q$ 。

在 Instant-Video 数据集上,将图注意力网络的堆叠层数  $L$  在  $\{1, 2, 3, 4\}$  范围内进行取值并实验,结果如图 7 所示(其余 2 组数据集上的实验结果与图 7 类似)。从图 7 可知,当网络堆叠层数为 3 时, RMSE 和 MAE 指标值最小,模型推荐性能最好。这是因为当网络堆叠层数过多时,图中节点学习到的特征表示具有极高的相似性,网络层数增加,不仅会造成模型复杂度增加,还会直接导致模型结果过拟合,对模型性能的提升起到反向抑制的效果,但如果网络堆叠层数太少,根据图神经网络的消息传递机制,图节点接收到的领域节点信息过少,致使图模型无法学习到更多有效特征,模型性能欠佳。

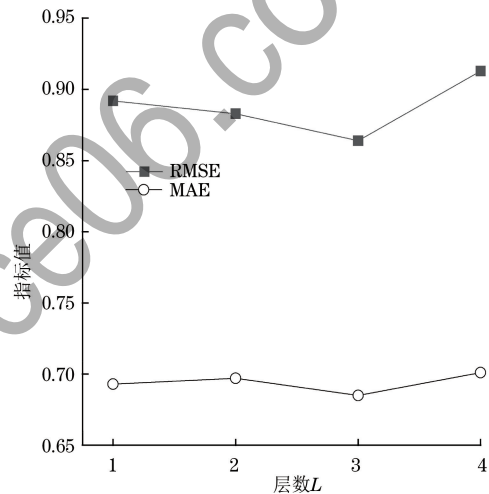


图 7 图注意力网络堆叠层数对模型性能的影响

Fig.7 The effect of the number of stacked layers of graph attention network on the performance of the model

在数据集 Instant-Video 上,将 Transformer 层数  $M$  在  $\{1, 2, 3, 4, 5, 6\}$  范围内进行取值并实验,结果如图 8 所示(其余 2 组数据集上的实验结果与图 8 类似)。从图 8 可知,当  $M=4$  时, RMSE 和 MAE 指标值均为最小,这表明此时模型复杂度、特征重用和注意力机制可能达到了一种平衡状态,使得模型性能达到了最优,而将层数调整至 5 层或 6 层时模型效果逐渐下降,这是因为在 Transformer 中,每一层都在某种程度上重用了前一层的特征,当增加额外的层数时,这些层并没有学习到新的有用特征,而且当模型层数过多时,还可能引入了一些不必要的噪声数据,从而影响模型性能。此外,从 RGT 模型的结构上看,最初输入 Transformer 编码层的特征是经过上层图模型增强提取过的特征,因此,过多的层数堆积会导致网络冗余,进而提高模型的复杂度,最终可能会引发过拟合现象。

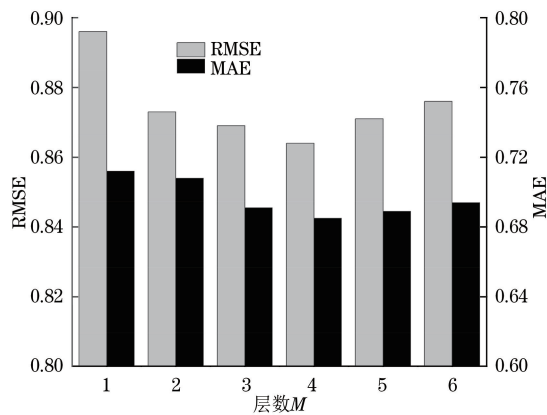


图 8 Transformer 编码器层数对模型性能的影响

Fig.8 The effect of the number of Transformer encoder layers on the performance of the model

为探究图中节点在不同潜在空间下的特征学习表示对模型推荐性能的影响,将注意力头数  $Q$  在  $\{1, 2, 4, 8, 16\}$  范围内进行取值并实验,结果如图 9 所示。从图 9 可知,当设置注意力头数为 4 时,模型的推荐性能在 Instant-Video 和 Digital-Music 数据集上表现出了最优效果,而在 Toys and Games 数据集中,设置  $Q = 8$  时模型性能表现最好,这可能是因为与 Instant-Video、Digital-Music 数据集相比,Toys and Games 数据集的固有信息特征表现出来的差异性更加明显。

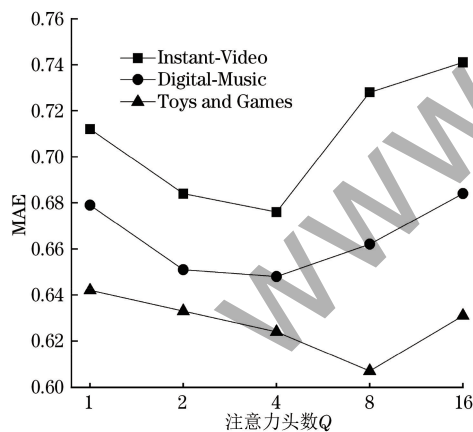


图 9 注意力头数对模型性能的影响

Fig.9 The effect of the number of attention heads on the performance of the model

进一步观察图 9 可以发现,当设置注意力头数为 1 或 16 时,模型性能表现均未达到理想状态,这表明图节点仅在单一的潜在空间下学习到的数据表征并不能最大程度地获取到足以代表该节点的特征表示,需要在多层潜在空间下进行更深层的挖掘,得到更丰富的特征表示,但过多的注意力头数也会造成模型复杂度提高,对性能表现起到反向作用。因此,在实验过程中,需要在注意力头数和模型性能之

间找到一个相对平衡点,以此保证模型效果达到最优。

## 4 结束语

评论文本数据作为一种辅助信息,蕴含着丰富的潜在信息,通过深度挖掘提取到与其相关的隐式特征并应用到推荐领域,能够很好地改善因评分矩阵数据稀疏性而导致的推荐精度下降问题。为了更好地挖掘用户与商品之间的深层潜在特征,进一步提高推荐质量,本文提出一种 RGT 推荐模型。实验结果表明,RGT 模型能够有效提取用户评论和用户历史行为数据的语义特征和商品关联特征,并较好地解决商品数据在时序上的先后关系,进一步降低评分预测的误差。但从数据的处理效率来看,本文构建的推荐模型主要利用评论文本数据完成推荐任务,模型时间复杂度较高,与只使用评分数据的推荐模型相比,本文模型所需的算力更高,执行效率较低。下一步将尝试引入更多非评分信息同时增强推荐模型的数据处理能力,进一步提高推荐算法的灵活性。

## 参考文献

- [1] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.
- [2] 邢玉莹,夏鸿斌,王涵. 缺失数据建模的改进型 ALS 在线推荐算法[J]. 计算机工程, 2018, 44(8): 212-217, 223.
- [3] XING Y Y, XIA H B, WANG H. Improved ALS online recommendation algorithm with missing data modeling[J]. Computer Engineering, 2018, 44(8): 212-217, 223. (in Chinese)
- [4] 纪成君,李蕊,王仕勤. 缺失数据下基于 SVDIFC 的协同过滤推荐算法[J]. 计算机应用研究, 2021, 38(10): 2994-2999.
- [5] JI C J, LI R, WANG S Q. Collaborative filtering recommendation algorithm based on SVDIFC under missing data[J]. Application Research of Computers, 2021, 38(10): 2994-2999. (in Chinese)
- [6] 周飞燕,金林鹏,董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.
- [7] ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251. (in Chinese)
- [8] 王永贵,尚庚. 融合注意力机制的深度协同过滤推荐算法[J]. 计算机工程与应用, 2019, 55(13): 8-14.
- [9] WANG Y G, SHANG G. Deep collaborative filtering recommendation with attention mechanism[J]. Computer Engineering and Applications, 2019, 55(13): 8-14. (in Chinese)
- [10] 吴静,谢辉,姜火文. 图神经网络推荐系统综述[J]. 计算机科学与探索, 2022, 16(10): 2249-2263.
- [11] WU J, XIE H, JIANG H W. Survey of graph neural network in recommendation system[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(10): 2249-2263. (in Chinese)
- [12] LEI D M, NG A Y, JORDAN M I. Latent Dirichlet allocation[J]. Journal of Machine Learning Research, 2003,

- 3: 993-1022.
- [8] 王雅静,郭强,邓春燕,等. 基于 LDA 主题模型的用户特征预测研究[J]. 复杂系统与复杂性科学, 2020, 17(4): 9-15.  
WANG Y J, GUO Q, DENG C Y, et al. Research on user traits predicting based on LDA topic model[J]. Complex Systems and Complexity Science, 2020, 17(4): 9-15. (in Chinese)
- [9] 颜端武,梅喜瑞,杨雄飞,等. 基于主题模型和词向量融合的微博文本主题聚类研究[J]. 现代情报, 2021, 41(10): 67-74.  
YAN D W, MEI X R, YANG X F, et al. Research on microblog text topic clustering based on the fusion of topic model and word embedding [J]. Journal of Modern Information, 2021, 41(10): 67-74. (in Chinese)
- [10] 穆晓霞,董星辉,柴旭清,等. 融合 LDA 主题模型和支持向量机的商品个性化推荐方法[J]. 郑州大学学报(理学版), 2022, 54(3): 34-39.  
MU X X, DONG X H, CHAI X Q, et al. Commodity personalized recommendation method integrating LDA topic model and support vector machine[J]. Journal of Zhengzhou University (Natural Science Edition), 2022, 54(3): 34-39. (in Chinese)
- [11] 田保军,刘爽,房建东. 融合主题信息和卷积神经网络的混合推荐算法[J]. 计算机应用, 2020, 40(7): 1901-1907.  
TIAN B J, LIU S, FANG J D. Hybrid recommendation algorithm by fusion of topic information and convolution neural network[J]. Journal of Computer Applications, 2020, 40(7): 1901-1907. (in Chinese)
- [12] 张永宾,赵金陵. 融合 LDA 与注意力的网络信息个性化推荐方法[J]. 计算机仿真, 2022, 39(12): 528-532.  
ZHANG Y B, ZHAO J L. Personalized recommendation method of network information integrating LDA and attention[J]. Computer Simulation, 2022, 39(12): 528-532. (in Chinese)
- [13] 宋晓丽,贺龙威. 基于改进自编码器的在线课程推荐模型[J]. 计算机系统应用, 2022, 31(3): 288-293.  
SONG X L, HE L W. Online course recommendation model based on enhanced auto-encoder[J]. Computer Systems & Applications, 2022, 31(3): 288-293. (in Chinese)
- [14] 子健,李俊,岳兆娟,等. 基于自编码器与属性信息的混合推荐模型[J]. 数据与计算发展前沿, 2021, 3(3): 148-155.  
CHEN Z J, LI J, YUE Z J, et al. Hybrid recommendation model based on autoencoder and attribute information[J]. Frontiers of Data & Computer, 2021, 3(3): 148-155. (in Chinese)
- [15] 周传华,于猜,鲁勇. 融合评分矩阵和评论文本的深度神经网络推荐模型[J]. 计算机应用研究, 2021, 38(4): 1058-1061, 1068.  
ZHOU C H, YU C, LU Y. Recommendation model of deep neural network combining rating matrix and review text[J]. Application Research of Computers, 2021, 38(4): 1058-1061, 1068. (in Chinese)
- [16] 任胜兰,郭慧娟,黄文豪,等. 基于注意力机制交互卷积神经网络的推荐方法[J]. 计算机科学, 2022, 49(10): 126-131.  
REN S L, GUO H J, HUANG W H, et al. Recommendation method based on attention mechanism interactive convolutional neural network [J]. Computer Science, 2022, 49(10): 126-131. (in Chinese)
- [17] 刘羽茜,刘玉奇,张宗霖,等. 注入注意力机制的深度特征融合新闻推荐模型[J]. 计算机应用, 2022, 42(2): 426-432.  
LIU Y Q, LIU Y Q, ZHANG Z L, et al. News recommendation model with deep feature fusion injecting attention mechanism[J]. Journal of Computer Applications, 2022, 42(2): 426-432. (in Chinese)
- [18] 刘振鹏,尹文召,王文胜,等. HRS-DC: 基于深度学习的混合推荐模型[J]. 计算机工程与应用, 2020, 56(14): 169-175.  
LIU Z P, YIN W Z, WANG W S, et al. HRS-DC: hybrid recommendation model based on deep learning[J]. Computer Engineering and Applications, 2020, 56(14): 169-175. (in Chinese)
- [19] 陈彬,张荣梅,张琦. DCFM: 基于深度学习的混合推荐模型[J]. 计算机工程与应用, 2021, 57(3): 150-155.  
CHEN B, ZHANG R M, ZHANG Q. DCFM: hybrid recommendation model based on deep learning[J]. Computer Engineering and Applications, 2021, 57(3): 150-155. (in Chinese)
- [20] 陈蕾,刘铭. 引入用户关注的图推荐模型的研究[J]. 系统工程, 2019, 37(2): 21-29.  
CHEN L, LIU M. The research on graph based recommendation model importing users attention [J]. Systems Engineering, 2019, 37(2): 21-29. (in Chinese)
- [21] 党伟超,姚志宇,白尚旺,等. 基于图模型和注意力模型的会话推荐方法[J]. 计算机应用, 2022, 42(11): 3610-3616.  
DANG W C, YAO Z Y, BAI S W, et al. Session recommendation method based on graph model and attention model[J]. Journal of Computer Applications, 2022, 42(11): 3610-3616. (in Chinese)
- [22] 邹程辉,李卫疆. 融合知识图谱和评论文本的个性化推荐模型[J]. 计算机工程与科学, 2023, 45(1): 181-190.  
ZOU C H, LI W J. A personalized recommendation model integrating knowledge graph and comment text [J]. Computer Engineering & Science, 2023, 45(1): 181-190. (in Chinese)
- [23] 张若一,金柳,马慧芳,等. 融合相似用户影响效应的知识图谱推荐模型[J]. 计算机工程与科学, 2023, 45(3): 520-527.  
ZHANG R Y, JIN L, MA H F, et al. A knowledge graph recommendation model incorporating the influence effect of similar users[J]. Computer Engineering & Science, 2023, 45(3): 520-527. (in Chinese)
- [24] INGMA D P, BA J. Adam: a method for stochastic optimization [EB/OL]. [2023-07-05]. <https://www.xueshufan.com/publication/2964121744>.
- [25] GUO H F, TANG R M, YE Y M, et al. DeepFM: a factorization-machine based neural network for CTR prediction[EB/OL]. [2023-07-05]. <https://arxiv.org/pdf/1703.04247.pdf>.
- [26] KIM D, PARK C, OH J, et al. Convolutional matrix factorization for document context-aware recommendation[C]// Proceedings of the 10th ACM Conference on Recommender Systems. New York, USA: ACM Press, 2016: 11-23.
- [27] ZHENG L, NOROOZI V, YU P S. Joint deep modeling of users and items using reviews for recommendation [C]// Proceedings of the 10th ACM International Conference on Web Search and Data Mining. New York, USA: ACM Press, 2017: 425-434.
- [28] CHENC, ZHANG M, LIU Y Q, et al. Neural attentional rating regression with review-level explanations[C]// Proceedings of 2018 World Wide Web Conference. New York, USA: ACM Press, 2018: 1583-1592.
- [29] 冯兴杰,崔桂颖. 基于交互注意力的可解释性推荐方法[J]. 计算机应用与软件, 2022, 39(10): 292-298, 328.  
FENG X J, CUI G Y. An interpretable recommendation based on interactive attention[J]. Computer Applications and Software, 2022, 39(10): 292-298, 328. (in Chinese)
- [30] MA H, LIU Q. In-depth recommendation model based on self-attention factorization[J]. KSII Transactions on Internet and Information Systems, 2023, 17(3): 121-132.