

# 基于细粒度特征交互的推荐模型

杨振宇<sup>1,2</sup>, 刘国敬<sup>2</sup>, 王 钰<sup>2</sup>

<sup>1</sup>中国矿业大学信息与控制工程学院 徐州 中国 221116

<sup>2</sup>齐鲁工业大学(山东省科学院)计算机科学与技术学院 济南 中国 250353

**摘要** 随着电子商务的兴起,用户在网购的同时留下了大量的评论。用户评论通常包含丰富的用户兴趣和项目属性等语义信息,反应了用户对项目特征的偏好。近年来,许多基于深度学习的方法通过利用评论进行推荐,并取得了巨大成功。这些工作主要是采用注意机制来识别对评分预测很重要的词或方面。它们单一的从评论中提取特征信息,并通过用户和物品的特征交互得到预测分数。然而,过度的聚合可能会导致评论中细粒度信息的丢失。此外,现有的模型要么忽略了用户和项目评论的相关性,要么只在单个粒度上构建评论特性交互,这导致用户和项目的特征信息不能被有效而全面地捕获。针对上述问题,在本文我们考虑通过从评论的多个粒度捕获特征信息,然后为用户和物品进行多粒度下的特征交互,可以实现更好的评分预测和解释性。

为此,我们提出了一种新的用于评分预测的细粒度特征交互网络(FFIN)。首先,模型并没有将用户的所有评论聚合成一个统一的向量,而是将用户和物品的每条评论单独建模,通过堆叠的扩展卷积分层地为每个评论文本构建多层次表示,充分地捕获了评论的多粒度语义信息;其次,模型在每个语义层次上构建用户和物品评论的细粒度特征交互,这有效避免了单粒度交互导致的次级重要信息被忽略的问题;最后,由于用户的评论行为通常是主观且个性化的,我们没有使用注意力机制来识别重要信息,而是通过类似于图像识别的层次结构来识别高阶显著信号,并将其用于最终的评分预测。我们在6个来自Amazon和Yelp的具有不同特征的真实数据集上进行了广泛的实验。我们的结果表明,与最近提出的最先进的模型相比,所提出的FFIN在预测精度方面获得了显著的性能提升。进一步的实验分析表明,多粒度特征的交互,不仅突出了评论中的相关信息,还大大提高了评分预测的可解释性。

**关键词** 特征交互; 评论推荐; 多粒度; 卷积神经网络

**中图法分类号** TP391.3; TP183 DOI号 10.19363/J.cnki.cn10-1380/tn.2021.09.11

## Recommendation model based on fine-grained feature interaction

YANG Zhenyu<sup>1,2</sup>, LIU Guojing<sup>2</sup>, WANG Yu<sup>2</sup>

<sup>1</sup> School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China

<sup>2</sup> School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

**Abstract** With the rise of e-commerce platforms, users left a large number of reviews while shopping online. These reviews usually contain rich semantic information about user interests and item attributes, reflecting user preferences for item features. In recent years, many deep learning-based solutions have been proposed by exploiting reviews for recommendation, and have achieved great success. The attention mechanism is mainly adopted in these works to identify words or aspects that are important for rating prediction. They singularly extract feature information from reviews, and obtain prediction ratings through the feature interaction between users and items. However, excessive aggregation may result in the loss of fine-grained information in reviews. In addition, the existing models either ignore the relevance of user and item reviews or construct review feature interactions only at a single granularity, which results in the features information of users and items not being captured efficiently and comprehensively. In order to address the above issue, in this paper we consider that by capturing feature information from multiple granularities of each review, and then performing feature interactions at multiple granularities for users and items could enable a better rating prediction with explainability.

To this end, we propose a novel fine-grained feature interaction network (FFIN) for rating prediction in this paper. In the first place, instead of aggregating all the user's reviews into a uniform vector, the model deals with each review for the user and the item separately, and construct a multi-level representation for each review text hierarchically by stacked extended convolution, which adequately captures the multi-granularity semantic information of the review. In the second place, the model constructs the fine-grained feature interaction of user and item reviews at each semantic level, which effectively avoids the problem that secondary important information is overlooked due to single-grained interaction. In the end, the user's review is usually subjective and personalized, instead of using the attention mechanism to identify impor-

**通讯作者:** 杨振宇, 博士生, 副教授, Email: yzy@qlu.edu.cn.

本课题得到山东省重大科技创新工程(No. 2020CXGCO10102)资助。

收稿日期: 2021-04-20; 修改日期: 2021-07-31; 定稿日期: 2021-08-10

tant informations, the high-order salient signals are identified by a hierarchical structure similar to image recognition and used in the final rating prediction. Extensive experiments are conducted over six real-world datasets with diverse characteristics. Our results demonstrate that the proposed FINN obtains substantial performance gain over recently proposed state-of-the-art models in terms of prediction accuracy. At the same time, further experimental analysis shows that the interaction of multi-granularity features not only highlights the relevant information in the reviews, but also greatly improves the interpretability of the rating prediction.

**Key words** feature interaction; review recommendation; multi-granularity; convolutional neural network

## 1 研究背景

现如今, 许多社交媒体网站和电子商务系统都允许用户通过撰写评论文本的方式来表达其对所购买商品或服务的使用体验, 同时还倡导进行数字评分<sup>[1-4]</sup>。这些评论文本中包含的丰富信息可以揭示商品的特征以及每个用户的偏好, 并在增强平台推荐系统的个性化方面起到了越来越重要的作用。目前, 许多推荐系统都是利用评论文本来获得更好的评分预测性能, 而不是使用原始而稀疏的用户-物品评分<sup>[5]</sup>。

早期的研究侧重于主题建模<sup>[6-7]</sup>或非负矩阵分解<sup>[8-9]</sup>来推导评论文本中的潜在特征。相比较于传统的基于用户-物品交互<sup>[10-13]</sup>的协同过滤模型, 这些技术具有一定的优越性, 并能获得更好的推荐性能。然而, 由于这些工作使用了词袋表示<sup>[14-15]</sup>模式而没有考虑评论上下文信息, 导致出现了明显的信息损失。

随着深度学习在自然语言处理等领域取得的巨大成功, 许多基于该技术的推荐模型极大的提高了评分预测的性能。在这些工作中, 卷积神经网络<sup>[16-18]</sup>架构分别从用户和物品对应的评论中提取潜在特征。通过使用一个密集的词嵌入矩阵来表示评论, 随后应用一个固定大小的滑动窗口来捕获上下文信息。基于卷积神经网络的技术更好的促进了对评论中语义信息的理解, 导致对现有基于词袋模式的评分预测方法的显著改进。然而, 一个饱受诟病的缺点是人们难以有效的理解神经网络所提取的特征, 这限制了推荐系统的可解释性<sup>[19-21]</sup>。

基于评论的推荐系统的关键在于对评论文本中不同粒度信息的捕获。通常而言, 同一用户对不同物品往往具有不同的偏好, 这些偏好都反映在他所撰写的评论中。同时, 评论中的重要语义特征被隐含在不同粒度的文本片段中。图 1 中的举例说明了这一问题。不同粒度的文本(如: 单词、短语和句子)可以揭示用户对不同物品的偏好。例如单词: “best”、“great”、“disappointing”; 短语: “quality great battery life”、“screen looks amazing” 和句子: “I usually go with product red but it looked a bit orange this year.”。



图 1 评论中的多粒度举例

Figure 1 Multi granularity examples in the reviews

然而, 现有的工作通常通过整合用户和物品的评论文本来为每个用户或物品学习单一的特征向量, 然后通过因子分解等手段来实现评分预测。例如, Seo 等人<sup>[22]</sup>(D-Attn)利用具有局部和全局双重注意力的卷积神经网络来建模用户偏好或物品属性, 并生成两个角度的用户和物品的特征向量。然后, 通过聚合各自的特征向量, 模型生成了具有可解释性的用户和物品表示用于评分预测。Lu (TARMF)等人<sup>[23]</sup>通过优化过的矩阵分解和基于注意力的 GRU<sup>[24]</sup>网络, 从历史评分和评论文本中学习用户偏好和物品属性。Liu (DAML)等人<sup>[25]</sup>基于评分和评论提出了一种双重注意相互学习模型。通过局部注意力和相互注意力来共同学习评论的特征, 以增强模型的可解释性。然后将评分特征和评论特征集成到一个统一的神经网络模型中, 通过神经因子分解机实现特征的高阶非线性交互, 完成最终的评分预测。尽管这些模型在推荐性能上有所提高, 但在捕捉细粒度的用户偏好和物品属性方面存在着局限性, 因为它们仅仅是为用户物品分配一个单一的特征向量, 导致评论中多个粒度下的信息没有被明显的捕获。

在本论文中, 提出了一种细粒度特征交互网络 (FFIN), 它是一种新颖的基于评论的推荐体系结构, 能够解决上述问题。FFIN 的优势在于两个核心: 多层次的用户/物品表示和细粒度的特征交互。该模型不是用单个抽象向量来表示每个用户, 而是在一个统一的模块中使用分层扩展卷积来构建基于评论的多层次表示。通过分层叠加扩展的卷积, 每一层的接收

输入宽度呈指数增长, 而参数的数量仅线性增加。同时, 每一层的输出被保留为跨不同长度的文本段的特征映射, 不会因为使用任何形式的池化或步进卷积而造成覆盖损失。通过这种方法, 模型可以在不同粒度上通过局部相关性和长期依赖性逐步获得评论中的语义特征, 包括词、短语和句子层次。

此外, 为了避免信息丢失, FFIN 在每个语义粒度上构建用户评论和物品评论的特征交互。评论的多粒度特征在不同的层次展现了不同的信息, 基于此的多粒度特征交互不仅提高了模型的准确性, 还避免了单粒度交互导致信息丢失的局限性。在实际应用中, 该模型基于评论文本的分层表示方法, 对每一对评论分别从词到句的层次上构建一个关联矩阵。通过这种方式, 可以识别用户和物品评论中隐含的多个粒度特征, 并以最小的损失进行交互, 从而为预测准确的评分提供足够的内容相关性线索。然后, 将每个粒度的评论对的多个关联矩阵合并成一个 3D 图像, 其通道表示不同粒度下用户和物品评论特征交互的相关程度。通过类似于基于 3D 卷积的图像识别层次, 识别高阶显著信号来预测用户对物品评分。模型在亚马逊和 Yelp 的真实数据集上进行了广泛的实验。结果表明 FFIN 在评论推荐方面具有的优越性。此外, 在一系列的分析实验中发现了 FFIN 能够有效的构建多个粒度特征的交互, 突出了评论中的相关信息, 提高了评分预测的可解释性。综上所述, FFIN 的优点在于: (1)以多层次扩张卷积结构来编码用户和物品评论, 避免了评论中细粒度信息的丢失; (2)在多个粒度下构建用户和物品评论的特征交互, 并使用 3D 卷积来融合处理多粒度信息, 有效的突出了评论中多个粒度下的相关信息; (3)不仅在性能方面有优越性, 还具有强力的可解释性。

## 2 相关工作

基于评级的预测始终是推荐系统研究的热点<sup>[26-28]</sup>。在这一问题上的主要研究思路是紧紧围绕着评分预测来持续展开的, 而且截至目前已有许多优秀的模型在这一课题上取得了优秀的表现。早期研究中, 矩阵分解(Matrix Factorization, MF)是相当主流且受欢迎的方法。它能够通过利用用户和物品的潜在特征来建模用户的显式反馈。例如, 经典的 Probabilistic Matrix Factorization(PMF)<sup>[29]</sup>以非线性因子作为模型的基础来展开讨论, 它使用与用户相关的系数, 将用户的偏好建模成一个系列向量的线性组合。随后研究者们发现, 隐式反馈信息同样有助于用户的偏好建模, 因此提出了 SVD++<sup>[30]</sup>, 该方法认

为, 除了用户对项目的显式历史评分记录具有偏好显示外, 浏览或收藏等隐反馈信息一定程度上同样可以从侧面反映用户的偏好, 但当该数据数量大且稀疏的时候, 推荐结果就表现得不那么令人满意; 矩阵分解的另一个缺点是无法为模型提供可解释性, 当用户与物品产生显式交互时, 尽管模型能够预测得到用户可能给出的评分, 但是却无法解释是什么因素在起作用。

随着深度学习技术的发展, 在评分预测的任务上又产生了新的生机。ConvMF<sup>[18]</sup>, 它将矩阵分解和神经网络相结合, 在使得用户隐向量与项目隐向量做内积尽可能逼近真实评分的同时, 对物品隐向量做了约束, 也就是让物品隐向量跟 CNN 学得特性尽可能的接近。为提高评分预测推荐系统的可解释性, DeepCoNN<sup>[31]</sup>首次提出了一个分别建模用户评论和物品评论的模型。它使用了卷积神经网络分别以并行的形式对用户和物品对评论文本进行处理, 最后利用因子分解机实现评分对预测。ANR<sup>[2]</sup>模型旨在从多角度(Asspect)来建模用户和商品。它将不同的理解角度抽象为模型的不同参数, 这使模型可以通过隐式的方式来抽取评论文本中的 Aspect。NARRE<sup>[5]</sup>模型在基于用户和物品的并行结构的基础上利用注意力机制实现了对每一条评论的贡献的判定, 也就是判别哪些评论对用户的偏好建模是更有用的。

形式上看, 深度学习阶段的模型都是通过并行地对评论文本提取特征来最终得到评分预测结果。但是这些工作只是单方面的为评论文本提取了特征表示, 并基于此做单粒度的特征交互, 这难免会丢失文本中的重要信息。因此, 如何充分挖掘文本中各粒度的有用信息, 并以此来实现特征的多粒度交互将是我们的工作重点。

## 3 基于细粒度特征交互的推荐模型

该模型有两个并行的神经网络组成, 一个用于对用户评论集合进行建模, 另一个用于物品评论建模。为了清楚起见, 本文的主要符号总结在表 1 中。首先, 用户和物品评论都是通过词嵌入的方式转换成连续的词向量矩阵。随后, 这些词向量矩阵被输入到层次扩张卷积结构中进行编码, 并输出多粒度的评论表示。接下来, FFIN 计算多个粒度下的用户和物品评论表示的关联矩阵; 将这些关联矩阵融合成一个 3D 交互图像, 并执行 3D 卷积操作, 输出高阶显著信息。最终, FFIN 将这一高阶显著信息输入到因子分解机中实现评分预测。用于评分预测的细粒度特征交互网络(FFIN)的结构, 如图 2 所示。

表 1 相关变量注释表

Table 1 Comment table of related variables

$S_u^N$	用户对应的原始评论文本集合
$S_i^M$	物品对应的原始评论文本集合
$N$	用户和评论集合的评论数量
$M$	物品评论集合的评论数量
$s^l$	评论文本在 $l$ 粒度下的特征表示
$s_u^n$	第 $n$ 个用户原始的评论文本
$s_i^m$	第 $m$ 个物品原始的评论文本
$A_{n,m}^l[i, j]$	用户与物品评论中第 $i$ 个特征与第 $j$ 个特征在 $l$ 粒度下的相关性
$\hat{r}_{u,i}$	用户和物品交互的预测评分

本小节将依据图 2 展示的模型结构依次描述本论文提出的细粒度特征交互网络(FFIN)。首先, 本节对基于评论的推荐问题进行了定义; 其次, 详细描述了评论表示模块中层次扩张卷积结构对原始评论文本的编码过程; 再次, 展示了用户和物品的多粒度的评论表示特征的交互过程; 最后, 介绍了一个带有因子分解机(FM)的评分预测层实现用户对物品的评分。

### 3.1 问题定义

基于评论的推荐问题可以表述为: 对于给定的用户  $u$  和物品  $i$ , 它们都对应一组原始的评论文本集

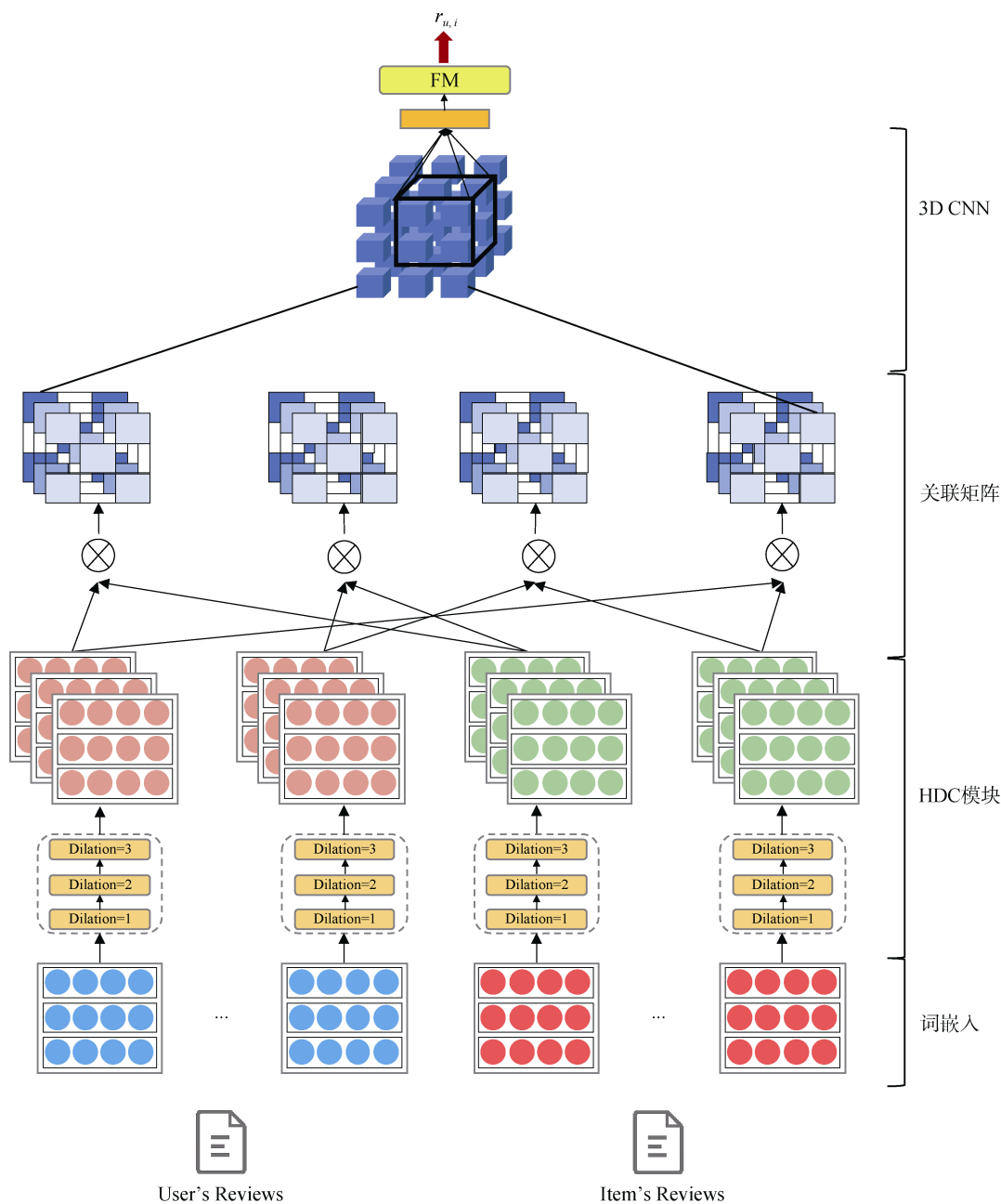


图 2 FFIN 模型结构示意图

Figure 2 Schematic diagram of FFIN model structure

合  $S_u^N$  和  $S_i^M$ , 其中  $N, M$  对应用户和物品评论集合的评论数量。目的是建立一个预测模型, 通过一组用户和物品的评论集合来预测用户对该物品的评分。设一组用户和物品评论集合中单个评论的最大长度分别为  $T_u$  和  $T_i$ , 则第  $n$  个用户和第  $m$  个物品原始的评论文本可以分别表示为:  $s_u^n$  和  $s_i^m$ 。

### 3.2 评论表示模块

由于用户和物品的多粒度评论表示的产生过程是类似的, 因此本节仅展示单个评论文本的编码步骤, 且  $s_u^n$  简化为  $s$ , 评论长度设为  $T$ 。

受到 Lin 等人<sup>[32]</sup>所提出的多层扩张卷积结构的启发, 本节设计了一种层次扩张卷积(hierarchical dilated convolution, HDC)<sup>[33-34]</sup>来直接从评论中学习多个语义粒度的评论表示。

对于给定的评论文本  $s = [w_1, w_2, \dots, w_T]$ , 模型首先通过一个嵌入矩阵将其映射为相应的词向量矩阵, 即  $X = [x_1, x_2, \dots, x_T]$ , 其中  $x_t \in R^d$  是该文本中的第  $t$  个单词的词嵌入表示,  $d$  是词向量的维度大小。然后, HDC 以该词向量矩阵为输入, 捕获评论中的多粒度语义特征。

不同于标准卷积在每一步对输入的一个连续子序列进行卷积, 扩张卷积通过每次跳过  $\delta$  个输入元素, 而拥有更为广泛的接受域, 其中  $\delta$  是扩张率。对于上下文中心单词  $x_t$  和大小为  $2w+1$  的卷积核  $W$ , 扩张卷积操作可以被公式为:

$$F(x_t) = \text{ReLU} \left( W \bigoplus_{k=0}^w x_{t \pm k\delta} + b \right) \quad (1)$$

其中,  $\oplus$  是向量拼接、 $b$  是偏置项、ReLU 是非线性激活函数。每个卷积层的输出是前一层输入的加权组合。与 Lin 等人的设置类似, 本模块以  $\delta=1$  为开始(等同于标准卷积), 以确保不遗漏原始输入序列中的任何元素。之后, 通过以更大的扩张率层次叠加扩张之后的卷积, 卷积文本的长度以指数形式扩展, 只需使用少量的层和适量的参数就可以覆盖不同接受域的语义特征。

此外, 为了防止梯度的消失或爆炸, 模块在每个卷积层的最后应用层归一化<sup>[35]</sup>。由于可能会有不相关的信息引入到长距离的语义单元中, 模型根据实际验证中的性能设计多层次的扩张率。每个堆叠层  $l$  的输出都保存为文本在特定粒度水平上的特征图, 公式为:

$$s^l = [x^l, x^l, \dots, x^l]_{l=1}^T, s^l \in R^{T \times f} \quad (2)$$

其中  $f$  是每一层卷积滤波器的数量。假设有  $L$  层堆叠的扩张卷积, 多粒度评论表示可以定义为  $[s^1, s^2, \dots, s^L]$ 。通过这种方式, HDC 以较小的扩张率从单词和短语层次逐步收获词义和语义特征, 以较大的扩张率从句子层次捕获长期的依赖关系。模型的评论编码模块不仅在并行能力上优于循环神经网络, 而且相比于完全基于注意力的方法能够显著减少内存的消耗。

经过 HDC 模块编码, 单个用户和物品的多粒度评论表示可以分别被表述为:

$$s_u^n = [s_u^1, s_u^2, \dots, s_u^L], s_u^n \in R^{L \times T_u \times f} \quad (3)$$

$$s_i^m = [s_i^1, s_i^2, \dots, s_i^L], s_i^m \in R^{L \times T_i \times f} \quad (4)$$

其中,  $s_u^n$  和  $s_i^m$  分别为用户和物品评论集合  $S_u$ 、 $S_i$  中的第  $n$  个和第  $m$  个评论文本。

### 3.3 特征交互模块

给定的用户和物品的多粒度评论表示集合  $S_u = \{s_u^n\}_{n=1}^N \in R^{N \times L \times T_u \times f}$  和  $S_i = \{s_i^m\}_{m=1}^M \in R^{M \times L \times T_i \times f}$ , 模型通过在每个粒度上构建关联矩阵实现用户和物品之间特征的交互。给定用户和物品的第  $n$  个和第  $m$  个多粒度评论表示, 第  $l$  层粒度的关联矩阵可以通过如下的点乘运算实现:

$$A_{n,m}^l = \frac{s_u^n[l] \cdot (s_i^m[l])^T}{\sqrt{f}}, A_{n,m}^l \in R^{T_u \times T_i} \quad (5)$$

其中,  $l=[1, 2, \dots, L]$  是语义粒度的数目。  $A_{n,m}^l$  中的元素  $A_{n,m}^l[i, j]$  代表着用户与物品评论中第  $i$  个特征与第  $j$  个特征在  $l$  粒度下的相关性。

为了总结整个用户和物品的评论中的多粒度特征, FFIN 将用户和物品的所有交互得到的关联矩阵融合成一个 3D 交互图像, 公式为:

$$Q = \{Q_{k,i,j}\}_{n \times m \times T_u \times T_i} \quad (6)$$

其中,  $n \times m$  代表关联矩阵在单个粒度下的总数量, 3D 交互图像中的每个像素点由以下方式得到:

$$Q_{k,i,j} = [A_{n,m}^l[i, j]]_{l=0}^L \quad (7)$$

具体来讲, 每个像素点都是所有粒度下特征的拼接向量, 表示用户与物品在多个粒度下的交互程度。

由于用户的评分行为通常是带有主观性的个性化, 因此在面对不同的物品时会体现出不同的偏好。受 Wang 等人<sup>[34]</sup>和 Zhou 等人<sup>[36]</sup>在新闻推荐和对话系统问题上的启发, 模型根据类似于图像识别的组成层次结构, 采用分层的 3D 卷积神经网络<sup>[37-40]</sup>和最大

池化操作, 从整个图像中识别出突出的匹配信号。三维卷积是典型的二维卷积的扩展, 其滤波器和步长都是三维立方体。形式上, 第  $t$  层的第  $z$  个特征图上的高阶像素在  $(k, i, j)$  处的计算方式为:

$$Q_{k,i,j}^{(t,z)} = ELU \left( \sum_{z'} \sum_{w=0}^{W_t-1} \sum_{h=0}^{H_t-1} \sum_{r=0}^{R_t-1} K_{w,h,r}^{(t,z)} \cdot Q_{k+w,i+h,j+r}^{(t-1,z')} + b^{(t)} \right) \quad (8)$$

其中,  $ELU^{[41]}$  是非线性激活函数。 $z'$  代表着上一层的特征图,  $K^{(t,z)} \in R^{W_t \times H_t \times R_t}$  和  $b^{(t)}$  分别是 3D 卷积核和偏置项, 大小为  $W_t \times H_t \times R_t$ 。随后, 通过一个最大池化操作提取出显著的信息:

$$\hat{Q}_{k,i,j}^{(t,z)} = \max \left( Q_{k:k+P_w^{(t,z)}-1, i:i+P_h^{(t,z)}-1, j:j+P_r^{(t,z)}-1}^{(t,z)} \right) \quad (9)$$

其中,  $P_w^{(t,z)}$ ,  $P_h^{(t,z)}$  和  $P_r^{(t,z)}$  是 3D 池化操作的大小。最后一层的输出是将用户和物品评论之间的综合交互向量进行拼接, 表示为  $\hat{y} \in R^e$ 。

### 3.4 评分预测模块

在本小节中, 模型将上述特征交互模块产生的预测向量  $\hat{y}$  输入到全连接神经网络和传统的因子分解机(FM)<sup>[42]</sup>中来实现最终的用户-物品评分预测。全连接神经网络的实现过程可以被公式如下:

$$\hat{o} = \hat{y}W_0 + b_0 \quad (10)$$

其中,  $W_0 \in R^{n \times e}$  和  $b_0 \in R^n$  分别是全连接网络的权重参数和偏置项。

FM 接受一个实值特征向量, 并使用分解参数对特征之间的相互作用进行建模。其定义如下:

$$\hat{r}_{u,i} = \hat{w}_1 + \sum_{i=1}^{|\hat{o}|} \hat{w}_i \hat{o}_i + \sum_{i=1}^{|\hat{o}|} \sum_{j=i+1}^{|\hat{o}|} \langle \hat{v}_i, \hat{v}_j \rangle \hat{o}_i \hat{o}_j \quad (11)$$

其中,  $\hat{o}_j \in R^n$  是实值输入特征向量。 $\langle \cdot, \cdot \rangle$  是点乘操作。参数  $\{v_1, \dots, v_n\}$ ,  $v \in R^n$  是用于建模交互队列  $(x_i, x_j)$  的分解参数。 $\hat{w}_1$  和  $\sum_{i=1}^n \hat{w}_i$  分别是全局偏置项和线性回

归组件, FM 的权重参数。FM 的输出  $\hat{r}_{u,i}$  是一个标量, 代表用户和物品交互的预测评分。

## 4 实验结果与分析

在本小节中, 实验在 6 个来自亚马逊和 Yelp 的具有不同特征的真实数据集上进行了一系列的评估验证。首先, 本节介绍了 6 个数据集的统计细节、模型的参数设置以及一些用于对比的基线模型; 随后, 展示了 FFIN 与对比模型在 6 个数据集上的对比结果并进行了一些讨论。最后, 从多个角度对模型进行了定量分析, 并对模型进行了可视化的研究。

### 4.1 数据集和预处理

本实验的数据集有两个不同的来源, 描述如下:

(1) Amazon Product Reviews<sup>[43]</sup>: 亚马逊是一个知名的电子商务平台, 用户可以为他们购买的产品写评论。本实验依据不同的领域和不同的尺度选择了 5 个具有不同特征的数据集, 分别是: Musical Instruments(Music)、Video Games(Video)、Office Products(Office)、Grocery & Gourmet Food(Grocery) 和 Tools & Home Improvement(Tools)。其中 Tools 是最大的, 它拥有超过 24 万用户的 200 万的评论, 而最小的数据集 Music 只有大约 20 万的评论。

(2) Yelp Dataset Challenge<sup>[44]</sup>: Yelp 是一个针对餐馆、酒吧、温泉等企业的在线点评平台。由于原始数据非常大且稀疏, 本实验对其进行了预处理, 以确保所有用户和项目至少有 5 个评分。即便如此, 它仍然是所有数据集中最大的数据集。它包含了大约 20 万用户的 300 多万条评论。

这些数据集的评分都是从 1~5 的整数。因为评论的长度和数量具有长尾效应, 本实验统一只保留长度和数量覆盖 85% 的用户和物品。此外, 针对于评论文本词典的构建, 本实验采用了类似于 CARL<sup>[21]</sup>的方式: 1) 删除停止词和文档频率高于 0.7 的词; 2) 选取频率最高的前 40000 个词作为词汇表; 3) 删除每个原始文档中除词汇表之外的词。数据集信息如表 2 所示。

表 2 六个数据集在预处理之后的统计信息

Table 2 Statistical information of six data sets after preprocessing

数据集	用户数量	物品数量	评分数量	用户评论的平均词汇数	物品评论的平均词汇数
Musical	27528	10620	231344	128.84	229.70
Video	55217	17408	497419	163.12	293.65
Office	101498	27965	800144	98.71	200.67
Grocery	127487	41320	1143470	99.26	191.04
Tools	240796	73649	2070309	117.86	216.91
Yelp16	167150	100139	1217208	133.21	155.20



为了评估所有算法的性能, 本文采用了均方误差(Mean Square Error, MSE), 它被广泛应用于推荐系统的评分预测任务。均方误差分数越小, 表明模型的性能更强。给定模型的预测评分  $\hat{r}_{u,i}$  和真实评分  $r_{u,i}$ , 均方误差可以被计算如下:

$$MSE = \frac{1}{N} \sum_{n=1}^N (\hat{r}_{u,i} - r_{u,i})^2 \quad (12)$$

其中,  $N$  是在测试集上用户-物品对的样本数量。

## 4.2 对比基线

为了评估模型的评分预测性能, 本节将 FFIN 与目前五种先进的模型进行了比较, 分别是 PMF<sup>[29,45]</sup>、NMF<sup>[8]</sup>、DeepCoNN<sup>[31]</sup>、DAML<sup>[25]</sup> 和 MPCN<sup>[46]</sup>。前两种模型在训练时仅使用评分信息, 后四种模型是具有代表性的评估模型。表 3 列出了对比模型的差异。

(1) PMF(Probabilistic Matrix Factorization): 概率矩阵分解是一种标准的矩阵因式分解方法, 它利用高斯分布对初始用户和物品的潜在表示进行初始化。

(2) NMF(Non-negative Matrix Factorization): 它使分解后的所有分量均为非负值(要求纯加性的描

述), 并且同时实现非线性的维数约减。非负矩阵分解仅使用评分矩阵作为算法的输入。

(3) DeepCoNN(Deep Cooperative Neural Networks): 深度协同神经网络可能是研究者们第一尝试通过深度学习技术来构建基于评论的推荐模型, 它使用两个并行的卷积架构从用户和物品的评论中推断潜在的特征表示。用户和物品的特征表示被连接起来, 并用作因子分解机(FM)的输入, 用于评分预测。

(4) DAML(Dual Attention Mutual Learning): 双重注意相互学习侧重于关注用户和物品评论之间的局部信息以及两者的交互。通过利用卷积神经网络的局部注意力和相互注意力的模块, DAML 模型的可解释性得到增强。然后, DAML 将评分特征和评论特征整合到统一的框架中, 利用神经因子分解机实现特征的高阶非线性交互, 完成最终的评分预测。

(5) MPCN: 这是一种多指针学习方案, 该模型从用户和项目的评论中提取重要的评论, 然后逐一匹配它们。它不仅可以利用信息最丰富的评论进行预测, 还可以进行更深层次的文字互动。

表 3 与基线模型的差异对比  
Table 3 Comparison with baseline model

差异	PMF	NMF	DeepCoNN	DAML	MPCN	FFIN
评分信息	√	√	√	√	√	√
评论文本	/	/	√	√	√	√
特征交互	/	/	/	√	√	√
多粒度特征交互	/	/	/	/	/	√

## 4.3 参数设置

本实验将数据集随机拆分为训练集(80%)、验证集(10%)和测试集(10%)。验证集用于调整超参数, 最后的性能比较在测试集上进行。基线算法的参数初始化与相应论文中的参数相同, 然后仔细调优, 以达到最佳性能。实验在 PyTorch 框架中实现本论文提出的模型。所有模型都使用 Adam<sup>[47]</sup> 优化器进行训练。对于所有模型, 实验都进行了 20 个周期的训练, 一旦模型性能在 5 个周期内没有提高训练将立即停止, 并将性能最好的模型的测试结果报告在开发集上。对于基于深度学习的模型: DeepCoNN、DAML、MPCN 和 FFIN, 学习率从 [0.005, 0.01, 0.015, 0.02] 选择。为了避免出现过拟合现象, 模型在 [0.1, 0.2, 0.3, 0.5, 0.7] 实验中选择并调整退化率。批处理大小根据数据集的大小从 50、100 和 150 中选择。评论中词嵌入的维度为 300, 经过比较该模型使用预训练的

Glove 嵌入向量进行初始化。对于所有使用卷积结构作为评论编码的模型, 卷积层的滤波器个数设为 100, 卷积滤波器的窗口大小设为 3。

对于 FFIN, 受限于 GPU 内存的限制, 实验设置每个用户和物品的评论集合数量都不得大于 5。实验测试了以不同扩张率叠加 1~5 个 HDC 层。最终使用了 [1,2,3] 三层层次结构(每个卷积层的扩张率), 因为它在验证集中获得了最佳性能。用于评论表示的卷积滤波器的窗口大小和数量分别为 3 和 100。对于交叉交互模块, 模型采用两层合成的方法提取出三维交互图像的高阶显著特征, 第一、二层三维卷积滤波器的个数和窗口大小分别为 32-[3,3,3]、16-[3,3,3], 它们的步长都为 [1,1,1]。接下来使用大小为 [3,3,3] 的最大池化, 其步长为 [3,3,3]。

## 4.4 实验结果

表 4 展示了本论文提出的模型与基线模型的比

较结果。

可以做几项观察。首先, 仅利用评分的 PMF 模型在所有的 6 个数据集上都是最差的。具体到两个数据集 Tools 和 Yelp16, 性能差距更加明显。这一点

也同样反映在没有使用评论文本的 NMF 上。这证明评论中蕴含的语义信息有助于捕获用户偏好和物品属性。PMF 和 NMF 忽略了评论文本的作用导致它们在建模用户偏好和物品属性方面出现很大地偏差。

表 4 FFIN 与基线的性能对比(MSE 值)

Table 4 Performance comparison between FFIN and baseline (MSE value)

模型	Musical	Video	Office	Grocery	Tools	Yelp16
PMF	1.721	1.819	1.983	2.103	2.231	2.574
NMF	1.685	1.888	1.756	2.013	2.220	2.330
Deep CoNN	1.322	1.537	1.650	1.401	1.465	1.593
DAML	1.105	1.274	1.454	1.245	1.313	1.432
MPCN	<b>0.971</b>	1.255	<b>0.856</b>	1.187	1.106	1.365
FFIN	1.013	<b>1.145</b>	0.955	<b>1.024</b>	<b>1.100</b>	<b>1.165</b>

其次, 所有的深度学习模型在充分利用评论文本之后, 性能都出现了明显的提升。这再次印证了评论文本中语义信息的重要性。作为较早期出现的模型, DeepCoNN 在所有 6 个数据集上的表现处于最劣势的位置, 这是合理的, 因为它在特征提取的过程中仅使用了卷积神经网络来建模评论文本, 没有使用词级注意力等手段对重要的词和评论进行加权, 导致其可解释性较差。相比之下, DAML 在均方误差指标上则要好的多, 它不仅使用了能够突出重要词的局部注意力机制, 还通过交互注意力机制来动态的学习用户和物品评论的潜在表示。因此, 相比于 DeepCoNN, 更加具有可解释性。MPCN 通过指针网络实现的评论级别和词级别的交互, 它不仅可以深层次地利用信息最丰富的评论进行预测, 还可以进行更深层次的文字互动。作为目前最先进的模型, MPCN 在性能方面都取得了最好的基线效果。

最后, 如表 3 所示, FFIN 除了 Musical 和 Office 数据集的性能略差一些, 在另外 4 个数据集上都取得了最佳成绩。FFIN 通过层次扩张卷积来编码原始的评论, 生成多粒度的评论表示, 最大程度地利用了评论各个粒度的语义信息。值得注意的是, FFIN 在 Grocery 数据集上获得了显著的改进, 因为 Grocery 是第二稀疏的数据集, 评论信息相对较少(参考表 1)。这表明通过在多个粒度下构建关联矩阵, 用户和物品评论中难以联系的特征在模型中被挖掘出来, 从而提高了评分预测性能。即使与 DAML、MPCN 两种最近提出的先进模型相比, FFIN 在均方误差指标上还是分别获得了 17.6%和 3.6%的整体改进。这证明 FFIN 能够精确地提取出评论中细粒度的信息, 从而避免了不必要的信息损失, 这对提升模型评分

的性能是显著的。总体而言, FFIN 能够有效地从评论中建模这一评分行为, 从而实现评级预测。

4.5 定量分析

本节首先讨论了四种不同类型的词嵌入对 FFIN 的性能影响, 即: Word2Vec 嵌入<sup>[48]</sup>、Glove 嵌入<sup>[49]</sup>、Random 嵌入和 BERT 嵌入<sup>[50]</sup>。图 3 显示了在 6 个数据集上所有不同的词嵌入的均方误差得分。

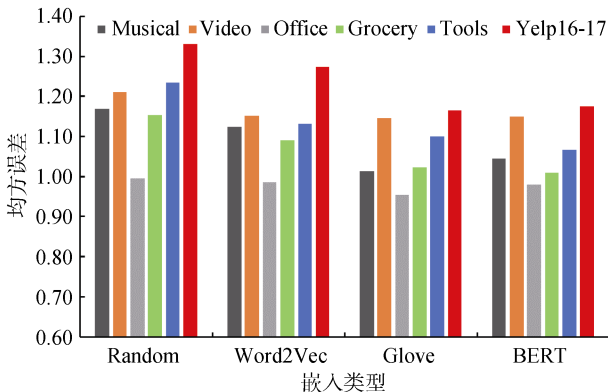


图 3 不同类型的嵌入对模型性能的影响

Figure 3 The impact of different types of embeddings on model performance

通过以上数据可以总结以下几点:

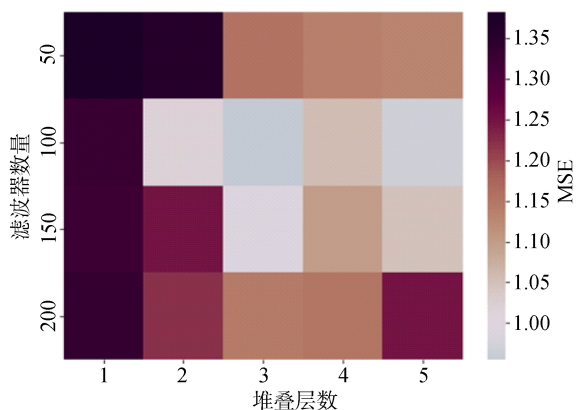
- (1) 与 Word2Vec、Glove 和 BERT 3 种预训练好的词嵌入相比, FFIN 在使用 Random 嵌入时的性能表现在所有数据集上都是比较差的。这与实际的预期一致, 因为预训练好的词嵌入总是能够带来更准确的语义信息。
- (2) 与 Glove 和 BERT 相比, Word2Vec 嵌入在所有数据集上的表现略微差一些。
- (3) 相对于 BERT 嵌入, Glove 嵌入在所有数据



集上的表现更好。因此, 本实验选它作为默认的词嵌入来初始化原始的评论文本。

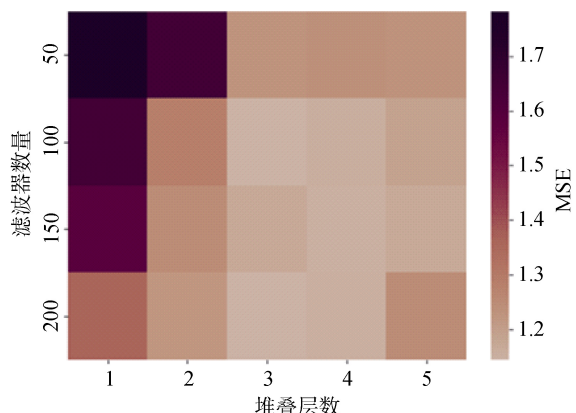
本节随后比较了在 HDC 评论表示模块中扩张卷积滤波器的数量和堆叠 HDC 层数的不同组合对模型性能的影响。图 4 展示了在 Video 和 Office 数据集上的结果。其中较深的区域代表较大的数值, 这表明模型性能越差。

如图 4 所示, 可以很容易的观察到在每层过滤器数量不同的设置中, 有一个一致的趋势, 即在前几个堆叠层的时候, 模型性能有一个显著的改善, 这种增益在当层数到达 3 时, 达到顶峰。然后, 当层数增长到 5 层时, 性能明显下降了很多。因此, 可以确定 HDC 层数的多少对于细粒度信息的提取和推荐的准确性确实很重要。另外, 还可以观察到在 HDC 层数不断变化的过程中, 模型性能也会出现一个先上升后下降的趋势。这两种现象出现的原因可能在



(a) HDC 各层次结构在 Video 数据集上的表现

(a) The performance of HDC's various hierarchical structures on the Video data set



(b) HDC 各层次结构在 Office 数据集上的表现

(b) The performance of HDC's various hierarchical structures on the Office data set

图 4 HDC 各层次结构数据集上的表现

Figure The performance of HDC's various hierarchical structures

于当 HDC 层数和滤波器数量较少时, 模型无法全面的捕获评论中的细粒度信息; 随着层数和数量的上升, 模型所捕获语义信息的粒度越来越精细, 因此模型的推荐性能也在提升。而当 HDC 层数和滤波器数量过大时, 细粒度信息往往会产生冗余和噪声, 这无助于模型的性能。

通过分析, 本实验将 HDC 堆叠的层数和卷积滤波器数量的最佳设置分别为 3 和 100。在这种情况下, 每层的卷积滤波器的扩张卷积的接受域在[3-7-13]之间(扩张率为[1,2,3]), 与评论词序列的平均长度相比, 通过局部交互的分层组成, 足以对多粒度的语义信息进行建模。

本实验进一步探索了带有最大池化操作的三维卷积神经网络对处理交互图像  $Q$  的影响。如图 5 所示为 FFNN 在 6 个数据集的对比结果, 其中  $a\_b$  表示 3D 卷积的层次结构中第一层和第二层的卷积滤波器数量分别为  $a$  和  $b$ 。

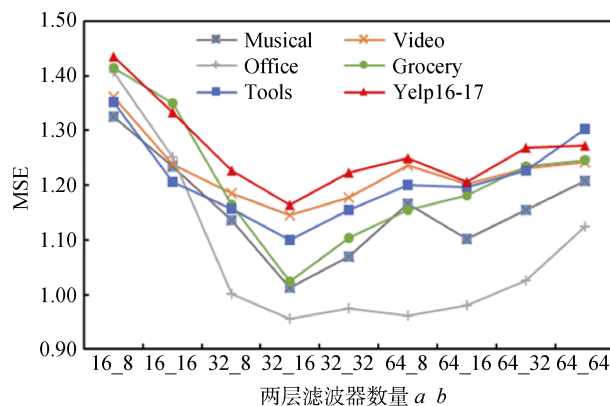


图 5 3D 卷积各层次结构在各个数据集上的表现

Figure 5 The performance of each level of 3D convolution on each data set

如图所示, 给定第一层的滤波器数量  $a$ , 由于可以提取更多的高阶信息, 因此性能首先随着第二层的滤波器数量  $b$  越大而增加。然后性能开始下降, 可能是因为模型中引入了更多的噪声模式(如[32\_8, 32\_16, 32\_32])。此外, 在  $b$  值相同而  $a$  值不同的层次结构中也存在类似的趋势(如[16\_8, 32\_8, 64\_8])。

本实验还进行了其他验证, 如: 改变 3D 卷积的层数([1,2,3,4])或者窗口大小([2,3,4])。最终的结果表明最优的三维卷积层次结构为两层, 第一层卷积滤波器数量为 32, 窗口大小为[3,3,3]; 第二层卷积滤波器数量为 16, 窗口大小为[3,3,3]。

#### 4.6 可视化研究

本小节进一步研究了构建多粒度评论表示和进行多粒度特征交互的有效性。

图 6 显示了用户和物品评论之间的多粒度交互矩阵(公式 5), 其中  $M_l$  表示第  $l$  层 HDC 输出用户和物品评论表示的交互矩阵。

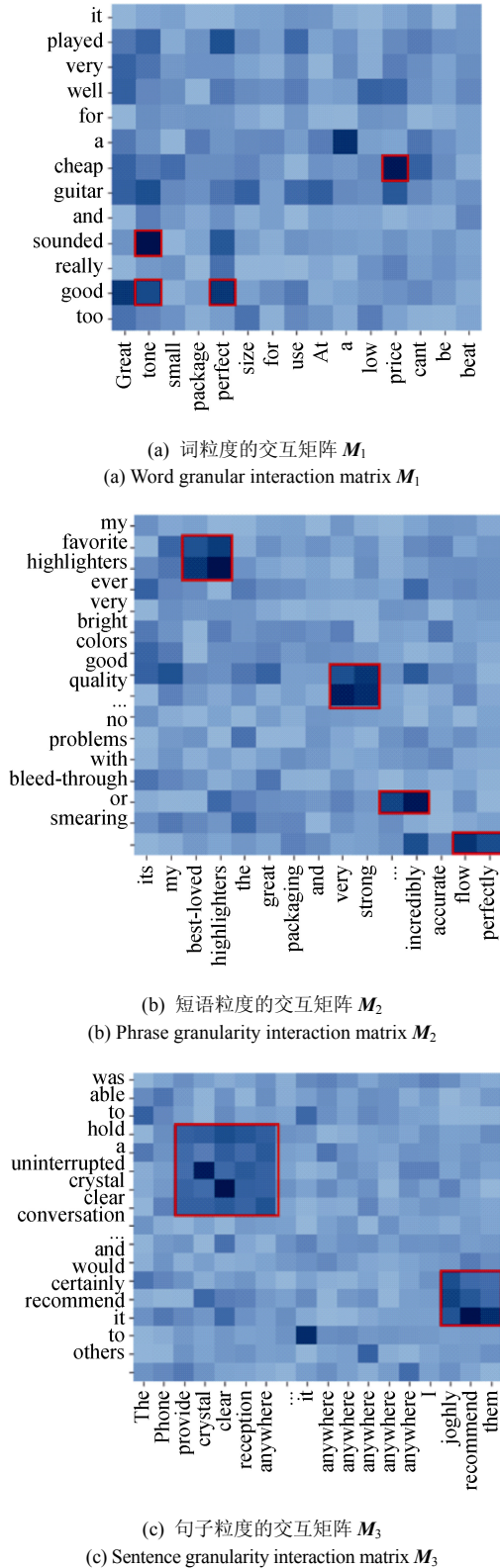


图 6 交互矩阵

Figure Interaction matrix

基于上图中数据观察到, 在第一层的交互矩阵

捕捉到的语义信息主要是词汇相关性。例如“sounded”、“tone”、“good”、“cheap”、“perfect”和“price”这几个词在  $M_1$  中的相关性较高, 并被赋予了较高的值, 这可能是由于它们在词嵌入编码中具有相似的同现信息。

与词粒度不同的是, 短语粒度的交互矩阵  $M_2$  通常能够识别出更为复杂的语义结构和潜在的长期依赖。这可以看出, “favorite highlighters”和“best-loved highlighters”、“good quality”和“very strong”的交互区域显著的获得了较大的相关性分数。同样地, 在更高级的交互矩阵  $M_3$  中也具有类似的相关性。“hold a uninterrupted crystal clear conversation”和“provide crystal clear reception anywhere”、“highly recommend them”和“certainly recommend it”这两对句子的表达具有高度相关的含义。

综上所述, 本论文提出的模型有能力识别评论中高度相关的词、短语和句子, 而忽略信息量较少的部分, 这有利于 FFIN 更准确地捕捉用户偏好和物品属性。这也充分证明了该模型具有强大的可解释性。

## 5 总结

在本文中, 提出了一种基于多粒度表示和交互的神经网络推荐架构。与之前的工作首先将用户和物品的评论整合到一个单一的表达向量中, 然后做单粒度的交互不同, 本论文提出的模型可以通过在多层次语义粒度上执行每对评论表示之间的交互捕捉更细粒度的相关信息。在从亚马逊和 Yelp 收集的真实数据集上进行的大量实验表明, 该模型具有较为先进的推荐性能。在一系列的探究实验表明 FFIN 能够构建具有丰富语义的多粒度评论表示, 并能通过关联矩阵构建全面的交互, 显著地提高了评分预测的可解释性。在未来, 本研究将尝试利用更多与用户和物品紧密相关的数据源如: 图片, 来进一步提升推荐的准确性。

## 参考文献

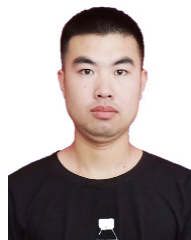
- [1] Cheng Z Y, Ding Y, Zhu L, et al. Aspect-Aware Latent Factor Model: Rating Prediction with Ratings and Reviews[C]. *The 2018 World Wide Web Conference*, 2018: 639-648.
- [2] Chin J Y, Zhao K Q, Joty S, et al. ANR: Aspect-Based Neural Recommender[C]. *The 27th ACM International Conference on Information and Knowledge Management*, 2018: 147-156.
- [3] Li C L, Quan C, Peng L, et al. A Capsule Network for Recommendation and Explaining what You Like and Dislike[C]. *The 42nd International ACM SIGIR Conference on Research and De-*

- velopment in Information Retrieval, 2019: 275-284.
- [4] Liu H T, Wu F Z, Wang W J, et al. NRPA: Neural Recommendation with Personalized Attention[C]. *The 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019: 1233-1236.
  - [5] Chen C, Zhang M, Liu Y Q, et al. Neural Attentional Rating Regression with Review-Level Explanations[C]. *The 2018 World Wide Web Conference*, 2018: 1583-1592.
  - [6] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet Allocation[J]. *The Journal of Machine Learning Research*, 2003, 3: 993-1022.
  - [7] Wang C, Blei D M. Collaborative Topic Modeling for Recommending Scientific Articles[C]. *The 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2011: 448-456.
  - [8] Févotte C, Idier J. Algorithms for Nonnegative Matrix Factorization with the B-Divergence[J]. *Neural Computation*, 2011, 23(9): 2421-2456.
  - [9] McAuley J, Leskovec J. Hidden Factors and Hidden Topics: Understanding Rating Dimensions with Review Text[C]. *The 7th ACM conference on Recommender systems*, 2013: 165-172.
  - [10] Tan Y, Zhang M, Liu Y, et al. Rating-boosted Latent Topics: Understanding Users and Items with Ratings and Reviews[C]. *IJCAI*, 2016, 16: 2640-2646.
  - [11] He X N, Liao L Z, Zhang H W, et al. Neural Collaborative Filtering[C]. *The 26th International Conference on World Wide Web*, 2017: 173-182.
  - [12] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann Machines for Collaborative Filtering[C]. *The 24th international conference on Machine learning - ICML '07*, 2007: 791-798.
  - [13] Georgiev K, Nakov P. A Non-Iid Framework for Collaborative Filtering with Restricted Boltzmann Machines[C]. *International Conference on Machine Learning*, PMLR, 2013: 1148-1156.
  - [14] Harris Z S. Distributional Structure[J]. *WORD*, 1954, 10(2/3): 146-162.
  - [15] Ruch P, Baud R, Geissbühler A. Evaluating and Reducing the Effect of Data Corruption when Applying Bag of Words Approaches to Medical Records[J]. *International Journal of Medical Informatics*, 2002, 67(1/2/3): 75-83.
  - [16] Chen Y. Convolutional neural network for sentence classification [D]. University of Waterloo, 2015.
  - [17] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences [J]. arXiv preprint arXiv:1404.2188, 2014.
  - [18] Kim D, Park C, Oh J, et al. Convolutional Matrix Factorization for Document Context-Aware Recommendation[C]. *The 10th ACM Conference on Recommender Systems*, 2016: 233-240.
  - [19] Deshpande M, Karypis G. Item-Based Top- N Recommendation Algorithms[J]. *ACM Transactions on Information Systems*, 2004, 22(1): 143-177.
  - [20] Ling G, Lyu M R, King I. Ratings Meet Reviews, a Combined Approach to Recommend[C]. *The 8th ACM Conference on Recommender systems*, 2014: 105-112.
  - [21] Wu L B, Quan C, Li C L, et al. A Context-Aware User-Item Representation Learning for Item Recommendation[J]. *ACM Transactions on Information Systems*, 2019, 37(2): 1-29.
  - [22] Seo S, Huang J, Yang H, et al. Interpretable Convolutional Neural Networks with Dual Local and Global Attention for Review Rating Prediction[C]. *The Eleventh ACM Conference on Recommender Systems*, 2017: 297-305.
  - [23] Lu Y C, Dong R H, Smyth B. Coevolutionary Recommendation Model: Mutual Learning between Ratings and Reviews[C]. *The 2018 World Wide Web Conference*, 2018: 773-782.
  - [24] ho K, van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation [J]. arXiv preprint arXiv:1406.1078, 2014.
  - [25] Liu D H, Li J, Du B, et al. DAML: Dual Attention Mutual Learning between Ratings and Reviews for Item Recommendation[C]. *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019: 344-352.
  - [26] Li N, Yin L, Peng Q X. Learning to Rank for Review Rating Prediction[J]. *Applied Mechanics and Materials*, 2014, 556: 6286-6289.
  - [27] Lim Y J, Teh Y W. Variational Bayesian Approach to Movie Rating Prediction[J]. *Proceedings of KDD Cup and Workshop*, 2007: 15-21.
  - [28] Lei X J, Qian X M, Zhao G S. Rating Prediction Based on Social Sentiment from Textual Reviews[J]. *IEEE Transactions on Multimedia*, 2016, 18(9): 1910-1921.
  - [29] Mnih A, Salakhutdinov R R. Probabilistic Matrix Factorization[C]. *Advances in Neural Information Processing Systems*, 2008: 1257-1264.
  - [30] Koren Y. Factor In the Neighbors[J]. *ACM Transactions on Knowledge Discovery from Data*, 2010, 4(1): 1-24.
  - [31] Zheng L, Noroozi V, Yu P S. Joint Deep Modeling of Users and Items Using Reviews for Recommendation[C]. *The Tenth ACM International Conference on Web Search and Data Mining*, 2017: 425-434.
  - [32] Lin J Y, Su Q, Yang P C, et al. Semantic-Unit-Based Dilated Convolution for Multi-Label Text Classification [J]. arXiv preprint arXiv:1808.08561, 2018.
  - [33] Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions [J]. arXiv preprint arXiv:1511.07122, 2015.
  - [34] Wang H Y, Wu F Z, Liu Z, et al. Fine-Grained Interest Matching for Neural News Recommendation[C]. *The 58th Annual Meeting of*

- the Association for Computational Linguistics*, 2020: 836-845.
- [35] Ba J L, Kiros J R, Hinton G E. Layer Normalization[EB/OL]. 2016: arXiv: 1607.06450[stat.ML]. <https://arxiv.org/abs/1607.06450>.
- [36] Zhou X Y, Li L, Dong D X, et al. Multi-Turn Response Selection for Chatbots with Deep Attention Matching Network[C]. *The 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2018: 1118-1127.
- [37] Ji S W, Xu W, Yang M, et al. 3D Convolutional Neural Networks for Human Action Recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(1): 221-231.
- [38] Tran D, Bourdev L, Fergus R, et al. Learning Spatiotemporal Features with 3D Convolutional Networks[C]. *2015 IEEE International Conference on Computer Vision*, 2015: 4489-4497.
- [39] Xu H J, Das A, Saenko K. R-C3D: Region Convolutional 3D Network for Temporal Activity Detection[C]. *2017 IEEE International Conference on Computer Vision*, 2017: 5794-5803.
- [40] Shou Z, Chan J, Zareian A, et al. CDC: Convolutional-de-Convolutional Networks for Precise Temporal Action Localization In Untrimmed Videos[C]. *2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 1417-1426.
- [41] Clevert D A, Unterthiner T, Hochreiter S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) [J]. arXiv preprint arXiv:1511.07289, 2015.
- [42] Rendle S. Factorization Machines[C]. *2010 IEEE International Conference on Data Mining*, 2010: 995-1000.
- [43] He R N, McAuley J. Ups and Downs: Modeling the Visual Evolution of Fashion Trends with One-Class Collaborative Filtering[C]. *The 25th International Conference on World Wide Web*, 2016: 507-517.
- [44] Asghar N. Yelp Dataset Challenge: Review Rating Prediction [J]. arXiv preprint arXiv:1605.05362, 2016.
- [45] Koren Y, Bell R, Volinsky C. Matrix Factorization Techniques for Recommender Systems[J]. *Computer*, 2009, 42(8): 30-37.
- [46] Tay Y, Luu A T, Hui S C. Multi-Pointer Co-Attention Networks for Recommendation[C]. *The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018: 2309-2318.
- [47] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization [J]. arXiv preprint arXiv:1412.6980, 2014.
- [48] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations In Vector Space[EB/OL]. 2013.
- [49] Pennington J, Socher R, Manning C. Glove: Global Vectors for Word Representation[C]. *The 2014 Conference on Empirical Methods in Natural Language Processing*, 2014: 1532-1543.
- [50] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. arXiv preprint arXiv:1810.04805, 2018.



杨振宇 于 2007 年在齐鲁工业大学计算机应用技术专业获得硕士学位, 现在中国矿业大学控制工程与控制理论专业攻读博士学位。研究领域为机器学习、社会计算。研究兴趣包括: 自然语言处理、推荐方法、模式识别。Email: yzy@qlu.edu.cn



刘国敬 于 2017 年在齐鲁工业大学(山东省科学院)测控技术专业获得学士学位, 现在齐鲁工业大学(山东省科学院)计算机应用技术专业攻读硕士学位。研究领域为机器学习、自然语言处理。研究兴趣包括: 自然语言处理、推荐方法、图像处理。 Email: liugj101@163.com



王钰 于 2016 年在齐鲁工业大学(山东省科学院)软件工程专业获得学士学位。现在齐鲁工业大学(山东省科学院)电子信息专业攻读硕士学位。研究领域为人工智能、推荐方法。研究兴趣包括: 推荐方法、自然语言处理、模式识别。Email: qlu\_wy@163.com