

注意力与多视角融合的新闻推荐算法^①



范琳娟, 孙喆喆, 徐 飞, 周行行

(西安工业大学 计算机科学与工程学院, 西安 710021)

通信作者: 范琳娟, E-mail: 1530196267@qq.com

摘 要: 针对目前现有的新闻推荐系统未能充分考虑新闻的语义信息, 对新闻文本建模因子的单一性问题, 提出注意力与多视角融合的新闻推荐算法 (Attention-BodyTitleEvent, Attention-BTE). 利用 BERT 模型以及注意力机制分别对新闻标题、正文、事件向量化, 将三者融合即新闻向量化表示, 再对候选新闻和用户浏览新闻数据进行处理, 分别得到对应的候选新闻向量化和用户向量化, 并将其进行点乘得到用户点击候选新闻的概率, 即新闻推荐结果. 实验数据表明, 与其他的新闻推荐算法相比, 该模型在 $F1$ 指标上提高了约 6%.

关键词: 新闻推荐系统; 多视角; 注意力机制; 事件; 向量化; 个性化推荐

引用格式: 范琳娟, 孙喆喆, 徐飞, 周行行. 注意力与多视角融合的新闻推荐算法. 计算机系统应用, 2022, 31(12): 178-186. <http://www.c-s-a.org.cn/1003-3254/8890.html>

News Recommendation Algorithm Based on Fusion of Attention and Multi-perspective

FAN Lin-Juan, SUN Yong-Yong, XU Fei, ZHOU Hang-Hang

(School of Computer Science and Engineering, Xi'an Technological University, Xi'an 710021, China)

Abstract: The existing news recommendation system fails to sufficiently consider the semantic information of news, and modeling factors for news body suffers from unity problems. Attention-BodyTitleEvent (Attention-BTE), a news recommendation algorithm based on fusion of attention and multi-perspectives, is proposed in this study. The BERT model and attention mechanism are applied to vectorize the body, title, and event in the news respectively. The three parts are combined to represent news vectorization, and then the candidate news and user browsing news data are processed respectively to obtain the corresponding candidate news vectorization and user vectorization. Finally, dot multiplication is conducted to obtain the probability of users clicking on the candidate news, namely the news recommendation result. Experimental data demonstrate that Attention-BTE improves the index by about 6% compared with the other news recommendation algorithm.

Key words: news recommendation system; multi-perspective; attention mechanism; event; vectorization; personalized recommendation

随着网络环境的飞速发展, 信息技术得到了广泛应用, 人们的新闻阅读习惯逐渐从传统的报纸、电视等媒介转移到互联网上. 然而, 人们对新闻信息的阅览能力往往不及新闻的产生速度, 新闻信息环境日趋稠密, 最终导致“信息过载”^[1]. 所以, 如何高效、快捷、有效地为用户推荐感兴趣的新闻文本资源变得越来越重要^[2].

近年来个性化新闻推荐逐渐成为 Web 技术的热点, 推荐系统在实时资讯、新闻、微博、电影、音乐、博客、电商等都有大量的应用. 通过推荐系统能够快速高效准确地处理信息过载等问题, 分析出用户对新闻内容的兴趣评分结果和用户行为特征及利用历史数据自动构建的用户兴趣模型, 无须对用户刻意地

① 基金项目: 新型网络与检测控制国家地方联合工程实验室项目 (GSYSJ2016013)

收稿时间: 2022-04-13; 修改时间: 2022-05-22; 采用时间: 2022-06-29; csa 在线出版时间: 2022-08-26

进行兴趣调查,大幅度有效地降低了用户的负担,让推荐系统的知名度得到大大提高.个性化推荐功能在移动互联网行业中的被运用也十分的普遍,今日头条、美团大众点评网等均有个性化推荐.虽然人们对推荐算法已多次进行了深入的技术研发,但现有的推荐算法仍然具有局限:(1)新闻有强时效性.新闻平台每天会自动产生大量新的新闻,而旧的新闻也会随之迅速消失.这就造成了一个很大程度的冷启动问题,导致许多如协同过滤等依赖用户行为的推荐方法无法使用;(2)新闻文章具有丰富的正文,这些正文包含重要信息未被利用,无法很简单直观地直接通过ID等特征来表示新闻;(3)精准建模用户的兴趣度.用户的兴趣度多样化并随时间的流动而改变,需要大量基于用户反馈行为来挖掘和建模.但是新闻平台上往往没有显式的用户反馈,甚至隐式反馈也非常稀疏.因此,个性化新闻推荐已成为国内外众多科研机构研究的热点,获得了信息检索、数据挖掘和人工智能等领域国内外学术会议的高度关注^[3,4].

综上所述,本文提出注意力与多视角融合的新闻推荐算法(Attention-BTE).从新闻的“最小单位”即词语考虑,利用BERT模型以及注意力机制分别对新闻中的标题、正文、事件向量化,并进一步将三者融合得到新闻向量化.通过对用户历史浏览数据分析,从多视角建模得到用户向量化.即用户对新闻的点击率是新闻向量化和用户向量化做点乘的结果.本文将新闻事件作为对新闻内容建模的考虑因素,实验表明,Attention-BTE算法能有效提高新闻推荐结果的准确性.

1 相关工作

人们对新闻推荐研究起源于上个世纪,近年来,随着神经网络的兴起,对基于深度学习的新闻推荐算法^[5]也有大量的研究.

1.1 传统的新闻推荐方法

早期的新闻推荐通常依赖于新闻之间的关联性^[6]和语义相关性^[7].这两种方法都是对新闻进行建模,计算新闻之间的相似性,但是这两种方法都无法准确描述用户的兴趣度,它们只表示了新闻之间的相似性,缺少对用户的兴趣度表示.所以,在个性化的新闻推荐任务中如何挖掘和表示用户的新闻阅读兴趣是非常重要的问题.

协同过滤(CF)计算目标用户和邻居之间的行为相似度,再通过最近邻居预测出目标用户对某个项目

的可能评分,最后将评分的Top-N推送给目标用户.但是,由于新闻推荐平台每天都会产生大量的、全新的新闻,因此,新闻推荐系统存在严重的冷启动问题.但是,协同过滤方法难以有效地解决系统的冷启动问题,所以,并不适用于新闻推荐.基于此,很多新闻推荐都采用了基于新闻内容的推荐,如Goossen等^[8]结合TF-IDF与领域本体的语义进行推荐;Samarinas等^[9]引入一种使用单词嵌入来构建用户兴趣模型的方法,实现新闻个性化推荐,这两种方法都只是从单一视角对新闻编码,不能完整、精准的表达新闻的语义信息.

1.2 基于深度学习的新闻推荐算法

2017年雅虎新闻使用深度学习网络作为特征提取器对新闻内容进行研究,以此来发现用户的喜好,相比于传统的方法有效地提升了推荐系统的准确度^[10],但该方法仅从新闻词语角度对新闻向量化.Huang等^[11]提出新的深度混合推荐模型DMFL,从新闻的多个角度对新闻向量化,最后将各个角度的向量进行求和并求平均值,得到用户的特征向量,这种方法不能准确地对用户建模.

Google公司为了实现新闻推荐提出了Wide&Deep模型^[12],该模型能同时训练线性模型和深度网络,但是在现实场景中存在学习效率低等问题.Chen等^[13]提出了结合用户的位置和兴趣度对用户进行个性化新闻推荐,此方法主要对用户向量化,不能准确考虑新闻语义内容.Wang等^[14]提出通过知识感知神经网络(CNN)从新闻文章标题中学习新闻表征,然后根据候选新闻和每个浏览新闻之间的相似性从新闻表征中学习用户表征,此方法只从新闻标题角度对新闻向量化,还不能准确的对新闻内容建模.

综上所述,本文提出了注意力与多视角融合的新闻推荐算法(Attention-BTE),从多视角对新闻和用户建模,充分考虑新闻语义信息,提高新闻推荐准确率.

2 注意力与多视角融合的新闻推荐模型构建

注意力与多视角融合的新闻推荐模型包括新闻编码器、用户编码器、点击预测器3个模块.该模型框架图如图1所示,其中 D^c 表示候选新闻, r 表示候选新闻向量化, $D_1, \dots, D_i, \dots, D_N$ 表示用户浏览新闻, $r_1, \dots, r_i, \dots, r_N$ 表示用户浏览新闻向量化, q_n 表示query向量, $\alpha_1^n, \alpha_i^n, \alpha_N^n$ 表示用户对浏览新闻的喜爱程度, u 表示用户向量化, \hat{y} 表示用户对候选新闻的点击率.

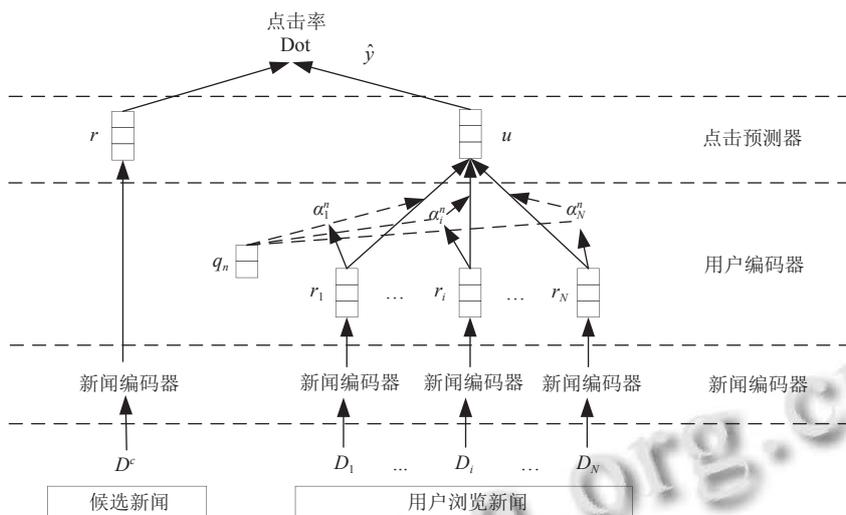


图1 注意力与多视角融合的新闻推荐模型框架

2.1 新闻编码器

新闻编码器从新闻标题、正文以及事件 3 个视角

学习新闻内容,对新闻向量化.新闻编码器框架如图2所示.

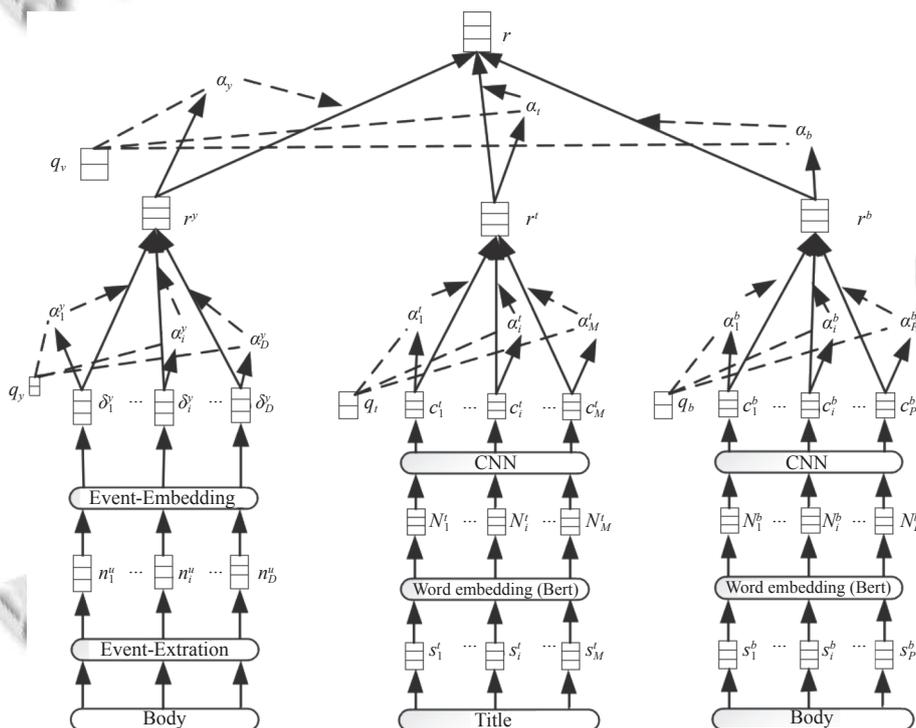


图2 新闻编码器框架

新闻编码器从两个方面考虑新闻信息。

(1) 新闻词语: 新闻标题、正文的每个词语对新闻内容准确表达的重要性不同, 本文通过注意力机制给标题、正文的每个词语赋予权重, 即词语的重要度, 再加权求和得到新闻标题、正文的向量。

(2) 新闻视角: 新闻不同视角包含的信息不同, 为了精准描述用户对新闻标题、正文、事件的关注度, 将标题、正文以及事件通过注意力机制赋予不同的权重, 再对各个视角加权求和, 得到新闻向量化。

新闻编码器由标题编码器、正文编码器、事件编

码器、注意力模块4部分构成。

2.1.1 标题编码器

标题编码器对新闻标题向量化。标题编码器包括预处理层、词向量训练层、CNN层、注意力层。

预处理层将新闻标题进行分词,利用Jieba分词工具对标题进行分词、去停用词等预处理操作。新闻标题分词后的词语表示为 $[s_1^t, s_2^t, \dots, s_M^t]$, M 是标题词语的数量。

词向量训练层是将预处理层的词语通过BERT模型^[15]得到标题词语的词向量,标题词语向量表示为 $[c_1^t, c_2^t, \dots, c_M^t]$, M 是标题词语的数量。

CNN层将词向量训练层的词向量通过CNN网络,提取词语之间的特征。卷积核的宽度与词嵌入维度一致,每次卷积操作的窗口取词数记为 h ,即卷积核 $\omega \in R^{h \times d}$ 。新闻标题第 i 个词语窗口滑动的卷积结果为:

$$N_i^t = \text{ReLU}(\omega w_{i:i+h-1}) + b \quad (1)$$

其中, ReLU 为非线性激活函数; $w_{i:i+h-1}$ 为每次卷积操作的取词数; $b \in R$ 为偏置。

即标题词语表示为: $[N_1^t, \dots, N_i^t, \dots, N_P^t]$ 。

注意力层,利用注意力机制给每个词语赋予权重。表示新闻标题的第 i 个词语的注意力权重为 α_i^t ,其计算公式为:

$$\begin{cases} a_i^t = \frac{q_t^T \tanh(V_t \times c_i^t + v_t)}{\|q_t^T\|_{L2} \|\tanh(V_t \times c_i^t + v_t)\|_{L2}} \\ \alpha_i^t = \frac{\exp(a_i^t)}{\sum_{j=1}^M \exp(a_j^t)} \end{cases} \quad (2)$$

其中, V_t 表示权重, v_t 表示偏置项, c_i^t 表示新闻标题第 i 个词语的向量, q_t 表示query向量。新闻标题向量表示为:

$$r^t = \sum_{j=1}^M \alpha_j^t c_j^t \quad (3)$$

2.1.2 正文编码器

正文编码器对新闻正文向量化。正文编码器包括预处理层、词向量训练层、CNN层、注意力层。

预处理层将新闻正文进行分词,利用Jieba分词工具将正文切分为词语,新闻正文的词语表示为 $[s_1^b, s_2^b, \dots, s_P^b]$, P 是正文中词语的数量。

词向量训练层是将预处理层的分词通过BERT模型得到新闻正文的词向量,正文词语向量表示为 $[c_1^b, c_2^b, \dots, c_P^b]$, P 是正文中词语的数量。

CNN层将词训练层的词向量通过CNN网络,提取词语之间的特征。卷积核的宽度与词嵌入维度一致,每次卷积操作的窗口取词数记为 h ,即卷积核 $\omega \in R^{h \times d}$ 。新闻正文第 i 个词语窗口滑动的卷积结果为:

$$N_i^b = \text{ReLU}(\omega w_{i:i+h-1}) + b \quad (4)$$

其中, ReLU 为非线性激活函数; $w_{i:i+h-1}$ 为每次卷积操作的取词数; $b \in R$ 为偏置项。即正文词语表示为: $[N_1^b, \dots, N_i^b, \dots, N_P^b]$ 。

注意力层是利用注意力机制给新闻正文的词语赋予权重,来选择每篇新闻正文中重要的词语。表示新闻正文的第 i 个词语的注意力权重为 α_i^b ,其计算公式为:

$$\begin{cases} a_i^b = \frac{q_b^T \tanh(V_b \times c_i^b + v_b)}{\|q_b^T\|_{L2} \|\tanh(V_b \times c_i^b + v_b)\|_{L2}} \\ \alpha_i^b = \frac{\exp(a_i^b)}{\sum_{j=1}^P \exp(a_j^b)} \end{cases} \quad (5)$$

其中, V_b 表示权重, v_b 表示偏置项, c_i^b 表示新闻正文第 i 个词语的向量, q_b 表示query向量。新闻正文向量表示为:

$$r^b = \sum_{j=1}^P \alpha_j^b c_j^b \quad (6)$$

2.1.3 事件编码器

事件编码器对新闻事件向量化。事件编码器包括事件抽取层、事件向量化层、注意力层。

事件抽取是从非结构化的自然语言文本中自动抽取用户感兴趣的事件信息并以结构化的形式表示^[6]。本文通过百度中文依存句法分析工具(DDParser)抽取新闻事件,新闻事件表示为 $[n_1^u, n_2^u, \dots, n_D^u]$, D 表示事件的个数,每个事件表示为: $n_i^u = (o_{i1}, p_i, o_{i2})$, o_{i1} 表示主语, p_i 表示谓语, o_{i2} 表示宾语。

事件向量化层是将事件向量化。Li等^[16]提出了将事件元素的词向量进行拼接作为事件的向量表示。本文首先将事件抽取层的事件主语、谓语、宾语分别通过Jieba分词切分为对应的词语,再利用正文编码器词向量训练层得到主语、谓语、宾语对应的每个词语向量化并将其分别表示为对应词语向量的平均值,最后事件向量化表示为事件元素的向量拼接,即事件表示为 δ_i^v 。

注意力层给每个事件赋予权重。同一新闻中的不同事件对于新闻的向量化标准具有不同的重要性。本文在此使用注意力机制对新闻事件赋予权重,来衡量新闻中事件的重要程度,表示新闻的第 i 个事件的注意

力权重为 α_i^y , 其计算公式为:

$$\begin{cases} \alpha_i^y = \frac{q_y^T \tanh(V_y \times \delta_i^y + v_y)}{\|q_y^T\|_{L2} \|\tanh(V_y \times \delta_i^y + v_y)\|_{L2}} \\ \alpha_i^y = \frac{\exp(a_i^y)}{\sum_{j=1}^P \exp(a_j^y)} \end{cases} \quad (7)$$

其中, V_y 表示权重, v_y 表示偏置项, δ_i^y 表示第 i 个事件向量化, q_y 表示 query 向量. 新闻事件向量表示为:

$$r^y = \sum_{j=1}^y \alpha_j^y \delta_j^y \quad (8)$$

2.1.4 注意力模块

利用注意力机制给标题、正文、事件赋予权重, 加权求和得到新闻向量化, 设 α_t 表示新闻标题的权重, 其计算公式如式(9):

$$\begin{cases} \alpha_t = \frac{q_v^T \tanh(U_{vt} \times r^t + u_{vt})}{\|q_v^T\|_{L2} \|\tanh(U_{vt} \times r^t + u_{vt})\|_{L2}} \\ \alpha_t = \frac{\exp(a_t)}{\exp(a_t) + \exp(a_b) + \exp(a_y)} \end{cases} \quad (9)$$

其中, U_{vt} 表示权重, u_{vt} 表示偏置项, r^t 表示新闻标题向量化, q_v 表示 query 向量. 同理可得新闻正文、事件的权重为 α_b 、 α_y . 最后, 新闻向量化为:

$$r = \alpha_t r^t + \alpha_b r^b + \alpha_y r^y \quad (10)$$

2.2 新闻编码器算法

新闻编码器算法的输是新闻标题、正文, 输出是新闻的向量化. 该算法通过对新闻标题、正文、事件向量化, 从而对新闻向量化. (1) 标题向量化: 输入新闻标题, 输出标题向量化. (2) 正文向量化: 输入新闻正文, 输出正文向量化. (3) 事件向量化: 输入新闻正文, 输出新闻事件向量化. 将其通过注意力机制进行融合, 得到新闻向量化. 该算法具体描述如算法 1.

算法 1. 新闻编码器算法

输入: N -Title, N -Body

输出: N -Embedding

Step 1. 初始化所有参数

Step 2. N -Title 的分词记为: s_i^t , N -Body 的分词记为: s_i^b

Step 3. N -Event 记为 $n_i^e=(o_{i1}, p_i, o_{i2})$

Step 4. s_i^t 词向量化记为: c_i^t ; s_i^b 词向量化记为: c_i^b

//标题向量化

Step 5. for t steps do

Step 6. 从 N -Title 中随机抽取一小批正样本和负样本

Step 7. 利用式(1)、式(2)更新参数

Step 8. 利用式(3)计算 N -Title 向量记为: r^t

END

//正文向量化

Step 9. for t steps do

Step 10. 从 N -Body 中随机抽取一小批正样本和负样本

Step 11. 利用式(4)、式(5)更新参数

Step 12. 利用式(6)计算 N -Body 向量记为: r^b

END

//事件向量化

Step 13. for t steps do

Step 14. 从 N -Body 中随机抽取一小批正样本和负样本

Step 15. 利用式(7)更新参数

Step 16. 利用式(8)计算 N -Event 向量记为: r^y

END

//新闻向量化

Step 17. 利用式(9)更新参数

Step 18. 利用式(10)计算 N -Embedding, 记为: r

END

算法 1 中, N 表示新闻; N -Title 表示新闻 N 的标题; N -Body 表示新闻 N 的正文; N -Event 表示新闻 N 的事件; N -Embedding 表示新闻 N 的向量化; o_{i1} 表示 N -Event 主语; p_i 表示 N -Event 谓语; o_{i2} 表示 N -Event 宾语.

2.3 用户编码器

用户编码器对用户兴趣度向量化. 将用户阅读的新闻通过新闻编码器得到用户阅读的每篇新闻向量化, 记为: r_i ($i = 1, 2, \dots, n$), 通过注意力机制给用户阅读的每篇新闻赋予权重, 用来衡量用户对每篇新闻的喜爱程度, 表示第 i 篇新闻的注意力权重为 α_i^n , 其计算公式为:

$$\begin{cases} \alpha_i^n = \frac{q_n^T \tanh(W_n \times r_i + b_n)}{\|q_n^T\|_{L2} \|\tanh(W_n \times r_i + b_n)\|_{L2}} \\ \alpha_i^n = \frac{\exp(a_i^n)}{\sum_{j=1}^P \exp(a_j^n)} \end{cases} \quad (11)$$

其中, W_n 表示权重, b_n 表示偏置项, r_i 表示新闻标题向量化, q_n 表示 query 向量. 用户向量化为:

$$u = \sum_{j=1}^n r_j \alpha_j^n \quad (12)$$

2.4 点击预测器

点击预测器是用户对候选新闻点击概率的预测, 将通过新闻编码器得到的候选新闻向量和通过用户编码器得到的用户向量做点乘运算得到用户对候选新闻的点击率, 其计算公式为:

$$p = r \cdot u \quad (13)$$

按点击率高低将 Top-N 篇新闻推荐给用户.

2.5 系统推荐算法

系统推荐算法输入是候选新闻和用户浏览新闻,

输出是用户对候选新闻的点击率。利用新闻编码器和用户编码器分别得到候选新闻和用户兴趣度向量化,将其点乘得到用户对候选新闻的点击率。该算法具体描述如算法2。

算法2. 系统推荐算法

输入: N, U 输出: p

Step 1. 初始化所有参数

Step 2. N 通过新闻编码器计算新闻向量化 r Step 3. U 通过用户编码器得到用户向量 u Step 4. 利用式(13)计算 p 。

END

算法2中, N 表示候选新闻, U 表示用户浏览新闻, p 表示用户对候选新闻的点击率, r 表示 N -Embedding, u 表示用户向量。

2.6 模型训练

本模型采用负样本技术进行训练。对于用户浏览过的新闻,视为正样本新闻,将在同一界面出现但用户并没有点击的新闻视为负样本新闻,随机抽取 K 个负样本新闻,预测正样本新闻 \hat{y}_i^+ 和 K 个负样本新闻 $[\hat{y}_1^-, \hat{y}_2^-, \dots, \hat{y}_K^-]$ 的点击率,使用 Softmax 激活函数对点击率进行标准化来计算正样本的点击概率:

$$p_i = \frac{\exp(\hat{y}_i^+)}{\exp(\hat{y}_i^+) + \sum_{j=1}^K \exp(\hat{y}_{i,j}^-)} \quad (14)$$

其中, \hat{y}_i^+ 表示第 i 个正样本新闻的点击率, $\hat{y}_{i,j}^-$ 表示第 i 个正样本新闻和第 j 个负样本新闻在同一界面中出现的情况下第 j 个负样本新闻的点击率。假设 S 表示正样本训练集,计算模型训练的损失函数 L 为:

$$L = - \sum \log(p_i) \quad (15)$$

3 实验结果和分析

3.1 数据集

本论文在中文数据集搜狐新闻和英文数据集 MIND (Microsoft news dataset) 上进行实验。

搜狐新闻是在搜狐新闻网站上爬取 2017 年 6 月部分用户的历史浏览记录,其包括用户、新闻标题、新闻正文、浏览时间戳。我们使用 6 月 25 日至 6 月 30 日的数据构建测试集,使用 6 月 1 日至 6 月 24 日的数据构建训练集,训练集中 6 月 24 日的数据构建验证集。

MIND^[17] 是一个大规模新闻推荐英文数据集,由

微软新闻从其 2019 年 10 月 12 日至 11 月 22 日匿名用户日志中收集而得,其包含用户、新闻标题、新闻正文、新闻摘要、新闻类别、新闻实体。本文使用 11 月 15 日至 11 月 22 日的数据构建测试集,使用 10 月 12 日至 11 月 14 日的数据构建训练集,训练集中 11 月 14 日的数据构建验证集。详细统计信息如表 1 所示。

表1 数据集情况统计

类别	中文数据集	英文数据集
用户数	1 000	1 000 000
新闻数	100 000	161 013
平均新闻标题长度	11.00	14.28
平均新闻正文长度	550.00	564.00

3.2 实验模型参数

本实验基于 TensorFlow 实现模型,具体的模型参数如表 2 所示。

表2 超参数取值

名称	取值
vec_dim	128
num_filters	128
Kernal_size	3、4、5
lr	0.0001
batch_size	64
epochs	20
query_dim	128

3.3 评价指标

本论文设 T 为测试集,为用户 u 进行新闻推荐, F_u 为推荐结果,假设用户 u 在测试数据集里感兴趣的新闻集合记为 T_u 。本文的性能指标以准确率 (Precision, P)、召回率 (Recall, R) 以及 $F1$ 指数作为参考。

准确率指新闻推荐系统推荐的新闻列表中符合用户偏好的内容占比,准确率越高,说明新闻推荐越准确,其计算公式如下:

$$P = \frac{\sum_u |F_u \cap T_u|}{\sum_u |F_u|}$$

召回率指用户在与诸多新闻实行交互行为的过程中,由推荐系统推荐而来的新闻占有比率,即用户浏览的新闻中,其为系统推荐新闻的占有率,其计算公式如下:

$$R = \frac{\sum_u |F_u \cap T_u|}{\sum_u |T_u|}$$

推荐系统的衡量标准是准确率、召回率越高越优,但是由于这两个指标具有某种程度的矛盾关联,而 $F1$ 则是将两者相结合的评价指标.因此,本论文使用 $F1$ 值来衡量推荐系统的性能优劣, $F1$ 值越高意味着综合性能越优,计算公式如下:

$$F1 = \frac{1}{\frac{1}{P} + \frac{1}{R}}$$

3.4 实验结果

3.4.1 新闻推荐列表 N 对 P 、 R 、 $F1$ 值的影响

中文数据集的测试集新闻数量是 853 名用户的 3 654 篇点击新闻,英文数据集的测试集是 920 056 名用户的 6 325 篇点击新闻.

由于不同的推荐列表长度会直接影响结果的 P 、 R 、 $F1$ 值.为验证在不同推荐列表长度下 Attention-BTE 的有效性,在测试集中,本次实验取推荐列表 N 的范围为 {10, 20, 30, 40, 50}, 然而, N 的取值不能无限大,因为其不能脱离测试集的限制.图 3、图 4 分别是中英文数据集在不同推荐列表长度 N 的情况下的 P 、 R 、 $F1$ 值的对比.

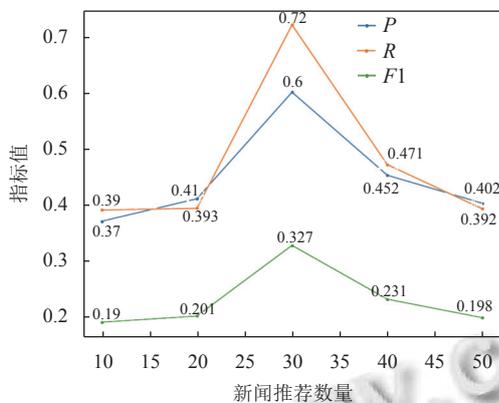


图 3 中文数据集下不同新闻推荐数量的比较结果

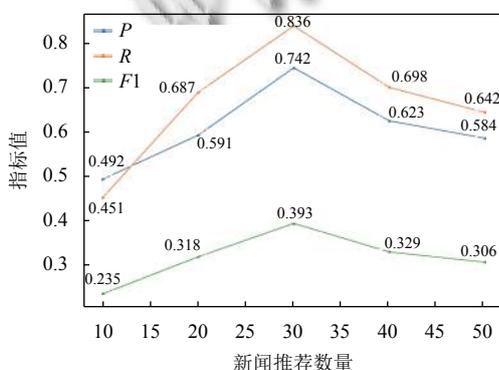


图 4 英文数据集下不同新闻推荐数量的比较结果

由图 3、图 4 可以得到结论,给用户推荐 30 篇新闻时, P 、 R 、 $F1$ 可以取得最大值,新闻推荐效果最优.随着推荐列表的增大, P 、 R 、 $F1$ 在逐渐增大,这是因为推荐列表长度的增加使推荐列表中包含更多与用户产生交互的项目,但是在推荐 40 篇、50 篇, P 、 R 、 $F1$ 值呈现减小趋势,这是因为推荐数量的增多而出现拟合现象.即过小的推荐数量不能增加更多与用户产生交互的项目,而在测试集数量少的情况下,过大的推荐数量则容易受到数量过度拟合的影响.

3.4.2 模型对比

在推荐列表数量相同的情况下,本方法与新闻推荐算法进行对比.如 CR、CF、DeepFM、Wide&Deep、CNN、NAML 模型,其中,中英文数据集下推荐的准确率、召回率和 $F1$ 值分别如表 3、表 4 所示.本文有如下观察.

表 3 不同模型在中文数据集上的比较结果

算法	P	R	$F1$
CR	0.500	0.607	0.274
CF	0.480	0.675	0.281
DeepFM	0.513	0.662	0.289
Wide&Deep	0.540	0.684	0.302
CNN	0.562	0.678	0.307
NAML	0.585	0.695	0.318
Attention-BTE	0.600	0.720	0.327

表 4 不同模型在英文数据集上的比较结果

算法	P	R	$F1$
CR	0.620	0.752	0.291
CF	0.622	0.756	0.291
DeepFM	0.685	0.790	0.317
Wide & Deep	0.693	0.802	0.321
CNN	0.698	0.793	0.321
NAML	0.725	0.823	0.338
Attention-BTE	0.742	0.836	0.352

第一,由于英文数据集的用户数量、新闻数量是大于中文数据集的,所以中文数据集中 CF 方法准确率低于 CR 方法的准确率,而在英文数据集中则与之相反.这可能是由于中文数据集的用户、新闻数量较少,CF 方法遇到了严重的用户-项目评分矩阵稀疏问题,因此数据集的大小也是影响新闻推荐准确性的关键.

第二,融合了标题、正文和类别(例如: NAML)的方法优于仅利用了新闻标题的方法(例如: DeepFM、Wide&Deep、CNN),这可能是丰富的内容信息有助于学习更准确的新闻表示,证明了融合多种新闻信息的

有效性。

第三, Attention-BTE 模型优于 CR、CF、DeepFM、Wide&Deep、CNN、NAML 模型, 因为本模型充分利用了新闻的主题、正文、事件, 并且将事件作为新闻信息的一个视角, 形成了信息量丰富的新闻表示, $F1$ 值提高 6% 左右。

3.4.3 不同视角对 P 、 R 、 $F1$ 的影响

在推荐列表相同的情况下, 比较不同视角的新闻推荐的准确率、召回率和 $F1$ 值。将 Attention-BTE 分别与标题、正文、事件、标题+正文、标题+事件、正文+事件比较 P 、 R 、 $F1$ 值, 比较结果如表 5 所示, 其图表表示如图 5。

表 5 中文数据集下不同视角的比较结果

视角	P	R	$F1$
标题	0.190	0.240	0.106
正文	0.280	0.290	0.142
事件	0.420	0.390	0.202
标题+正文	0.435	0.510	0.226
标题+事件	0.472	0.580	0.260
正文+事件	0.520	0.630	0.285
Attention-BTE	0.600	0.720	0.327

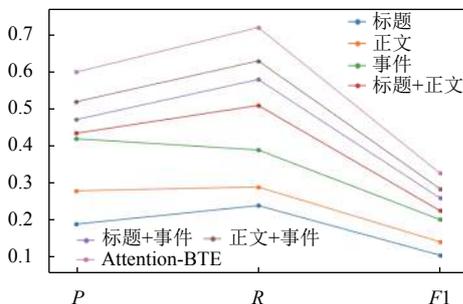


图 5 中文数据集下不同视角比较结果的图形表示

本文首先比较了不同视角的新闻信息, 如标题、正文、事件、标题+正文、标题+事件、正文+事件对推荐结果的影响。从图 5 可以看出, 事件相比于标题、正文取得了更好的表现, 这是因为事件是从新闻正文中提取而得, 事件是新闻正文信息的提炼, 所以, 事件对推荐指标值的影响是大于标题、正文。其次, 正文的实验表现明显优于标题, 因为正文可以提供更丰富、详细的内容信息, 证明丰富的文本内容可以带来更好的推荐效果。最后, 与仅使用标题相比, Attention-BTE 模型在中文数据集上的 $F1$ 提高约 22%, 证明了融合不同类型视角的有效性, 同时证明了不同新闻视角在学习新闻表示的过程中起着相辅相成的互补作用。

3.4.4 注意力与多视角融合的有效性

本文通过注意力机制计算新闻标题、正文、事件三者的权重。在实验中, 对新闻标题、正文、事件 3 个视角进行分析, 三者的注意力权重标准差如图 6 所示。

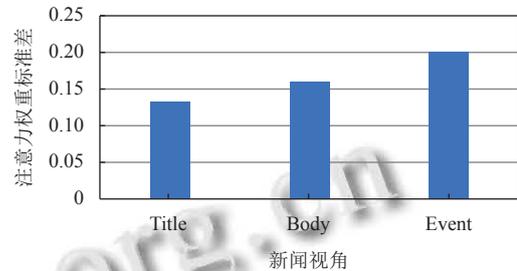


图 6 本文模型在中文数据集上新闻标题、正文、事件的标准差

实验结果表明, 事件的注意力权重最大, 其次是正文, 最后是标题。原因: (1) 事件是对新闻内容的精准概括。(2) 正文表示新闻的原本意思, 包含比标题更丰富的信息。(3) 标题对于学习新闻表示来说是模糊的和无信息的。在这些情况下, 事件可以为推荐提供丰富的补充信息, 验证了注意力多视角学习的有效性。

4 结论

新闻推荐研究具有重要的现实意义和应用需求, 新闻推荐的准确性在于是否充分考虑了新闻的内容信息, 本文提出注意力与多视角融合的新闻推荐算法 (Attention-BTE), 将新闻标题、正文、事件通过注意力机制融合, 得到新闻向量化。首先, 因为词语是组成新闻标题、正文、事件最基本的单位, 同一词语在不同的新闻中具有不同的含义, 所以, 从最基本的词语入手, 给每个词语都赋予不同的权重, 能充分考虑新闻语义信息。其次, 能准确对用户向量化。由于每个用户对新闻感兴趣的角度不同, 使用注意力机制可以精准描述用户对新闻感兴趣的视角, 从而精准对用户向量化。最后, 通过注意力机制给新闻的标题、正文、事件赋予权重, 充分考虑各个角度对新闻向量化准确性的比率。利用真实的实验数据, 通过实验验证了该方法的有效性。在后续的研究中, 可以使用神经网络的方法计算事件的向量化, 以进一步提高新闻推荐的准确性。

参考文献

- 孟祥武, 胡勋, 王立才, 等. 移动推荐系统及其应用. 软件学

- 报, 2013, 24(1): 91–108.
- 2 祝婷, 秦春秀, 马晓悦, 等. 基于本体与 LDA 主题模型的文
本资源推荐方法研究. 情报杂志, 2015, 34(11): 150–156.
[doi: [10.3969/j.issn.1002-1965.2015.11.028](https://doi.org/10.3969/j.issn.1002-1965.2015.11.028)]
 - 3 Kazai G, Yusof I, Clarke D. Personalised news and blog
recommendations based on user location, Facebook and
Twitter user profiling. Proceedings of the 39th International
ACM SIGIR Conference on Research and Development in
Information Retrieval. Pisa: ACM, 2016. 1129–1132.
 - 4 Karkali M, Pontikis D, Vazirgiannis M. Match the news: A
firefox extension for real-time news recommendation.
Proceedings of the 36th International ACM SIGIR
Conference on Research and Development in Information
Retrieval. Dublin: ACM, 2013. 1117–1118.
 - 5 de Souza PMG, Jannach D, da Cunha AM. Contextual hybrid
session-based news recommendation with recurrent neural
networks. IEEE Access, 2019, 7: 169185–169203.
 - 6 Ahn D. The stages of event extraction. Proceedings of the
Workshop on Annotating and Reasoning about Time and
Events. Sydney: ACL, 2006. 1–8.
 - 7 Capelle M, Hogenboom F, Hogenboom A, *et al.* Semantic
news recommendation using WordNet and Bing similarities.
Proceedings of the 28th Annual ACM Symposium on
Applied Computing. Coimbra: ACM, 2013. 296–302.
 - 8 Goossen F, Ijntema W, Frasincar F, *et al.* News
personalization using the CF-IDF semantic recommender.
International Conference on Web Intelligence. Sogndal:
ACM, 2011. 10.
 - 9 Samarinas C, Zafeiriou S. Personalized high quality news
recommendations using word embeddings and text
classification models. EasyChair Preprints, 2019.
 - 10 Okura S, Tagami Y, Ono S, *et al.* Embedding-based news
recommendation for millions of users. Proceedings of the
23rd ACM SIGKDD International Conference on
Knowledge Discovery and Data Mining. Halifax: ACM,
2017. 1933–1942.
 - 11 Huang ZH, Yu C, Ni J, *et al.* An efficient hybrid
recommendation model with deep neural networks. IEEE
Access, 2019, 7: 13790–137912.
 - 12 Cheng HT, Koc L, Harmsen J, *et al.* Wide & deep learning
for recommender systems. Proceedings of the 1st Workshop
on Deep Learning for Recommender Systems. Boston: ACM,
2016. 7–10.
 - 13 Chen C, Meng XW, Xu ZH, *et al.* Location-aware
personalized news recommendation with deep semantic
analysis. IEEE Access, 2017, 5: 1624–1638. [doi: [10.1109/
ACCESS.2017.2655150](https://doi.org/10.1109/ACCESS.2017.2655150)]
 - 14 Wang HW, Zhang FZ, Xie X, *et al.* DKN: Deep knowledge-
aware network for news recommendation. Proceedings of the
2018 World Wide Web Conference. Lyon: International
World Wide Web Conferences Steering Committee, 2018.
1835–1844.
 - 15 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of
deep bidirectional Transformers for language understanding.
Proceedings of the 2019 Conference of the North American
Chapter of the Association for Computational Linguistics:
Human Language Technologies. Minneapolis: ACL, 2019.
4171–4186.
 - 16 Li ZY, Ding X, Liu T. Constructing narrative event
evolutionary graph for script event prediction. Proceedings of
the 27th International Joint Conference on Artificial
Intelligence. Stockholm: AAAI Press, 2018. 4201–4207.
 - 17 Wu FZ, Qiao Y, Chen JH, *et al.* MIND: A large-scale dataset
for news recommendation. Proceedings of the 58th Annual
Meeting of the Association for Computational Linguistics.
Online: ACL, 2020. 3597–3606.

(校对责编: 牛欣悦)