

# 基于注意力机制的深度学习推荐算法<sup>①</sup>



申晋祥, 鲍美英

(山西大同大学 计算机与网络工程学院, 大同 037009)

通讯作者: 申晋祥, E-mail: [836938642@qq.com](mailto:836938642@qq.com)

**摘要:** 针对目前基于评论文本的推荐算法存在文本特征和隐含信息提取能力不足的问题, 提出一种基于注意力机制的深度学习推荐算法. 通过分别构建用户和项目的评论文本表示, 利用双向门控循环单元提取文本的上下文依赖关系以获得文本特征表示, 引入注意力机制, 更准确的获取用户兴趣偏好和项目属性特征. 将生成的用户和项目评论数据的两组隐含特征分别输入全连接层处理, 再合并到同一个向量空间进行评分预测, 得到推荐结果. 在 Yelp 和 Amazon 两个公开数据集中进行实验, 结果表明所提出的算法与其他算法相比, 具有更好的推荐性能.

**关键词:** 深度学习; 双向门控循环单元; 注意力机制; 特征提取; 推荐算法

引用格式: 申晋祥, 鲍美英. 基于注意力机制的深度学习推荐算法. 计算机系统应用, 2021, 30(6): 220–225. <http://www.c-s-a.org.cn/1003-3254/7945.html>

## Deep Learning Recommendation Algorithm Based on Attention Mechanism

SHEN Jin-Xiang, BAO Mei-Ying

(College of Computer and Network Engineering, Shanxi Datong University, Datong 037009, China)

**Abstract:** This study proposes a deep learning recommendation algorithm based on attention mechanism to solve the problem that the current recommendation algorithms based on comment texts have insufficient extraction of text features and implicit information. The comment text representations of users and items are constructed, and the context dependency of texts is extracted by bidirectional gated recurrent units for text feature representations. Moreover, the attention mechanism is introduced to obtain the interest preference of users and the attribute features of items more accurately. The two sets of hidden features of the generated user and item comment data are respectively input into the fully connected layer and then merge into the same vector space for rating prediction. As a result, the recommendation results are obtained. Experiments on two public data sets, Yelp and Amazon, show that the proposed algorithm has better recommendation performance than other algorithms.

**Key words:** deep learning; bidirectional gated recurrent unit; attention mechanism; feature extraction; recommendation algorithm

## 1 引言

随着大数据、云计算以及互联网技术的快速发展, 各种网络应用使得网络数据爆炸式增长, 面对海量蕴含丰富价值的数据如何能够快速获得自己所感兴趣的

信息变得越来越不容易. 推荐系统能够从大规模数据中分析提取出用户兴趣偏好的所需信息, 为用户提供个性化推荐, 是目前广泛认可的有效解决“信息过载”问题的一种策略.

① 基金项目: 国家自然科学基金 (11971277); 山西省教育科学“十三五”规划项目 (GH-18045); 山西大同大学校级科研项目 (2017K7)

Foundation item: National Natural Science Foundation of China (11971277); Education Scientific Research Project of Shanxi Province during Thirteenth Five-Year Plan (GH-18045); Scientific Research Fund of Shanxi Datong University (2017K7)

收稿时间: 2020-10-08; 修改时间: 2020-11-02; 采用时间: 2020-11-06; csa 在线出版时间: 2021-06-01

虽然推荐系统已被广泛应用,但仍存在很多问题,如数据稀疏、冷启动、推荐多样性等问题.充分利用用户评论文本,对评论文本进行深度挖掘可以有效的提高推荐性能.针对目前推荐系统对评论文本内容潜在特征和隐含信息提取能力不足导致的推荐精度下降问题,利用深度学习对评论文本挖掘和提取的优势,并融入注意力机制.提出一种基于注意力机制的深度学习推荐算法以提高推荐系统的推荐精度.

## 2 相关研究

为解决传统推荐算法存在的评分数据稀疏问题,基于评论文本的推荐算法相继出现,利用从评论文本中提取的特征信息来提高项目的预测评分.

目前大多数基于评论文本的推荐算法<sup>[1,2]</sup>都是基于词袋模型,不能对评论文本的上下文以及单词顺序特征进行很好的提取.应用深度学习技术的推荐算法<sup>[3-7]</sup>通过卷积神经网络结合评论文本词序以及文本上下文信息,能够很好的提取评论文本的潜在语义表示进行推荐,有效的提高了推荐性能.在此基础上进一步引入注意力机制,可以加强重点部分信息的关注度并投入更多注意力资源,从而获取更多需要重点关注内容的特征信息.基于注意力机制的相关研究<sup>[8-12]</sup>通过对评论文本不同方式的建模,进一步提高文本处理效率.

基于以上研究,提出一种基于注意力机制的深度学习推荐算法(Attention-based Deep Learning User and Item, ADLUI),主要贡献如下:

(1) 分别构建用户和项目的评论文本表示,利用双向门控循环单元捕获文本中的词序信息,提取文本的上下文依赖关系以获得文本特征表示,将其输入到全连接层获取最终评论文本特征表示以进行预测评分.

(2) 在分别构建的用户和项目评论文本的双向门控循环单元中引入注意力机制,识别文本特征表示中不同特征的重要性,更准确的获取用户兴趣偏好和项目属性特征.

(3) 在 Yelp 和 Amazon 两个真实数据集中进行实验,结果表明所提出的算法与其他先进算法相比,具有更好的推荐性能.

## 3 基于注意力机制的深度学习推荐模型

评论文本中,一些单词和句子所表达的语义特征对用户的偏好和项目的特性侧重会有所不同,因此所

提模型对用户和项目评论文本分别建模,采用双向门控循环单元分别生成用户和项目特征表示,引入注意力机制有选择的对重要特征表示部分进行提取,最后将两个部分生成的评论文本特征表示再合并到同一个向量空间完成深度部分预测评分. ADLUI 推荐模型结构如图 1 所示,接下来对推荐模型进行详细介绍.

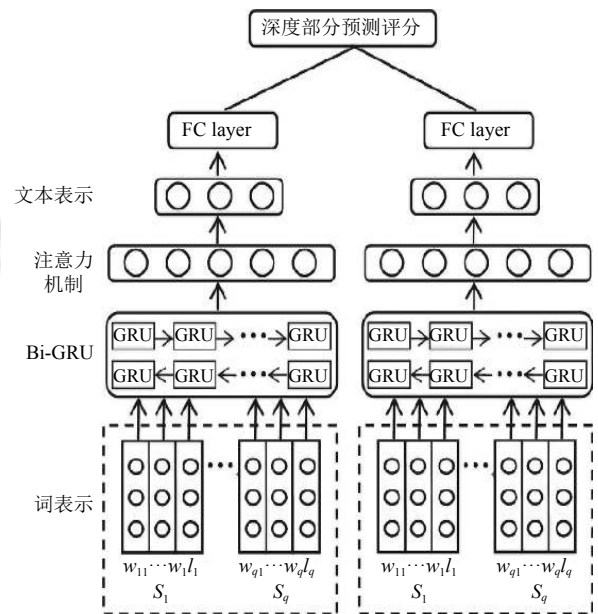


图 1 ADLUI 推荐模型结构

### 3.1 门控循环单元和注意力机制

#### 3.1.1 门控循环单元

门控循环单元(Gated Recurrent Unit, GRU)能够捕捉时间序列中时间步距较大的依赖关系,同时有效解决梯度消失问题. GRU 模型有两个门:更新门(\$Z\_t\$)和重置门(\$r\_t\$).当前新状态\$\tilde{h}\_t\$之间的线性内插值利用上一状态\$h\_{t-1}\$与新序列信息计算,权重矩阵相乘,重置门结果\$r\_t\$与\$h\_{t-1}\$进行哈达玛积\$\odot\$,然后把两部分信息相加通过\$\tanh\$激活函数将结果缩小至\$(-1, 1)\$.计算\$\tilde{h}\_t\$如式(1)所示.

$$\tilde{h}_t = \tanh(W_h w_t + U_h(r_t \odot h_{t-1}) + b_h) \quad (1)$$

网络最后计算\$h\_t\$向量,\$h\_t\$将保留当前单元信息并传递到下一个单元中.计算如式(2)所示.

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2)$$

基于GRU对评论文本建模,将文本中句子的每个单词\$w\_k^j\$都映射为\$w\_k^j \in R^d\$,GRU把输入序列\$w\_k^j \in R^d\$映射到一组隐藏状态\$h\_q^j\$,按GRU公式计算.

### 3.1.2 注意力机制

注意力机制可以有效突出评论文本重点, 提取对句子重要的用户或项目特定单词. 通过注意力机制根据隐藏状态  $s_k, \dots, s_p$  提取上下文向量  $c_k$ , 并结合当前隐藏状态  $h_k$ . 先前状态  $s_j$  的注意力分数  $a_k$  可由式 (3) 计算得出.

$$a_k = \text{softmax}(f_{\text{att}}(h_k, s_j)) \quad (3)$$

上下文向量  $c_k$  利用  $a_k$  加权计算如式 (4) 所示.

$$c_k = \sum_{j=1}^{l_k} a_{kj} s_j \quad (4)$$

## 3.2 模型实现过程

### 3.2.1 词嵌入矩阵

利用 Word2Vec 中的 Skip-gram 算法训练词嵌入矩阵, 假定评论文本由  $q$  个句子  $\{s_1, s_2, \dots, s_q\}$  组成, 其中第  $k$  个句子  $s_k$  包含  $l_k$  个单词,  $\{w_{k_1}, w_{k_2}, \dots, w_{k_{l_k}}\}$ , 则单词级嵌入矩阵  $w_{kj}$  可表示为  $w_{kj} = \{w_{k_1}, w_{k_2}, \dots, w_{k_{l_k}}\}$ .

### 3.2.2 序列数据编码方法

双向门控循环单元 (Bidirectional Gated Recurrent Unit, Bi-GRU) 前向和后向处理序列数据全面考虑了评论文本的上下文信息来捕获文本语义特征. 前向隐藏状态如式 (5) 所示, 后向隐藏状态如式 (6) 所示.

$$\vec{h}_t = \overrightarrow{\text{Bi-GRU}}(w_t, \vec{h}_{t-1}) \quad (5)$$

$$\overleftarrow{h}_t = \overleftarrow{\text{Bi-GRU}}(w_t, \overleftarrow{h}_{t+1}) \quad (6)$$

假定序列向量  $[w_1; w_2; \dots; w_j]$ , 则前向 GRU 读取序列从  $w_1$  到  $w_j$ , 后向 GRU 读取序列是从  $w_j$  到  $w_1$ .

### 3.2.3 引入注意力机制

评论文本中每个单词在其所在的句子中具有不同的重要性, 引入注意力机制的目的就是有效提取句子中的重要单词. 假定某用户 (user) 所发表的评论文本中第  $k$  个句子  $H_k^u = (h_{k_1}^u, h_{k_2}^u, \dots, h_{k_q}^u)$ , 其中  $h_{k_j}^u$  表示句子  $H_k^u$  中第  $j$  个单词的隐藏状态, 假定每个单向 GRU 的隐藏单元为  $p$ ,  $H_k^u \in R^{2p \times q}$ , 分别将每个用户 (user) 和每个项目 (item) 映射为连续的实数值向量  $u \in R^{d_u}$  和  $I \in R^{d_I}$ , 其中  $d_u$  表示用户嵌入的维度,  $d_I$  表示项目嵌入的维度. 则句子的注意力可由式 (7) 表示.

$$a_{kj}^u = \text{softmax}(w_w^u \tanh(W_{wh}^u H_k^u + W_{wu}^u u + b_w^u)) \quad (7)$$

式中,  $w_w^u \in R^{1 \times d_a}$ ,  $d_a$  为可任意设置的一个超参数,  $W_{wu}^u \in$

$R^{d_a \times 2p}$  和  $W_{wh}^u \in R^{d_a \times 2p}$  为权重矩阵,  $a_{kj}^u$  为  $1 \times q$  向量,  $a_{kj}^u$  各个维度均可表示相对应位置词的注意力多少. 因此, 通过用户评论文本中词级隐藏状态的加权和组成获得基于用户评论文本的表示如式 (8) 所示.

$$s_k^u = \sum_{j=1}^{l_k} a_{kj}^u h_{kj}^u \quad (8)$$

类似基于用户评论文本的过程同理可得基于项目评论文本的表示.

### 3.2.4 全连接层处理

将用户和项目的评论文本的特征表示分别输入全连接层 (Fully Connected layer, FC) 处理, 基于注意力机制的用户和项目语义向量分别用  $r_u$  和  $r_I$  表示, 如式 (9) 和式 (10) 所示.

$$r_u = \text{ReLU}(W_{fc}^u d^u + b_{fc}^u) \quad (9)$$

$$r_I = \text{ReLU}(W_{fc}^I d^I + b_{fc}^I) \quad (10)$$

最后将潜在的表示  $(r_u, r_I)$  映射到同一个向量空间并通过内积预测评分, 如式 (11) 所示.

$$\hat{r}_{u,I} = r_u^T r_I \quad (11)$$

## 4 实验与结果分析

### 4.1 实验数据集

实验使用两个不同域的公开数据集 Yelp (<https://www.kaggle.com/c/yelp-recsys-2013/data>) 和 Amazon (<http://jmcauley.ucsd.edu/data/amazon>) 对比不同算法的性能, 原始数据集中有百万级的商品评论信息且没有进行类别筛选和数据整理, 先进行数据预处理, 将评论记录少于 5 条的用户以及被评论记录少于 5 条的商品剔除, 数据集中的评分值是 1 至 5 之间的整数, 然后对数据集类别进行筛选. 预处理后的 Yelp 和 Amazon 数据集统计信息如表 1 和表 2 所示.

表 1 Yelp 数据集统计信息

| 特征      | 餐厅评价数  |
|---------|--------|
| Users   | 1827   |
| Items   | 1932   |
| Reviews | 89 753 |

实验中将数据集 80% 作为训练集, 10% 作为验证集, 10% 作为测试集, 进行 5 次交叉实验取平均结果作为最终结果.

表2 Amazon数据集统计信息

| 特征     | Instant | Musical     | Office   | Grocery and Gourmet |
|--------|---------|-------------|----------|---------------------|
|        | Video   | Instruments | Products | Food                |
| Users  | 3876    | 1562        | 5371     | 12059               |
| Items  | 2798    | 1058        | 2592     | 8763                |
| Review | 48623   | 13286       | 62582    | 169256              |

## 4.2 实验评估指标

实验采用均方根误差 ( $RMSE$ ) 和归一化折损累计增益 (Normalized Discounted Cumulative Gain,  $NDCG$ ) 作为评估各推荐算法性能的评价指标. 均方根误差如式 (12) 所示.

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (P_j - R_j)^2}{N}} \quad (12)$$

式中,  $N$  表示实验测试项目数,  $P_j$  表示预测评分,  $R_j$  表示实际评分,  $RMSE$  值越小说明性能越好.

$NDCG$  考虑结果集中项目的位置相关性, 如式 (13).

$$NDCG@M = \frac{1}{Z} \sum_{j=1}^M \frac{2^{rel_j} - 1}{\log_2(j+1)} \quad (13)$$

式中,  $rel_j$  表示处于位置  $j$  的推荐结果的相关性,  $Z$  表示归一化因子.

## 4.3 推荐模型参数分析

ADLUI 推荐模型参数包含 Epoch、词嵌入维度  $d_w$  和注意力维度  $d_A$ , 深入分析参数设置对模型的影响. 在训练数据中 Epoch 产生的结果如图 2 所示.

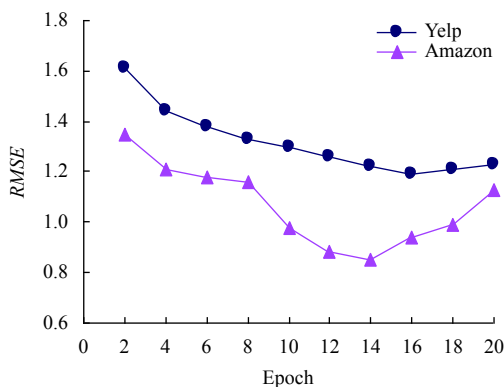


图2 训练数据中 Epoch 产生的结果

从训练中可以看出, 刚开始  $RMSE$  值相对较大, 那是因为参数更新的数量小还不能确定最优参数值, 随着训练继续, Epoch 的增大,  $RMSE$  值逐步减小. 从两个数据集的  $RMSE$  值变化情况可看出, Amazon 的  $RMSE$  值波动变化比 Yelp 大, 原因是 Amazon 数据集的数据

量大且数据多样性, 导致训练误差大, Epoch 值的变化使  $RMSE$  有明显波动. 当 Epoch 值取 13 到 17 时实验结果最优. 为了避免 Epoch 值过大而发生数据过拟合现象以导致推荐性能下降, 在验证数据中再次分析 Epoch 的取值, 验证数据中 Epoch 产生的结果如图 3 所示, Epoch 取值 15 结果最优.

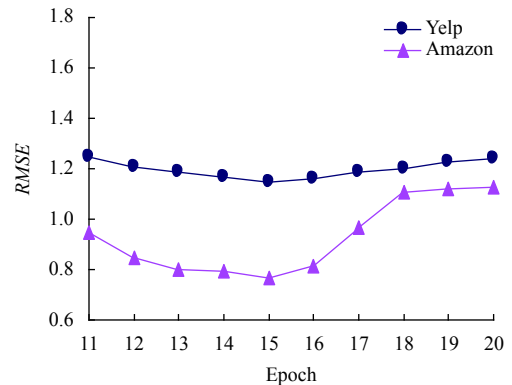
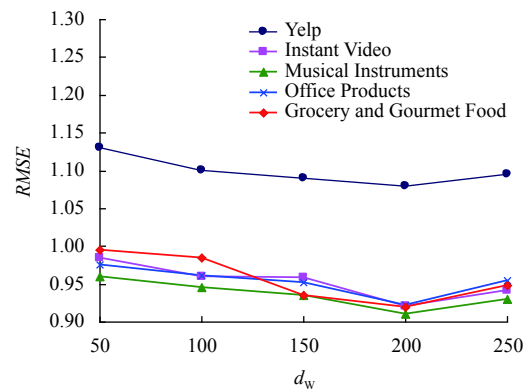


图3 验证数据中 Epoch 产生的结果

词嵌入维度  $d_w$  和注意力维度  $d_A$  通过数据集进行实验, 观察不同数据集中  $RMSE$  值的最优结果. 词嵌入维度  $d_w$  实验结果如图 4 所示,  $d_w$  取值 200 结果最优.

图4 词嵌入维度  $d_w$  实验结果

注意力维度  $d_A$  实验结果如图 5 所示,  $d_A$  取值 100 结果最优.

## 4.4 各算法比较分析

所提 ADLUI 推荐模型与现有模型 TopicMF<sup>[1]</sup>、CDL<sup>[4]</sup>、ConvMF<sup>[5]</sup>、DeepCoNN<sup>[6]</sup>、ACoNN<sup>[8]</sup> 进行比较. 在 Yelp 和 Amazon 数据集中, 各模型  $RMSE$  值的比较结果如表 3 所示, 各模型  $NDCG$  的比较结果如表 4 所示.

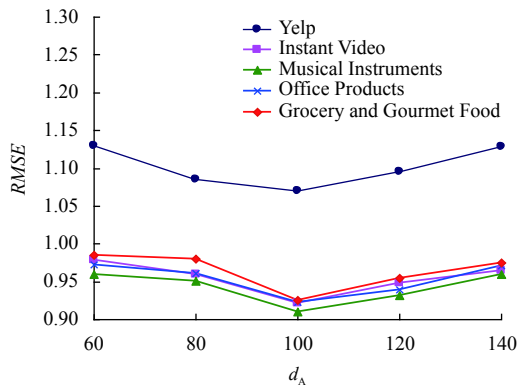
图5 注意力维度  $d_A$  实验结果

表3 各模型 RMSE 值的比较结果

| 模型       | Yelp  | Amazon |
|----------|-------|--------|
| TopicMF  | 1.108 | 0.935  |
| CDL      | 1.105 | 0.929  |
| ConvMF   | 1.096 | 0.927  |
| DeepCoNN | 1.093 | 0.924  |
| ACoNN    | 1.086 | 0.916  |
| ADLUI    | 1.082 | 0.911  |

表4 各模型 NDCG 值的比较结果

| 模型          | Yelp   |         | Amazon |         |
|-------------|--------|---------|--------|---------|
|             | NDCG@5 | NDCG@10 | NDCG@5 | NDCG@10 |
| TopicMF     | 0.344  | 0.375   | 0.474  | 0.531   |
| CDL         | 0.351  | 0.383   | 0.481  | 0.537   |
| ConvMF      | 0.356  | 0.389   | 0.484  | 0.545   |
| DeepCoNN    | 0.366  | 0.395   | 0.497  | 0.553   |
| ACoNN       | 0.374  | 0.403   | 0.504  | 0.562   |
| ADLUI       | 0.378  | 0.411   | 0.509  | 0.571   |
| Increase(%) | 1.06   | 1.95    | 0.98   | 1.58    |

通过实验结果验证了各模型的性能, TopicMF 基于主题模型考虑不到评论文本上下文及词序特征, 性能表现最差. CDL 模型是将深度学习中的堆栈自编码器与矩阵分解模型结合的一个分层贝叶斯模型. 结合评论文本建模采用自编码学习项目的特征表示, 性能优于 TopicMF. ConvMF 是将 CNN 与概率矩阵分解相结合, 利用 CNN 提取评论文本潜在特征. 相对于 CDL 更进一步, 基于 CNN 考虑词序和评论文本上下文信息进行获取潜在语义表示. DeepCoNN 是采用两个并行的 CNN 分别提取用户行为特征和评论文本中的项目属性信息, 模型的最后一层应用分解机捕获交互信息完成评分预测. 性能优于 ConvMF. ACoNN 是对 DeepCoNN 模型的改进, 在 DeepCoNN 模型中加入一层基于注意力机制的权值更新层实现对评论文本的进一步优化,

加强网络对重点内容的关注有效提高推荐的准确性. 本文所提算法 ADLUI 分别构建用户和项目的评论文本表示, 利用 Bi-GRU 捕获文本中的词序信息, 提取文本的上下文依赖关系以获得文本特征表示, 引入注意力机制, 识别文本特征表示中不同特征的重要性, 将生成的用户和项目评论数据的两组隐含特征分别输入全连接层处理获取最终特征表示, 然后再合并到同一个向量空间以进行预测评分, 有效结合用户和项目信息, 与其他的算法相比较具有更优的推荐性能. 在 Yelp 和 Amazon 两个数据集中, NDCG@10 结果较其他最好算法提高了 1.95% 和 1.58%, 实验结果表明, ADLUI 算法有效提高推荐性能.

## 5 结论

提出一种基于注意力机制的深度学习推荐算法, 充分利用评论文本信息, 挖掘用户潜在的兴趣偏好和项目属性特征, 实验结果表明, 所提算法能够有效提高推荐的精度, 为用户提供更加准确和优质的推荐项目. 下一步将考虑用户兴趣偏好会随时间而变化对推荐算法完善进行研究.

## 参考文献

- 1 Bao Y, Fang H, Zhang J. TopicMF: Simultaneously exploiting ratings and reviews for recommendation. Proceedings of the 28th AAAI Conference on Artificial Intelligence. Québec City, QC, Canada. 2014. 2–8.
- 2 Cheng ZY, Ding Y, Zhu L, et al. Aspect-aware latent factor model: Rating prediction with ratings and reviews. Proceedings of the 2018 World Wide Web Conference. Lyon, France. 2018. 639–648.
- 3 Elkahky AM, Song Y, He XD. A multi-view deep learning approach for cross domain user modeling in recommendation systems. Proceedings of the 24th International Conference on World Wide Web. Florence, Italy. 2015. 278–288.
- 4 Wang H, Wang NY, Yeung DY. Collaborative deep learning for recommender systems. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Sydney, Australia. 2015. 1235–1244.
- 5 Kim D, Park C, Oh J, et al. Convolutional matrix factorization for document context-aware recommendation. Proceedings of the 10th ACM Conference on Recommender Systems. Boston, MA, USA. 2016. 233–240.

- 6 Zheng L, Noroozi V, Yu PS. Joint deep modeling of users and items using reviews for recommendation. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Cambridge, UK. 2017. 425–434.
- 7 Costa F, Ouyang SX, Dolog P, *et al.* Automatic generation of natural language explanations. Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion. Tokyo, Japan. 2018. 57–56.
- 8 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.
- 9 Wu L, Ge Y, Liu Q, *et al.* Explainable social contextual image recommendation with hierarchical attention. arXiv preprint arXiv: 1806.00723, 2018.
- 10 颜梦香, 姬东鸿, 任亚峰. 基于层次注意力机制神经网络模型的虚假评论识别. 计算机应用, 2019, 39(7): 1925–1930. [doi: [10.11772/j.issn.1001-9081.2018112340](https://doi.org/10.11772/j.issn.1001-9081.2018112340)]
- 11 邢长征, 赵宏宝, 张全贵, 等. 融合评论文本层级注意力和外积的推荐方法. 计算机科学与探索, 2020, 14(6): 947–957. [doi: [10.3778/j.issn.1673-9418.1906067](https://doi.org/10.3778/j.issn.1673-9418.1906067)]
- 12 胡玉琦, 李婧, 常艳鹏, 等. 引入注意力机制的 BiGRU-CNN 情感分类模型. 小型微型计算机系统, 2020, 41(8): 1602–1607. [doi: [10.3969/j.issn.1000-1220.2020.08.006](https://doi.org/10.3969/j.issn.1000-1220.2020.08.006)]