

文章编号:1007-5321(2021)04-0115-06

DOI:10.13190/j.jbupt.2020-250

双重注意力充分组合评论特征的推荐模型

李 剑, 刘 鹏, 刘 维

(北京邮电大学 人工智能学院, 北京 100876)

摘要: 针对基于评论文本推荐系统的特征没有充分组合的问题,提出一种利用双重注意力实现评论特征组合的推荐模型。首先利用经情感分类任务微调后的编码模型对评论文本进行编码,得到对应的特征向量;然后利用双线性内积计算用户与商品之间评论特征向量的交叉注意力,实现用户和商品之间评论特征的交叉组合;再利用多头自注意力实现用户和商品对应评论特征的自组合,得到用户和商品最终的特征表示。在真实数据集上的实验结果表明,所提模型的均方误差相比其他模型下降了1.43%。

关键词: 特征组合; 编码模型; 双线性内积注意力; 多头自注意力

中图分类号: TP399

文献标志码: A

Recommendation Model for Fully Combining Review Features Based on Dual Attention

LI Jian, LIU Peng, LIU Wei

(School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: To fully combine the features of review text in recommendation systems, a recommendation model fully combining review features based on dual attention (FCRF) is proposed. First, the model encodes the review text by utilizing the encoder model, which is fine-tuned by the sentiment classification task to obtain the corresponding feature embedding. Secondly, to realize the cross combination between user features and item features, the bilinear inner product is adopted to calculate the cross attention of review feature embeddings between the user and the item. Thirdly, to obtain the final feature representation of the user and the item, the multi-head self-attention is adopted to realize the self-combination of the user's review features and the item's review features. Experimental results on four real-world datasets show that the mean square error of FCRF is 1.43% lower than that of the state-of-the-art model, which verifies the effectiveness of FCRF.

Key words: feature combination; encoder model; bilinear inner product attention; multi-head self-attention

因为涉及用户和商品之间的交互,通常需要充分考虑推荐系统的用户特征与商品特征之间的交叉组合,如在推荐领域广泛应用的因式分解机^[1]模型,即是在逻辑回归的基础上引入用户与商品之间

特征的二阶交叉组合,从而提高推荐效果。受因式分解机(FM, factor machine)^[1]启发,推荐领域内的大多数模型也都通过不同的特征组合方式来提升推荐效果。与此同时,随着推荐系统中业务和数据的

收稿日期: 2020-11-20

基金项目: 国家自然科学基金项目(U1636106); 北京市自然科学基金项目(4182006)

作者简介: 李 剑(1976—), 男, 教授, 博士生导师, E-mail: lijian@bupt.edu.cn.

复杂性日益增加,许多研究开始关注如何挖掘更多的有效特征,同时得益于自然语言处理技术的发展,推荐系统中的文本信息开始成为模型特征的重要来源,其中包括用户和商品的评论文本,因为用户对应的评论包好了用户对不同商品的不同属性的情感倾向,商品对应的评论表示在不同商品属性上所收到的不同用户的情感倾向,这种情感倾向的表示与匹配和用户商品之间的行为直接相关,因此,如何有效表示并匹配不同评论文本对应的特征是提升模型效果的关键。

对于评论文本特征的表示,有多种方法,如基于全局词向量^[2]或随机初始化的词向量进行构建,但在数据规模较大的场景下,会存在特征表达不充分的问题。而对于评论特征的匹配,由于大部分推荐模型都是基于人工构建的特征,特征的含义和下标是固定的,于是能使用对应的系数对特征进行组合,然后计算组合后特征向量之间的相似度,但在基于评论实现推荐的场景中,评论特征的含义是隐式的,无法对应固定的系数从而实现特征的组合,因此如何有效地组合评论特征是提升评论特征匹配效果的关键。

针对以上问题,提出一种利用双重注意力实现特征充分组合的推荐模型(FCRF, recommendation model for fully combining review features based on dual attention),启发于因式分解机模型中特征交叉组合的思想,利用注意力实现用户与商品对应评论特征之间充分的交叉组合。主要贡献包括如下4点:

1) 用BERT(bidirectional encoder representations from transformers)^[3]模型实现对输入评论文本的编码,并使用情感分类任务微调BERT模型;

2) 实现用户和商品之间评论特征的交叉;

3) 针对评论特征组合的方式,提出先交叉组合后自组合的简易思路,其中交叉组合可实现用户和商品之间评论特征的初步交叉,自组合可实现更高阶的组合;

4) 在真实数据集上的实验结果表明,相比当前基于评论的最优推荐模型,FCRF的均方误差下降了1.43%,证实了FCRF的优越性,并且通过组件分析,证实了FCRF中各个模块的有效性。

1 相关工作

随着深度学习在各个领域取得突破性进展,推荐领域也结合深度学习对相关模型进行了改进和创

新,如基于经典的因式分解机模型的各种改进模型^[4],就是在特征交叉思路的基础上,使用深度学习对特征实现了更高阶更有效的特征交叉组合。同时,除了研究有效的特征组合,基于深度学习表示特征也是一个重要研究方向。对此,有研究开始基于评论让神经网络自动地学习文本中隐含的特征。因此,目前基于评论实现推荐的相关工作主要分为两方面:一是如何有效地表示评论特征;二是如何有效地组合评论特征。

关于如何表示评论特征,有2种思路:第1种是在词的维度上拼接所有评论^[5-6];第2种是在句子维度上拼接所有评论^[7-8]。对于拼接后的评论,用预训练词向量对评论进行初步编码。而为了提取评论中词的上下文信息,Liu等^[6]使用双向长短期记忆网络对评论的词向量序列进行再编码。此外,在广义文本表示的研究方向上,近年来有许多卓有成效的研究,如使用BERT生成文本的句向量,在多个下游任务都取得了较大提升。

关于如何组合评论特征,有学者引入注意力机制计算用户(商品)评论特征向量之间的注意力^[7],从而实现用户(商品)对应评论特征的自组合以及计算用户与商品对应评论特征向量之间的交叉注意力^[6,8],但该交叉注意力只作为权重用于用户或商品评论特征的自组合,没有用于用户评论特征与商品评论特征的交叉。而在自然语言处理领域,关于文本特征组合的方式,已有大量研究,如基于向量内积计算词向量之间的注意力从而用于交叉组合^[9],使用多头自注意力对词向量进行自组合^[10]以及使用卷积神经网络计算文本中每个词邻域上的注意力从而用于抽取上下文等^[11],这些关于文本特征组合的研究结果还没有被充分地引入推荐领域对评论文本的处理中。

此外,Sachdeva等^[12]通过对比经典推荐模型与基于评论实现推荐的模型,验证了评论对于推荐效果的提升,最终证实了在用户商品交互数据较为稠密的情况下,在推荐各项指标上有显著提升。笔者基于交互数据较为稠密的Amazon 5-core数据集和豆瓣电影数据集进行了实验,其用户与商品对应的评论数最少为5,且所设计模型的目标场景也是评论信息较为丰富的推荐场景,因此能保证实验结果的有效性,也证实了模型的优越性。

2 FCRF 算法描述

给定用户和商品各自对应的所有评论 $R^u =$

$\{R_0^u, \dots, R_i^u, \dots, R_n^u\}$ 与 $R^v = \{R_0^v, \dots, R_i^v, \dots, R_n^v\}$, 对于某一条评论, $R_i^u = \{w_0^u, \dots, w_j^u, \dots, w_l^u\}$, $R_i^v = \{w_0^v, \dots, w_j^v, \dots, w_l^v\}$, 其中 $w_j^u, w_l^v \in \mathbb{R}$, 表示评论中某个词的序号, 模型计算包含如下4个步骤.

1) 将 R^u 和 R^v 输入 BERT 编码层, 得到用户和商品对应评论的句向量 $r^u = \{r_0^u, \dots, r_i^u, \dots, r_n^u\}$ 与 $r^v = \{r_0^v, \dots, r_j^v, \dots, r_n^v\}$, 其中 $r^u, r^v \in \mathbb{R}^{n \times d}$, $r_i^u, r_j^v \in \mathbb{R}^d$, d 表示句向量的维度.

2) 将 r^u 与 r^v 输入特征交叉组合层, 得到 $\hat{r}^u = \{\hat{r}_0^u, \dots, \hat{r}_i^u, \dots, \hat{r}_n^u\}$ 与 $\hat{r}^v = \{\hat{r}_0^v, \dots, \hat{r}_j^v, \dots, \hat{r}_n^v\}$.

3) 在特征自组合层, 先将 \hat{r}^u, \hat{r}^v 与 r^u, r^v 在句子维度相连得到 e^u, e^v , 再将 e^u, e^v 进行自组合, 得到基于评论的表示向量 u^r 和 v^r , $u^r, v^r \in \mathbb{R}^d$.

4) 将 u^r, v^r 与对应的标识向量 u^i, v^i 相连后输入神经因式分解机层, 得到预估评分 y , 其中 $u^i, v^i \in \mathbb{R}^d$.

2.1 BERT 编码层

如图1编码层所示, 将多个评论文本经过 BERT 编码层得到每条评论对应的句向量, 有

$$r^u = f_{\text{BERT}}(R^u) \quad (1)$$

在编码层中使用的 BERT 模型为基于官方提供的预训练模型, 使用 Amazon 5-core 构建情感分类数

据集, 通过情感分类任务微调后得到的模型.

具体而言, 基于情感分类任务微调的模型为在 BERT 模型输出层后面添加一层全连接层, 以单条评论为输入, 对应的评分为标签的分类模型. 使用时首先将谷歌官方提供的预训练 BERT 模型的参数加载至情感分类模型的 BERT 部分, 然后利用情感分类数据集训练分类模型, 当分类模型充分训练后, 再用模型中 BERT 模块的参数去初始化 FRCF 对应的 BERT 模型编码层, 从而实现 BERT 模型编码层的参数初始化.

2.2 特征交叉组合层

特征交叉组合的目的是获取同时包含用户与商品评论信息的特征表示, 从而构造用户、商品之间更高阶的特征, 对比已有研究, 交叉注意力都用于用户或商品对应评论特征的自组合^[7,9], FRCF 首次实现了基于交叉注意力实现用户和商品评论特征的相互交叉.

对于用户与商品之间评论特征的交叉组合, 由于特征向量每一维含义是隐式的, 不同特征向量在同一维度上的含义并不严格对齐, 所以不适用因式分解机模型对2个向量计算元素积从而实现特征交叉的思路. 对此, 借鉴通过双线性内积计算特征向量之间交叉注意力的方法, 并基于交叉注意力用商品(用户)特征去动态地表示用户(商品), 实现特征初步交叉, 再在特征自组合层中, 实现特征更高阶的交叉组合.

给定用户和商品的评论特征向量 r^u 与 r^v , 使用双线性内积算交叉注意力, 有

$$a = r^u W^u (r^v W^v)^T = r^u W (r^v)^T \quad (2)$$

其中: $a \in \mathbb{R}^{n \times n}$, $W \in \mathbb{R}^{d \times d}$. W 隐式地包含了用户对评论特征的变换矩阵 W^u 与商品对应评论特征的变换矩阵 W^v ; 然后, 基于交叉注意力矩阵 a 计算用户和商品对应评论特征的交叉表示. 利用商品特征所表示的用户特征向量为

$$\hat{r}_i^u = \sum_{j=1}^n \frac{\exp(a_{ij})}{\sum_{k=1}^n \exp(a_{ik})} r_j^v, \forall i \in [1, 2, \dots, n] \quad (3)$$

利用用户特征所表示的商品特征向量为

$$\hat{r}_j^v = \sum_{i=1}^n \frac{\exp(a_{ij})}{\sum_{k=1}^n \exp(a_{kj})} r_i^u, \forall j \in [1, 2, \dots, n] \quad (4)$$

2.3 特征自组合层

特征自组合层对用户和商品交叉后的评论特征

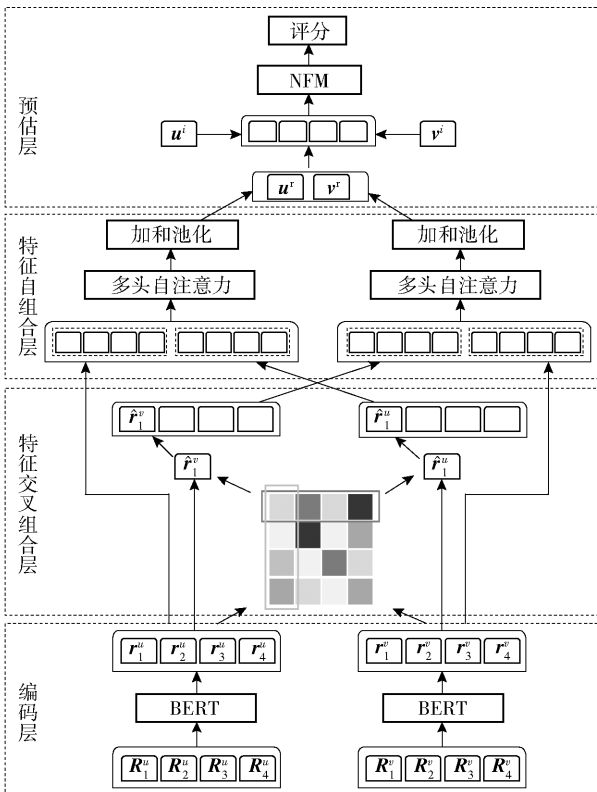


图1 FRCF 模型结构

实现自组合,从而使得每条评论特征向量能包含所有评论的特征信息,依据的直觉为理解用户(商品)的某一条评论需要基于用户(商品)所有评论综合推断.

在特征自组合层,分开计算用户和商品对应的评论特征向量且步骤相同,所以以用户评论特征向量的计算为例,首先将特征交叉组合层交叉后与交叉前的评论特征向量相连,有

$$\mathbf{e}^u = \{\mathbf{r}_0^u, \dots, \mathbf{r}_n^u, \hat{\mathbf{r}}_0^u, \dots, \hat{\mathbf{r}}_n^u\} \quad (5)$$

然后使用多头自注意力计算用户对应所有特征向量的自注意力,并基于自注意力进行自组合.

$$\begin{aligned} \mathbf{Q}^u &= \mathbf{e}^u \mathbf{W}_Q^u \\ \mathbf{K}^u &= \mathbf{e}^u \mathbf{W}_K^u \\ \mathbf{V}^u &= \mathbf{e}^u \mathbf{W}_V^u \end{aligned} \quad (6)$$

$$\mathbf{Z}^u = \text{softmax} \left(\frac{\mathbf{Q}^u \mathbf{K}^{u^T}}{\sqrt{d_k}} \right) \mathbf{V}^u \quad (7)$$

其中: $\mathbf{W}_Q^u, \mathbf{W}_K^u, \mathbf{W}_V^u \in \mathbb{R}^{d \times (n_h \times d_h)}$; $\mathbf{Z}^u \in \mathbb{R}^{2n \times (n_h \times d_h)}$, n_h 为多头自注意力的头数,其隐式地包含在参数矩阵的第 2 维中. 将 \mathbf{Z}^u 输入给前馈神经网络层,有

$$\text{FFN}(\mathbf{Z}^u) = \max(0, \mathbf{Z}^u \mathbf{W}_1^u + \mathbf{b}_1^u) \mathbf{W}_2^u + \mathbf{b}_2^u \quad (8)$$

再对以上自组合后的特征向量使用加和池化:

$$\mathbf{u}^r = \sum_i^{2n} \text{FFN}(\mathbf{Z}^u)_i \quad (9)$$

最终得到基于评论所表示的用户向量 \mathbf{u}^r . 基于评论所表示的商品向量 \mathbf{v}^r 的计算与之类似.

如前所述,针对评论特征的组合提出了一种先交叉组合后自组合的简易思路,为了证明该思路的有效性,对特征交叉组合层和自组合层分别进行了另外一种实现. 具体地,在特征交叉组合层使用加性注意力而不是双线性内积来计算交叉注意力,在特征自组合层使用文本卷积而不是多头自注意力来实现评论特征的自组合,这里不详细描述以上 2 种方法,在实验部分会进行对比和分析.

2.4 特征融合与评分预估

将基于 2.3 节得到的用户和商品的特征向量 \mathbf{u}^r , \mathbf{v}^r 与用户和商品对应的标识向量 \mathbf{u}^i 和 \mathbf{v}^i 相连,有

$$\mathbf{o} = [\mathbf{u}^r, \mathbf{v}^r, \mathbf{u}^i, \mathbf{v}^i] \quad (10)$$

其中 $\mathbf{o} \in \mathbb{R}^{2d}$, 为用户商品特征向量. 基于 NFM 对 \mathbf{o} 进行计算,得到最终的评分预估,NFM 的计算为

$$\hat{y}_{U,V} = \mathbf{m}_0 + \sum_{i=1}^{|\mathbf{o}|} \mathbf{m}_i \mathbf{o}_i + \hat{f}(\mathbf{o}) \quad (11)$$

其中: U, V 表示有交互行为用户商品集合 S 中的用户商品对, $\hat{f}(\mathbf{o})$ 为神经因式分解机中的特征高阶组

合项,有

$$\hat{f}(\mathbf{o}) = \mathbf{h}^T \sigma_L(\mathbf{W}_L(\dots \sigma_1(\mathbf{W}_1 \mathbf{f}(\mathbf{o}) + \mathbf{b}_1) \dots) + \mathbf{b}_L) \quad (12)$$

$$f(\mathbf{o}) = \frac{1}{2} \left[\left(\sum_{i=1}^{|\mathbf{o}|} \mathbf{o}_i \mathbf{c}_i \right)^2 - \sum_{i=1}^{|\mathbf{o}|} (\mathbf{o}_i \mathbf{c}_i)^2 \right] \quad (13)$$

最终使用均方误差训练模型:

$$L = \frac{1}{|S|} \sum_{(U,V) \in S} (\hat{y}_{U,V} - y_{U,V})^2 \quad (14)$$

3 实验及分析

3.1 数据集

为综合分析中英文评论下模型的效果,基于 Amazon 5-core 数据集 (<http://jmcauley.ucsd.edu/data/amazon/>) 中的电子音乐、电子游戏、工具和家具装修数据集以及豆瓣电影数据集 (<http://movie-data.csuldw.com/>) 的设计和实验,并参照 NARRE (neural attentional rating regression with review-level explanations)^[8] 进行数据预处理,然后进行过滤.

1) 为提取完整的评论信息,同时考虑模型大小,过滤掉文本长度大于 50 个中文字符的评论.

2) 对于 Amazon 5-core 数据集,过滤掉评论数小于 2 的用户或商品. 对于豆瓣电影数据集,过滤掉评论数小于 5 的用户和电影. 统计结果如表 1 所示.

表 1 数据集统计

统计指标	电子音乐	电子游戏	工具	豆瓣电影
用户数/个	15 695	10 518	9 442	157 830
商品数/个	11 643	6 945	7 275	47 830
评论数/个	144 526	50 685	40 059	975 285
评论/用户	9.2	4.8	4.2	6.2
评论/商品	12.4	7.3	6.9	20.4

按照 80%, 10%, 10% 的比例将上述数据集划分为数据集 A, B, C , 为了避免单一测试集评判的波动,让数据集 B 和 C 互为验证、测试集,数据集 A 作为训练集. 此外,抽取训练集 A 中的评论以及评分,作为用于 BERT 微调的情感分类数据集.

3.2 参数设置

对于 FCRF,在亚马逊数据集上,使用谷歌官方提供的精简版 BERT 模型初始化的 BERT 编码层,在豆瓣电影数据集上,使用第三方训练的 BERT (<https://github.com/ymcui/Chinese-BERT-wwm>) 初始化 BERT 编码层,输出句向量的维度都为 128,用

户和商品对应的最大评论数量为 20, ID 向量维度为 50, 模型中间网络层向量维度为 50, 学习率为 0.001, 批大小为 20, Dropout 系数为 0.8, L2 正则系数为 10^{-6} .

对于对比模型, 文本编码使用 GloVe 词向量, 输出维度为 300, 其他参数与 FCRF 一致.

3.3 对比实验

结合当前主流的协同过滤推荐模型与基于评论的推荐模型, 选取如下 3 类对比模型.

1) 协同过滤模型. 概率矩阵分解 (PMF, probabilistic matrix factor) 模型, 改良版的奇异值分解模型 (SVD + +, singular value decomposition plus).

2) 以文章形式的评论文本作为输入的模型. 该类模型将用户或商品对应的评论在词级别相连组成文章, 然后编码成文章向量作为输入. 该类模型选取 DeepCoNN (deep co-attention neural network) 模型和 DAML (dual attention mutual learning) 模型.

3) 以句子形式的评论序列作为输入的模型. 将每条评论单独编码得到句向量作为输入. 该类模型选取经典的 NARRE 模型与最新的 NRCA (neural unified review recommendation with cross attention) 模型作为对比模型. FCRF 也属于该类模型.

如 2.3 节所述, 为验证评论特征组合思路的有效性, 加入了由加性注意力实现特征交叉的 FCRF (FCRF_{add}) 以及由文本卷积实现特征自组合的 FCRF (FCRF_{conv}) 进行实验. 实验结果如表 2 所示.

表 2 不同模型下的均方误差对比

模型	电子音乐	电子游戏	工具	豆瓣电影	平均
PMF	0.343 8	1.957 0	1.882 2	0.885 4	1.267 1
SVD + +	0.433 3	1.160 3	0.837 0	0.674 3	0.777 6
DeepCoNN	0.413 5	1.188 1	0.850 5	0.648 1	0.775 1
DAML	0.266 0	1.019 0	0.788 7	0.554 7	0.657 1
NARRE	0.278 5	1.108 6	1.072 1	0.592 0	0.762 8
NRCA	0.278 5	1.041 6	0.809 5	0.563 2	0.673 2
FCRF _{add}	0.261 9	1.004 8	0.782 0	0.549 8	0.649 6
FCRF _{conv}	0.265 3	1.000 1	0.785 6	0.552 3	0.650 8
FCRF	0.2623	1.003 6	0.781 6	0.543 1	0.647 7
提升%	1.39%	1.51%	0.77%	2.14%	1.43%

根据实验结果可得, 利用评论文本的模型相比协同过滤模型 PMF、SVD + + 有显著提升, 使用注意力的模型明显优于未使用注意力的模型 DeepCoNN, 而效果最好的现有模型为基于文章向量的

DAML 模型, 优于基于句向量序列的模型 NARRE 和 NRCA 模型, 但 FCRF 作为基于句向量序列的模型, 其评判效果相比 DAML 模型提升了 1.43%, 相比 NRCA 模型提升 3.79%, 一方面说明目前基于句向量序列的模型对于抽取后评论特征向量的组合无效; 另一方面也证实了 FCRF 模型设计的有效性与优异性.

此外, 对于 FCRF_{add} 和 FCRF_{conv}, 其模型效果虽然稍逊于 FCRF, 但相比于现有最优 DAML 模型仍有显著优势, 证实了对于评论特征向量先交叉组合后自组合的简易思路的有效性和鲁棒性, 也说明了特征交叉组合层和特征自组合层具体实现时的灵活性. 随着关于文本之间注意力机制的相关研究进一步发展, FCRF 还有相应的提升空间.

值得一提的是, FCRF 在中文的豆瓣电影数据集上均方误差的提升 (2.14%) 普遍高于在英文的 Amazon 5-core 数据集上的提升, 一方面说明 FCRF 在中文评论场景下的有效性; 另一方面也说明了 FCRF 简易的特征组合思想使模型针对不同场景有较好的泛化性.

3.4 推理速度与内存

由于 FCRF 使用 BERT 模型作为编码层, 降低了推理速度并增加了内存占用. 选取表 2 中效果较好的 DAML, NRCA 模型和 FCRF 模型, 其中 FCRF 分为使用 BERT 编码层和不使用 BERT 编码层 (使用 GloVe 编码). 评判指标包括平均推理时间、平均消耗内存和 4 个数据集上的均方误差, 显卡为 GeForce GTX 1060, 实验结果如表 3 所示.

表 3 推理速度与内存

模型	推理时间/ms	内存/GB	MSE
DAML	51.0	3.5	0.657 1
NRCA	28.2	2.9	0.673 2
FCRF-GloVe	33.0	1.2	0.651 1
FCRF	80.3	4.7	0.647 7

由表 3 可得, 因为引入 BERT 编码层 FCRF 在推理时间和内存上不如 DAML 和 NRCA 模型, 但在去除 BERT 编码层后, FCRF 在模型效果依然优于 DAML 和 NRCA 模型的情况下, 推理时间和内存都有较大优势.

3.5 组件分析

FCRF 的主要创新点在于引入 BERT 编码层和特征组合层, 为了验证各模块有效性, 依次去除

BERT 层(-BERT)、BERT 情感微调(-Senti pre)、特征交叉组合层(-Interact att)、特征自组合层(-Self att)后模型的效果,实验结果见表4.

表4 FCRF 去除组件后的均方误差

模型	电子 音乐	电子 游戏	工具	豆瓣 电影	平均/下降
-BERT	0.266 7	1.017 4	0.786 9	0.549 9	0.655 2/0.011 6
-Senti pre	0.264 3	1.018 1	0.786 1	0.548 6	0.654 3/0.010 2
-Interact att	0.263 5	1.018 9	0.797 3	0.557 3	0.659 2/0.017 8
-Self att	0.264 5	1.033 4	0.797 5	0.560 1	0.663 9/0.025 0
FCRF	0.262 3	1.003 6	0.781 6	0.543 1	0.647 7

由实验结果可见,去除 BERT 情感微调后,模型效果下降 1.02%,而如果完全将 BERT 编码层替换为 GloVe 编码,模型效果只下降 1.16%,证实了情感微调的必要性;对于特征组合层,去除特征交叉组合层后,模型效果下降 1.78%,去除特征自组合层后,下降 2.50%,验证了特征组合层的必要性和有效性,尤其是特征自组合层,由于其对交叉后的评论特征向量实现高阶自组合,对模型的影响最大.

此外,在完全去除 BERT 编码层后 FCRF 在 4 个数据集上的均方误差为 0.655 2,优于 DAML 模型的 0.657 1,进一步验证了在不使用 BERT 编码层的情况下,特征组合层的优异效果.

4 结束语

受 FM 模型对特征同时使用一阶线性组合与二阶线性交叉思想的启发,提出一种利用双重注意力充分组合评论特征的推荐模型 FCRF,对用户与商品对应评论特征实现充分的交叉组合与自组合.此外,引入 BERT 预训练模型,并基于情感分类任务进行微调,从而增强对评论文本的初始编码.在多个包含评论的推荐数据集上,FCRF 取得了最为优异的实验效果,证实了模型设计的有效性与先进性.

参考文献:

- [1] Rendle S. Factorization machines with libFM[J]. ACM Transactions on Intelligent Systems and Technology, 2012, 3(3): 1-22.
- [2] Pennington J, Socher R, Manning C. Glove: global vectors for word representation[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics, 2014: 1532-1543.
- [3] Devlin J, Chang Mingwei, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//NAACL-HLT. Minneapolis: ACM, 2019: 4171-4186.
- [4] Guo Huifeng, Tang Ruiming, Ye Yunming, et al. DeepFM: a factorization-machine based neural network for CTR prediction[C]//Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence. Melbourne: International Joint Conferences on Artificial Intelligence Organization, 2017: 1725-1731.
- [5] Zheng Lei, Noroozi V, Yu P S. Joint deep modeling of users and items using reviews for recommendation[C]//Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge United Kingdom: ACM, 2017: 425-434.
- [6] Liu Donghua, Li Jing, Du Bo, et al. DAML: dual attention mutual learning between ratings and reviews for item recommendation[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. [S.l.]: ACM, 2019: 344-352.
- [7] Chen Chong, Zhang Min, Liu Yiqun, et al. Neural attentional rating regression with review-level explanations[C]//Proceedings of the 2018 World Wide Web Conference on World Wide Web-WWW'18. Lyon: ACM Press, 2018: 1583-1592.
- [8] Liu Hongtao, Wang Wenjun, Xu Hongyan, et al. Neural unified review recommendation with cross attention[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.]: ACM, 2020: 1789-1792.
- [9] Chen Qian, Zhu Xiaodan, Ling Zhenhua, et al. Enhanced LSTM for natural language inference[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics(ACL). Vancouver: [s. n.], 2017: 17-29.
- [10] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach: [s. n.], 2017: 5998-6008.
- [11] Pang L, Lan Y, Guo J, et al. Text matching as image recognition[C]//2016 Association for the Advancement of Artificial Intelligence(AAAI2016). Phoenix Arizona: AAAI Press, 2016: 2793-2799.
- [12] Sachdeva N, McAuley J. How useful are reviews for recommendation? a critical review and potential improvements // Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.]: ACM, 2020: 1845-1848.