

基于深度神经网络的个性化混合商品推荐模型^①



倪美玉, 曹为刚

(浙江科贸职业技术学院, 金华 321019)
通讯作者: 曹为刚, E-mail: ppcadea@163.com

摘要: 针对推荐系统广泛采用的协同过滤算法存在的稀疏性和冷启动问题, 提出了一种基于深度神经网络和动态协同滤波的推荐模型. 该模型采用预训练 BERT 模型结合双向 GRU 从用户和商品评论中提取隐含特征向量, 利用耦合 CNN 构建评分预测矩阵, 通过动态协同滤波融入用户兴趣变化的时间特征. 在亚马逊公开数据集上进行实验, 结果表明该模型提高了商品评分预测的准确性.

关键词: 商品推荐; BERT 模型; 双向 GRU; 耦合 CNN; 动态协同滤波

引用格式: 倪美玉, 曹为刚. 基于深度神经网络的个性化混合商品推荐模型. 计算机系统应用, 2021, 30(5): 184-189. <http://www.c-s-a.org.cn/1003-3254/7898.html>

Personalized Hybrid Recommendation Model Based on Deep Neural Network

NI Mei-Yu, CAO Wei-Gang

(Zhejiang Vocational College of Science and Trade, Jinhua 321019, China)

Abstract: Collaborative filtering algorithm widely used in the recommendation systems has the problems of sparseness and cold start. For this reason, a recommendation model based on deep neural networks and dynamic collaborative filtering is proposed in this study. The model combines a pre-trained BERT model with bidirectional GRU to extract hidden feature vectors from users and commodity reviews. Furthermore, coupled CNN is used to construct the score prediction matrix and the temporal changes in user interests are incorporated through dynamic collaborative filtering. Finally, the experiments on an Amazon's data set show that the proposed model increases the accuracy of commodity score prediction.

Key words: commodity recommendation; BERT model; bidirectional GRU; coupled CNN; dynamic collaborative filtering

随着信息技术和互联网技术的飞速发展, 数据规模急剧增加, 信息过载严重, 推荐系统是解决问题重要技术手段^[1,2]. 当前, 在各大电商平台中推荐系统中推荐算法的好坏对电子商务发展有着十分深远的影响. 个性化推荐是根据用户的需求, 从海量信息中挖掘出用户感兴趣的商品或服务, 并按照评分预测结果进行个性化展示. 推荐算法是个性化推荐的核心, 直接决定了推荐系统的推荐性能, 成为当前研究的热点问题^[3,4].

协同过滤是目前个性化推荐系统普遍采用的技术,

其基本思想是利用过去具有相同兴趣的用户会在未来选择相似的商品. 典型的协同过滤类算法是基于矩阵分解的协同过滤推荐模型, 该模型首先分别表征用户和商品特征, 然后借助用户和商品特征向量的点积预测评分矩阵, 具有较好的推荐性能, 但评分数据稀疏性问题一直制约传统协同过滤的瓶颈^[5]. 深度学习技术近年来被用于解决个性化商品推荐问题, 基于深度学习的混合推荐模型被相继提出, 例如 DeepCoNN^[6]、DLALSTM^[7]、NARRE^[8] 等. 这些模型均是首先利用深度学习技术将

① 收稿时间: 2020-09-10; 修改时间: 2020-10-09; 采用时间: 2020-10-13; csa 在线出版时间: 2021-04-28

用户对商品的评论数据变换到隐空间, 得出评分矩阵后再利用传统推荐模型完成个性化推荐, 但静态词向量编码影响了评分矩阵估计性能, 降低了推荐系统的准确性.

本文提出了一种基于深度神经网络和动态协同过滤的混合个性化商品推荐模型. 文章后续章节安排如下: 第1节总体介绍了文中混合个性化推荐模型架构, 第2节给出了基于深度学习的用户与商品隐含特征提取方法, 第3节构建了基于动态协同过滤的商品推荐模型, 第4节通过实验验证了推荐模型的参数设置和推荐性能, 最后是总结.

1 本文推荐模型

目前协同过滤的商品推荐算法的关键是对商品评分的预测, 主要存在以下3个方面问题. 首先, 评分矩阵稀疏性问题. 评论数据稀疏性是影响个性化商品推荐系统最终评分预测准确性的关键因素, 这是由于平台中的商品数量远远大于单个用户的评论数量, 导致评论数据呈现出十分明显的稀疏性, 进而影响用户与商品、用户与用户之间的关系量不足, 降低商品评分预测结果的准确性. 其次, 冷启动问题. 引发推荐系统冷启动主要有两个原因, 即新用户和新商品. 在商品推荐平台中, 每天都有新用户注册, 同样也有新商品上架. 系统无法对新商品或新用户的兴趣程度. 最后, 信息过期问题. 随着时间的推移, 用户的兴趣偏好或者商品的受欢迎程度均会出现变化, 而传统推荐算法没有考虑这种变化带来的影响.

针对上述问题, 本文提出了基于卷积神经网络和动态协同过滤的个性化商品推荐模型, 模型架构如图1所示.

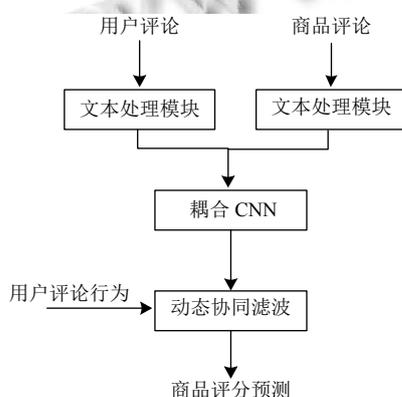


图1 推荐模型架构

首先将用户评论和商品评论文本信息用BERT模型和双向GRU进行处理, 利用并行卷积神经网络分别提取用户和商品的隐含特征向量; 然后, 引入用户对商品的评分行为, 采用TimeSVD++算法实现最终的商品评分预测. 该模型通过卷积神经网络在捕获用户和商品的隐含特征向量, 以降低用户对商品评分矩阵预测过程中存在的数据稀疏性和冷启动问题; 在协同过滤过程中融入时间序列, 降低商品推荐过程中信息过程带来的影响.

2 基于深度学习的用户与商品隐含特征提取

2.1 评论文本处理

词袋模型是典型的词分法评论文本处理模型, 但该模型对文本中的词语顺序不敏感, 经常将意思完全相反的文本用相同的词向量表示^[9]. 并且评论文本的噪声大、网络用语多也严重影响了词分法的文本处理性能. 为此, 文中采用双向GRU的文本处理模型^[10], 如图2所示. 由于用户评论和商品评论是具有相似结构的两个平行模型, 因此只介绍商品评论文本的处理过程.

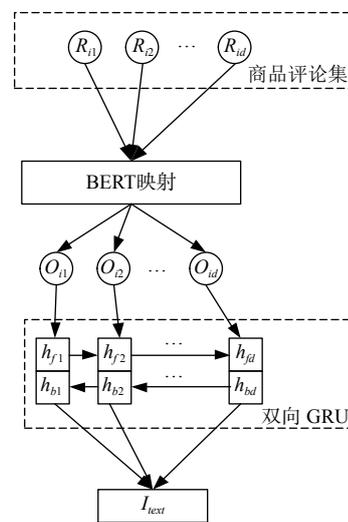


图2 商品评论处理过程

令 I 表示某商品的评论集:

$$I = \{R_{i1}, R_{i2}, \dots, R_{id}\} \tag{1}$$

其中, $R_{ij}(j = 1, 2, \dots, d)$ 表示单条评论, d 表示模型设置的最大评论个数. 将商品评论数据输入到BERT映射模型^[5], 将单条评论数据转换为向量形式:

$$c = \{O_{i1}, O_{i2}, \dots, O_{id}\} \tag{2}$$

在BERT映射过程中, 如果商品评论个数大于 d ,

则只映射该商品的前 d 条评论; 如果商品评论个数小于 d , 则通过填充零向量使映射后向量长度等于 d .

考虑到评论之间的内在联系, 采用双向 GRU 对 BERT 映射后的向量记性编码. 编码过程包括前向和后续两部分, 其中前向按照正序处理 ($O_{i1} \rightarrow O_{id}$), 后向按照逆序处理 ($O_{id} \rightarrow O_{i1}$), 分别获得编码后的序列 h_{fi} 和 h_{bi} . 接下来, 通过拼接获取评论向量特征:

$$h_i = h_{fi} \oplus h_{bi} \quad (3)$$

经过双向 GRU 编码后的特征向量可以表示为:

$$H_0 = \{h_1, h_2, \dots, h_d\} \quad (4)$$

已知每一条商品评论的内容均具有各自特征, 即均隐含了对商品的不同偏好的贡献, 因此采用自注意力机制对特征向量进行信息汇总, 汇总后得出的注意力向量可以表示为:

$$a = \text{softmax}(\omega_1 \times \tanh(\omega_2 \times H_0^T)) \quad (5)$$

其中, ω_1 和 ω_2 均表示维度符合要求的权值参数, $\text{softmax}()$ 表示基于注意力机制的归一化处理. 注意力向量中包含了评论对商品偏好的贡献, 将同一商品的隐藏特征向量进行加权求和后, 并后接一个全连接层, 即可得出评论集 u 对商品的最终隐含特征向量

$$I_{\text{text}} = W' \times aH_0 + b' \quad (6)$$

其中, W' 和 b' 分别表示全连接层的权值参数和偏置参数. 同理, 可以按照相同方式获得评论集对用户的最终隐含特征向量 U_{text} .

2.2 基于卷积神经网络的评分预测

按照上一节文本处理方式分别获得用户和商品的偏好表达式后, 本节首先基于卷积神经网络提取用户和商品的深层次特征, 并利用共享层将其映射到相同的特征空间后进行评分矩阵预测.

将最终隐含特征向量输入到卷积神经网络以提取其深层次特征, 假设卷积神经网络的卷积层的神经元个数为 m , 第 j 个神经元的卷积核为 $K_j \in R^{c \times d}$, 其中 d 为隐含特征向量的维度. 对于隐含特征向量 I_{text} (以商品为例), 各个卷积核的输出为:

$$\kappa_j = G(I_{\text{text}} * K_j + b_j) \quad (7)$$

其中, $G(\cdot)$ 和 b_j 分别表示卷积核的激活函数和偏置项, 文中激活函数采用线性整流函数 ReLU, 具体为:

$$G(x) = \max\{0, x\} \quad (8)$$

ReLU 能有效避免训练过程中出现梯度消失和梯

度爆炸现象, 并且能够显著提升随机梯度下降算法的收敛速度, 降低模型训练时间^[11].

接下来, 对卷积运算后的特征向量进行池化操作, 采用最大池化法提取特征向量中的最大值作为特定卷积内核的特征. 经过最大池化操作后, 卷积层的输出结果被转化为长度固定的向量, 池化操作可以表示为:

$$R_i = \max\{\kappa_1, \kappa_2, \dots, \kappa_{m-d+1}\} \quad (9)$$

根据上述卷积操作与池化操作过程可知, 卷积层将输出的特征向量经过最大池化操作后, 文本信息中特征最强的信息被筛选出来, 而特征较弱的信息在池化操作过程中被滤除. 为了保证文本信息的完整性, 文中采用多个并行卷积神经网络来获得文本信息的各种特征, 这是因为多个并行的卷积层能够提取出不同的局部信息, 并行卷积层的输出特征集合可以表示为:

$$R = \{R_1, R_2, \dots, R_M\} \quad (10)$$

其中, M 表示并行卷积层中卷积核的个数.

池化层的输出被导入全连接层, 令 W 表示全连接层的权值矩阵, 则全连接层的输出可以表示为:

$$X_I = f(W \times R + b) \quad (11)$$

其中, $f(\cdot)$ 和 b 分别为全连接层的激活函数和偏置项. 上述是以商品为例获得的深层特征 X_I , 同样将用户评论文本输入, 能够得到用户评论的深层特征 X_U .

虽然深层特征 X_I 和 X_U 均源自于评论数据, 但是对用户和商品两个不同映射空间的特征表示, 并不具有可比性, 必须在评分预测之前将其变换到相同的特征空间中. 为此, 文中采用共享层对用户评论数据的隐含特征和商品评论数据的隐含特征进行耦合操作, 即将 X_I 和 X_U 耦合为单个特征向量 $Z = (X_I, X_U)$.

3 基于动态协同过滤的商品推荐模型

动态协同过滤算法是在 SVD++ 算法的基础上, 增加了时间序列项, 因此被称为 TimeSVD++ 算法^[12]. TimeSVD++ 算法是由简单的因子分解模型演化而来的, 假设已经获得用户对商品的评分预测矩阵 $r \in R^{N \times M}$, 其中 N 和 M 分别表示用户个数和商品个数, 矩阵中 r_{ui} 代表用户 u 对商品 i 的评分预测值.

通过卷积神经网络处理评论文本信息, 得出的用户和商品的深层特征向量分别为 X_I 和 X_U , 利用共享层耦合后并采用因子分解机得出用户对商品的评分矩阵预测值 \hat{r}_{ui} . TimeSVD++ 算法首先需要降低 \hat{r}_{ui} 和 r_{ui} 之间

的误差,获取最优的评分预测,具体过程可以表示为:

$$\hat{r}_{ui} = X_I X_U \quad (12)$$

$$\min F = \sum_{r_{ui}} (\hat{r}_{ui} - r_{ui})^2 \quad (13)$$

在评分预测过程中加入偏置项,即构成SVD模型:

$$\hat{r}_{ui} = X_I X_U + u + b_u + b_i \quad (14)$$

式中, u 为训练过程中用户对商品评分预测值的均值, b_u 和 b_i 分别表示用户偏置项和商品偏置项,表示一个用户或一个商品的评分预测值的均值。

SVD++算法在SVD模型的基础上,通过隐式反馈信息加入了用户兴趣信息。也就是说,任意一个用户只要对某个商品进行过评论,不管评论内容的评分预测值高低,均表示用户对该商品感兴趣,感兴趣的程度用隐因子来表示 $Y_j = \{Y_{j1}, Y_{j2}, \dots, Y_{jF}\}$ 。此时,用户对商品的评分预测模型修正为:

$$\hat{r}_{ui} = \left[X_U + \frac{\sum_{j \in N(u)} Y_j}{\sqrt{|N(u)|}} \right] X_I + u + b_u + b_i \quad (15)$$

式中, $N(u)$ 表示用户 u 评价过的所有商品的集合。

TimeSVD++算法的基本思路是随着时间的推移,用户对商品的喜好程度也是变化的,即式(15)中 X_U 、 b_u 和 b_i 不再是固定量,而是一个关于时间的函数,但商品特征向量 X_I 不随时间变化。由此可见,TimeSVD++算法中融入了时间序列项,因此在预测用户对商品的评分过程中需要沿着时间轴划分出多个时间段,在各个时间段内对商品评分进行预测。划分后的时间段用 $Bin(t)$ 表示, X_U 、 b_u 和 b_i 值在同一个时间段内相同,不同时间段内不同。考虑时间序列项的TimeSVD++算法评分预测过程如下:

$$\hat{r}_{ui} = \left[X_U(t) + \frac{\sum_{j \in N(u)} Y_j}{\sqrt{|N(u)|}} \right] X_I + u + b_u(t) + b_i(t) \quad (16)$$

其中, $b_i(t)$ 和 $b_u(t)$ 分别表示商品和用户在 t 时刻的偏置,均由静态部分和动态部分组成,即:

$$b_i(t) = b_i + b_{i, Bin(t)} \quad (17)$$

$$b_u(t) = b_u + \alpha_u \cdot \text{sign}(t - t_u) \cdot |t - t_u|^\beta + b_{u,t} \quad (18)$$

式中, $b_{i, Bin(t)}$ 表示商品在时间段 $Bin(t)$ 内的偏置, t_u 表示用户给出所有评分的均值。

TimeSVD++算法在获取评分预测值后,通过式(19)最小化预测值与真实值之间的误差:

$$\min F = \sum_{r_{ui}} (\hat{r}_{ui} - r_{ui})^2 + \lambda \left[\|X_U\|^2 + \|X_I\|^2 + \|b_u(t)\|^2 + \|b_i(t)\|^2 + \|Y_j\|^2 \right] \quad (19)$$

4 实验分析

4.1 实验条件

为了验证本文提出的深度神经网络和动态协同滤波(DEEP-TimeSVD++)的个性化商品推荐模型有效性,使用亚马逊公共数据集中的子数据集 Toys_and_Games 和 Instant_Video,两个数据集中约有78.3%的商品评论数量少于20条,67.1%的商品评论数量少于10条,满足用户对商品评论稀疏性要求。

后续实验中采用均方误差(MSE)作为评价模型对商品评分预测准确性的指标,该指标定义如下:

$$MSE = \frac{1}{N} \sum_{u,i} (\hat{r}_{ui} - r_{ui})^2 \quad (20)$$

式中, N 表示实验过程中样本个数。模型的MSE越小,说明模型的商品评分预测性能越好。

文中实验环境的具体参数如表1所示。

表1 实验环境参数

| 项目 | 参数 |
|------|--------------|
| 操作系统 | Win10 64位旗舰版 |
| 编程平台 | PyCharm 2019 |
| 编程语言 | Python |
| CPU | Core i7 |
| 内存 | 8 GB |
| 硬盘 | 2 TB |

4.2 文本处理性能实验

针对评论文本的评分预测问题,文中提出采用BERT映射结合双向GRU(BERT-GRU)提出评论文本特征,本节将其与不同文本处理模型进行对比测试,包括预训练词向量结合CNN模型(Word-CNN)、预训练词向量结合CNN模型(Word-GRU)、BERT映射结合CNN模型(BERT-CNN)。4种文本处理模型的性能对比结果如表2所示。

根据4种模型的文本处理性能结果可知,加入BERT映射的文本处理模型相比传统预训练词向量模型的文本处理性能得到了进一步提高,这是因为传统

词向量模型在多义词影响下不能准确给出同一词语在不同环境下的真实含义,而采用BERT映射能够有效解决该问题,提升对评论文本的语义理解能力.对比BERT-CNN和BERT-GRU可知,双向GRU模型比传统CNN模型的文本处理性能更佳,这是因为某个商品的评论集处理本质上是一个序列处理问题,而双向GRU中的RNN递归结构比CNN结构更适合于表征序列特征,能够有效提取评论文本之间的联系,提高文本处理性能.

表2 文本处理性能对比结果

| 模型 | Toys_and_Games | Instant_Video |
|----------|----------------|---------------|
| Word-CNN | 0.82 | 0.94 |
| Word-GRU | 0.78 | 0.93 |
| BERT-CNN | 0.76 | 0.935 |
| BERT-GRU | 0.74 | 0.92 |

4.3 模型参数选择实验

本节对模型涉及到的隐因子、时间段划分和学习率进行测试,分析不同参数取值对实验结果的影响.其中隐因子的测试范围为[5,10,15,20],时间段划分的测试范围为[10,15,20,25].

图3为不同隐因子下本文提出的DEEP-TimeSVD++算法MSE指标的变化情况.结果表明,当隐因子取值为15时,DEEP-TimeSVD++算法对Toys_and_Games和Instant_Video的MSE分别为0.85和0.93,均达到最小.对比图3可知,当隐因子取值为5、10和20时,DEEP-TimeSVD++算法的MSE变化并不是很大,说明DEEP-TimeSVD++算法对隐因子设置敏感性低,综合考虑推荐性能,后续实验中将隐因子a设置为15.

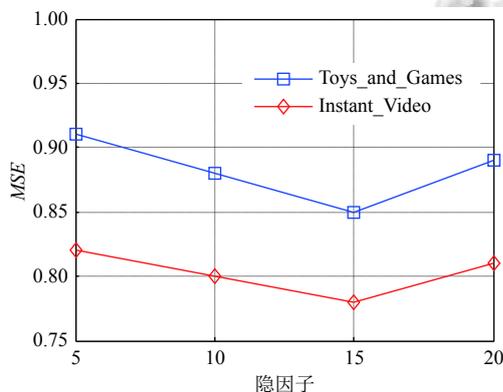


图3 隐因子对算法性能影响

图4为不同时间段划分的DEEP-TimeSVD++算法MSE指标的变化情况.结果表明,Toys_and_Games

和Instant_Video数据集在Bin(t)为20时MSE最小,为此文中Bin(t)选取为20.

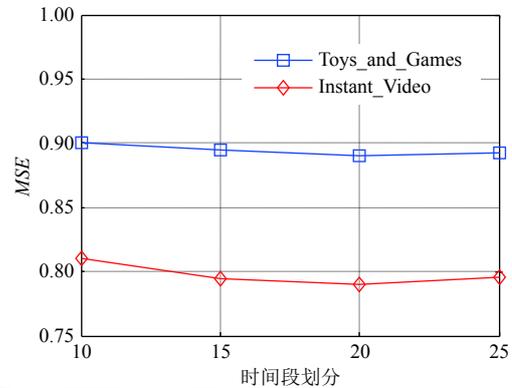


图4 时间段划分对算法性能的影响

4.4 评分预测效果对比实验

本节仍然采用亚马逊公共数据集中的Toys_and_Games和Instant_Video子集对评分预测推荐的准确性进行评价.首先采用文中提出的BERT-GRU评分预测模型处理评论文本,然后根据评分矩阵预测结果进行个性化商品推荐,分别采用因子分解机(FM)、SVD和SVD++模型进行对比测试.此外,对比方法还包括文献[13]提出的融合了边缘降噪自编码器和FM模型(mSDA-FM)的商品推荐算法.推荐评价指标除了采用MSE以外,还采用平均绝对误差(MAE),定义如下:

$$MAE = \frac{1}{N} \sum_{u,i} (\hat{r}_{ui} - r_{ui}) \quad (21)$$

分别采用本文提出的DEEP-TimeSVD++算法与各对比算法处理Toys_and_Games和Instant_Video,结果如图5和图6所示.

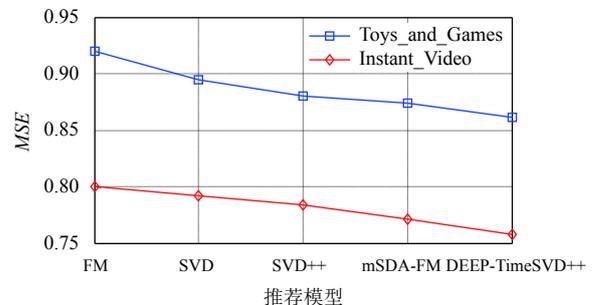


图5 MSE结果对比

实验结果表明,本文提出的DEEP-TimeSVD++算法在MSE和MAE两个性能指标上均优于对比方法.

由图5可知,在MSE指标上,DEEP-TimeSVD++算法对Toys_and_Games和Instant_Video相比性能较好的mSDA-MF分别提高了1.4%和3.8%。由图6可知,DEEP-TimeSVD++算法对Toys_and_Games和Instant_Video相比性能较好的SVD++算法在MAE指标上分别提高了2.5%和2.2%。上述实验结果证实了DEEP-TimeSVD++算法在商品推荐中具有更高的准确性。

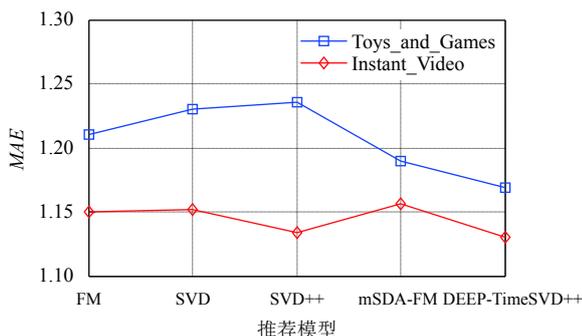


图6 MAE结果对比

5 结论

提出了一种基于深度神经网络和动态协同滤波的个性化商品推荐模型,该模型利用BERT模型和双向GRU替代传统词向量文本处理方式,能够有效降低数据稀疏性影响,实现深层隐含特征提取;采用耦合CNN网络预测评分矩阵,并基于融合时间序列项动态协同滤波实现商品评分预测。实验结果验证了该模型降低了数据稀疏性和冷启动问题的影响,考虑了用户兴趣随时间推移出现的变化,在推荐准确率上具有明显优势。

参考文献

- 1 Subramaniaswamy V, Manogaran U, Logesh R, *et al.* An ontology-driven personalized food recommendation in IoT-based healthcare system. *The Journal of Supercomputing*, 2019, 75(6): 3184–3216. [doi: 10.1007/s11227-018-2331-8]
- 2 汤伟韬,余敦辉,魏世伟.融合知识图谱与用户评论的商品推荐算法. *计算机工程*, 2020, 46(8): 93–100.
- 3 翁小兰,王志坚.协同过滤推荐算法研究进展. *计算机工程与应用*, 2018, 54(1): 25–31. [doi: 10.3778/j.issn.1002-8331.1710-0081]
- 4 成保梅.基于协同过滤的电子商务个性化推荐算法研究. *现代电子技术*, 2019, 42(20): 37–39, 44.
- 5 冯兴杰,曾云泽.基于评分矩阵与评论文本的深度推荐模型. *计算机学报*, 2020, 43(5): 884–900. [doi: 10.11897/SP.J.1016.2020.00884]
- 6 黄文明,卫万成,张健,等.基于注意力机制与评论文本深度模型的推荐方法. *计算机工程*, 2019, 45(9): 176–182.
- 7 刘慧婷,纪强,刘慧敏,等.基于双层注意力机制的联合深度推荐模型. *自然科学版*, 2020, 48(6): 97–105.
- 8 郭望,胡文心,吴雯,等.多层次场景感知评分预测研究. *计算机应用与软件*, 2019, 36(8): 145–154. [doi: 10.3969/j.issn.1000-386x.2019.08.026]
- 9 韩冬煦,常宝宝.中文分词模型的领域适应性方法. *计算机学报*, 2015, 38(2): 272–281.
- 10 梁志剑,谢红宇,安卫钢.基于BiGRU和贝叶斯分类器的文本分类. *计算机工程与设计*, 2020, 41(2): 381–385.
- 11 梁斌,刘全,徐进,等.基于多注意力卷积神经网络的特定目标情感分析. *计算机研究与发展*, 2017, 54(8): 1724–1735. [doi: 10.7544/issn1000-1239.2017.20170178]
- 12 薛峰,刘凯,王东,等.基于深度神经网络和加权隐反馈的个性化推荐. *模式识别与人工智能*, 2020, 33(4): 295–302.
- 13 Wang H, Wang NY, Yeung DY. Collaborative deep learning for recommender systems. arXiv: 1409.2944, 2015.