

引用格式:王凯丽,李海明. 基于多通道 CNN-BiGRU 与多特征融合方法[J]. 微电子学与计算机, 2022, 39(3): 41-49. [WANG K L, LI H M. A method with multi-channel CNN-BiGRU and Multi-feature fusion[J]. Microelectronics & Computer, 2022, 39(3): 41-49.]DOI:10. 19304/J. ISSN1000-7180. 2021. 0987

基于多通道 CNN-BiGRU 与多特征融合方法

王凯丽,李海明

(上海电力大学 计算机科学与技术学院,上海 200000)

摘 要: 新闻推荐是根据用户的阅读习惯,为其推送更符合需求的内容,然而现有的方法仍存在特征学习不足的问题. 针对此问题,提出了一种基于多通道 CNN-BiGRU 与多特征融合方法,主要由以下四部分组成:(1)词嵌入层. 在词向量中融入实体嵌入向量,弥补单独仅使用词向量的不足,完成多通道词向量的构建;(2)多通道 CNN-BiGRU 模型. 此部分模型使用卷积神经网络(CNN)提取语句的局部特征,使用双向门控循环单元(BiGRU)提取文本的长序列依赖关系,同时,借助最大池化操作以减少参数数量,避免冗余,借助注意力网络以关注重要词汇的特征,获取精确的特征表示;(3)多特征融合. 借助注意力机制融合多个新闻特征,并关注重要特征的内容,完成新闻表征的构建;(4)用户表征提取. 通过多头注意力机制提取用户历史浏览记录中新闻间的交互关系,以构建准确的用户表征,完成推荐. 实验结果表明,所提出的 CBMR 模型相对现有的 CNN、DKN、TANR 和 NRMS 模型,在 AUC、MRR、nDCG@5 和 nDCG@10 指标上表现更优异.

关键词: 卷积神经网络;双向门控循环单元;注意力机制;用户表征

中图分类号: TP391. 3

文献标识码: A

文章编号: 1000-7180(2022)03-0041-09

A method with multi-channel CNN-BiGRU and Multi-feature fusion

WANG Kaili, LI Haiming

(School of Computer Science and Technology, Shanghai University of Electric Power, Shanghai 200000, China)

Abstract: According to the reading habits of users, news recommendations can push more suitable content for them, but the existing methods still have the problem of insufficient feature learning. To solve this problem, a method based on multi-channel CNN-BiGRU and multi-feature fusion is proposed, which is mainly composed of the following four parts: (1) Word Embedding Layer. Entity embedding vectors are integrated into word vectors to make up for the deficiency of using word embedding vectors and construct multi-channel word vectors; (2) Multi-channel CNN-BiGRU Model. In the model, a convolutional neural network (CNN) is used to extract the local features of sentences, and a bidirectional gated recurrent unit (BiGRU) is used to extract the long sequence dependency of text. At the same time, the maximum pool algorithm is used to reduce the number of parameters and avoid redundancy, and the attention network is used to pay attention to the features of key words to extract accurate features; (3) Multi-feature Fusion. With the help of attention mechanism, integrate multiple news features and pay attention to important features to complete the construction of news representation; (4) User Encoder. The interaction of news in the browsing histories of a user is extracted by multi-headed self-attention mechanism to construct user representation and complete recommendation. The experimental results show that compared with CNN, DKN, TANR, and NRMS models, CBMR performs better in AUC, MRR, nDCG@5, and nDCG@10

收稿日期: 2021-07-23; 修回日期: 2021-08-17

基金项目: 国家自然科学基金面上项目(61772327);上海自然科学基金面上项目(20ZR1455900)

indicators.

Key words: convolutional neural network; bidirectional gated recurrent unit; attention mechanism; user representation

1 引言

由于网络上每天都有海量的新闻报道产生,一个用户不可能通过阅读所有的新闻文章来获得感兴趣的信息,因此,个性化的新闻推荐是解决新闻信息过载、实现用户个性化信息获取的一个重要方法^[1].传统推荐算法存在着人工提取特征不足、数据稀疏等问题,而基于深度学习的推荐算法能自动提取特征,具有强大的学习力,在新闻、图像、音乐等领域被广泛应用,并取得了良好效果.

基于深度学习的新闻推荐模型优先考虑如何学习新闻表征和用户表征.例如,文献[2]使用个性化注意力机制来建模用户对不同上下文和新闻的兴趣;文献[3]使用长短记忆网络(LSTM)捕获用户对新闻的兴趣;文献[4]使用门控制循环单元(GRU)捕捉新闻间的交互关系以对用户短期兴趣进行建模.然而,这三种模型存在一些不足:(1)在学习新闻表征中仅利用了标题这一个特征;(2)循环神经网络虽能提取文本序列信息,但局部特征提取不足.

根据对用户历史行为日志分析,可以有以下几个发现:首先,一篇新闻文章通常包含了多个特征,如标题、摘要和类别等,这些信息都可以进行新闻表示.其次,不同单词和特征间包含的信息量也是有差异的.标题中每个单词的重要性是不一样的,例如在标题“Xbox One On Sale This Week”中,单词“Xbox”和“Sale”要比单词“On”和“This”重要.不同特征中,标题所含信息是简明准确的,而摘要比标题信息量多且更具体化,分类则十分简短却是必不可少的.因此,需要关注重要单词和特征.最后,用户浏览记录中的新闻存在着交互关系,为了构建更准确的用户表征,需要挖掘新闻间的交互关系.

针对上述问题,提出了一种基于多通道 CNN-BiGRU 与多特征融合模型(Multi-channel CNN-BiGRU and Multi-feature Fusion for Recommendation, CBMR).本文主要工作有以下四个方面:

(1)将基于知识图谱的实体嵌入向量和词向量进行融合,产生新的词向量,使模型易于理解上下文的语义信息;

(2)提出一种基于多通道 CNN-BiGRU 模型.该模型既借助卷积神经网络 CNN 抽取句子的部分

特征,又借助双向门控循环单元 BiGRU 提取长序列信息,同时,运用池化层实现特征降维,并借助注意力机制,避免特征冗余,改善了特征学习过程,有助于提高模型的性能;

(3)采用基于注意力机制的多特征融合方法.使用注意力机制融合新闻的标题、摘要、主分类和子分类四个特征,关注重要特征的内容,避免干扰;

(4)提出一种多头自注意力机制构建用户表征的方法.该方法借助多头自注意力机制挖掘新闻间的交互关系,并引入注意力机制使关键新闻获取更多的关注.

通过在真实数据集上的实验,结果表明,本文所提出的 CBMR 模型能有效提升新闻推荐的性能.

2 相关工作

传统的推荐算法通常依托人工提取新闻和用户的表征,既需要储备相当多的专业知识,又消耗大量的时间和人力.与传统方法相比,深度学习具有自动提取特征、可移植性好、学习能力强等优点.现如今,大多数的新闻网站(如今日头条和微软新闻)都使用了深度学习技术,并取得了很好的成绩.

传统的新闻推荐的方法:(1)LR 模型^[5].该模型使用机器学习的逻辑回归进行点击率预测,但依赖于人工提取的特征组合,无法自动学习交叉特征.(2)FM 模型^[6].该模型通过隐向量做内积来表示组合特征,但是由于计算的复杂度,只能考虑到二阶的交叉特征.(3)Wide&Deep 模型^[7].该模型可以同时捕获低阶和高阶组合特征,并将线性模型和 Deep 模型进行结合,但是线性模型的输入依赖于人工提取的特征.(4)DeepFM 模型^[8].该模型是 Wide&Deep 模型的改进,不需要人工特征工程,能同时学习低阶和高阶的交叉特征,但是该交叉特征是未知的、不可控的.

基于深度学习的新闻推荐方法:(1)CNN 模型^[9].Kim 等人将卷积神经网络(CNN)应用于新闻推荐中,替代传统推荐中人工提取特征,自动捕获新闻的低阶和高阶特征,但会遇到梯度消失,新闻表征学习不足等问题.(2)DKN 模型^[10].Wang 等人将卷积神经网络(CNN)与知识图谱进行融合,利用知识图谱计算新闻标题中每个实体的嵌入值及实体上下

文的嵌入值,最后将新闻标题的词嵌入、实体嵌入和上下文嵌入一同作为 CNN 的输入来捕获新闻表征.该模型提升了新闻推荐的准确率,但未考虑新闻间的交互关系,难以捕获用户真实喜好.(3)TANR 模型^[11].该方法是一种基于分类意识的新闻推荐算法,通过分类意识的编码模型训练预测新闻类别,再将标题与类别进行结合训练.该模型融合了新闻的多个特征进行学习,但类别信息不是真实的数据,而是经过编码模型提取得到的,存在误差,无法保证推荐的准确性.(4)NRMS 模型^[12].Wu 等人借助多头自注意力机制捕获标题的上下文信息.与注意力机制不同,自注意力机制侧重于学习单词间的依赖关系,能捕获句子的内部结构,是一种自身向自身学习的过程.

不同于上述模型,本文提出的基于多通道

CNN-BiGRU 和多特征融合模型(CBMR),首先考虑了新闻的多个特征信息,其次充分发挥了注意力机制、CNN 和 BiGRU 的优势,通过多通道 CNN-BiGRU 模型提取文本的局部特征和长序列依赖关系,并借助注意力机制进行多特征融合,然后利用多头自注意力机制构建用户表征,最后完成新闻推荐的预测.

3 CBMR 模型

本文提出了一种基于多通道 CNN-BiGRU 和多特征融合模型 CBMR,主要用于新闻推荐中.模型主要由三个部分组成:(1)多通道 CNN-BiGRU 模型;(2)多特征融合;(3)用户表征提取.如图 1 所示.

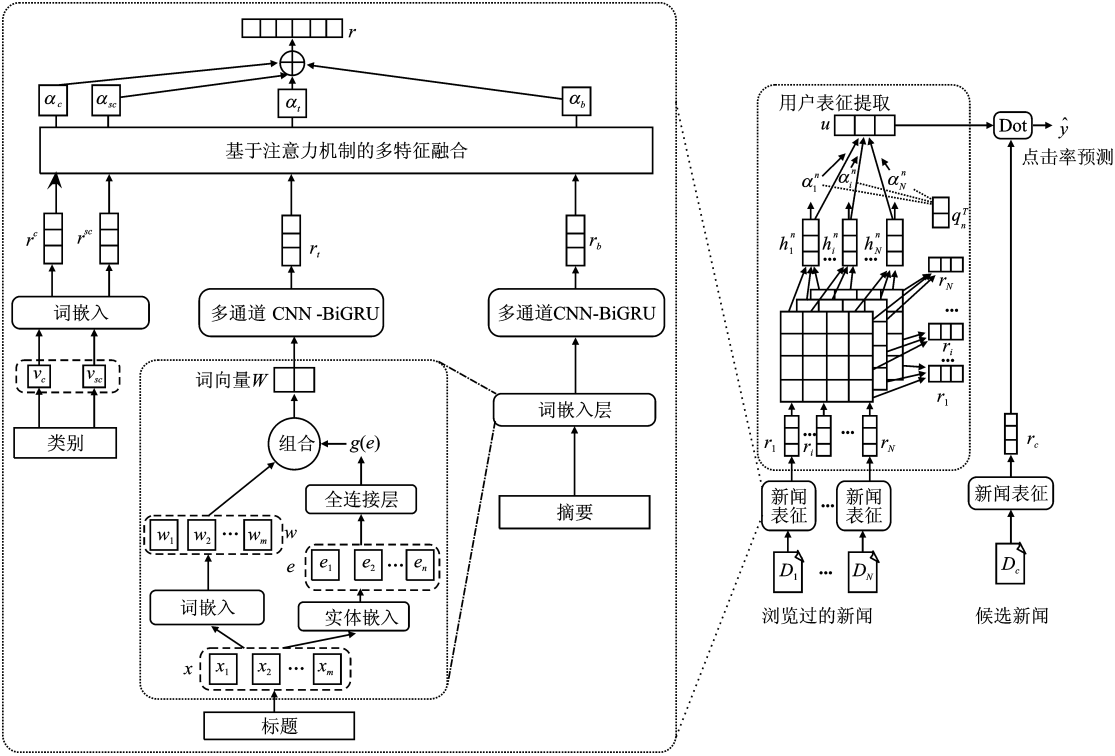


图 1 基于多通道 CNN-BiGRU 与多特征融合模型 CBMR

Fig. 1 Multi-channel CNN-BiGRU and Multi-feature Fusion for Recommendation, CBMR

3.1 词嵌入层

词嵌入层借助大型语料库上训练得到的预训练模型加载词向量,再与实体嵌入向量进行融合,生成新的词向量.

假设一个语句中有 m 个单词 $x = [x_1, x_2, \dots, x_m]$,经词嵌入后,将词序列转换成词向量,得到词向量为 $w = [w_1, w_2, \dots, w_m]$.在实体嵌入中,通过知识图谱推导出每个单词 $w_i \in R^{D \times 1}$ 关联的实体嵌

入信息 $e_i \in R^{K \times 1}$,其中 D 为词嵌入向量维度, K 为实体嵌入向量维度,记实体嵌入向量为 $e = [e_1, e_2, \dots, e_n]$,其中 n 是实体的个数.实体嵌入信息就是将新闻内容中被识别出的知识实体与来自维基百科的知识图进行关联,再经嵌入模型学习得到的实体向量表示.

将实体嵌入向量 e 作为全连接神经网络的输入,得到其隐藏表示 $g(e)$,使实体嵌入向量和词嵌

入向量处于同一维度中,具体计算公式如式(1):

$$g(e) = \tanh(Me + b) \quad (1)$$

式中, $M \in R^{D \times K}$ 是可训练参数矩阵, $b \in R^{D \times 1}$ 是偏置值, $\tanh$ 是非线性的激活函数。

然后,将得到的实体嵌入向量 $g(e)$ 和词嵌入向量 w 进行合并,产生新的词向量 W ,即 $W = [[w_1, g(e_1)], \dots, [w_m, g(e_m)]]$. 若存在第 i 个单词 w_i 无关联的实体信息,则将 e_i 设为 0.

3.2 多通道 CNN-BiGRU 模型

此部分模型有三部分组成:(1)卷积神经网络层 CNN;(2)双向门控循环单元层 BiGRU;(3)组合. 该模型主要用于提取新闻标题和摘要的特征. 框架如图 2 所示.

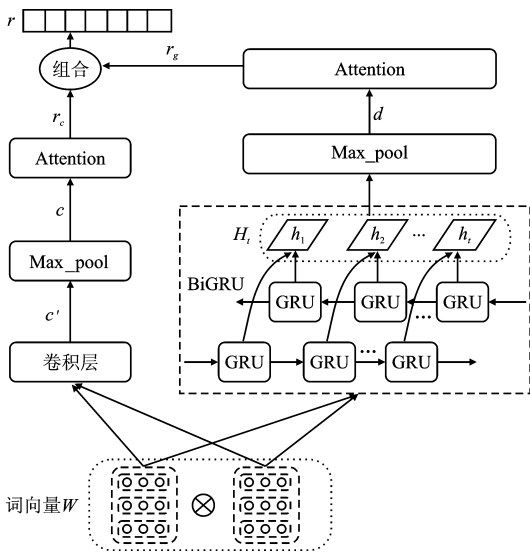


图 2 多通道 CNN-BiGRU 模型

Fig. 2 Multi-channel CNN-BiGRU model

(1) 卷积神经网络层 CNN

卷积核可以捕捉到多个连续词之间的特征,所以应用卷积神经网络来挖掘句子的局部特征. 本文的卷积神经网络主要分为卷积层,池化层和注意力层.

卷积层中,将词向量 W 作为输入,设第 i 个单词经过卷积层后得到的向量记为 c'_i ,具体计算公式如式(2):

$$c'_i = \text{ReLU}(f_i \times W_{i:i+k-1} + b_i) \quad (2)$$

式中, $f_i \in R^{N_f \times k \times D \times 2}$ 是卷积核, N_f 是卷积核的数量, D 是词向量的维度, $k \times D$ 是卷积核大小, k 是窗口大小,可取多个不同的值, $W_{i:i+k-1}^t$ 是第 i 个单词和距离 k 的单词的词向量, $b_i \in R^{N_f}$ 是偏置值. ReLU 是非线性的激活函数. 卷积之后,输出向量为 $c' = [c'_1, c'_2, \dots, c'_{m-k+1}]$

池化层中,选择最大池化方法,以减少模型参

数,降低过拟合风险,具体计算公式如式(3):

$$c = \text{Max}(c') \quad (3)$$

池化之后,输出向量为 $c = [c_1, c_2, \dots, c_{(N_f-1)/\text{stride}+1}]$,其中 N_f 是卷积核的数量, stride 是池化层步长.

在标题“Xbox One On Sale This Week”中,由单词“Xbox”和“Sale”可知该条新闻与游戏销售有着密切关系,进一步说明单词“Xbox”和“Sale”要比“One”、“On”等单词重要,更值得被关注. 因此,在池化层之后,采用 Yang 等人^[13]提出的单词层面的注意力机制,确保将注意力放至关键单词上,而忽略无关紧要的单词信息,获取准确特征表示.

注意力层中,设标题中第 i 个单词的注意力权重的计算公式如式(4):

$$\begin{aligned} a_i &= q_i^T \tanh(V_t \times c_i + v_t) \\ \alpha_i &= \frac{\exp(a_i)}{\sum_{j=1}^M \exp(a_j)} \end{aligned} \quad (4)$$

式中, V_t 和 v_t 是参数, q_i^T 表示注意力查询向量. 将所有单词的注意力进行加权求和得到新的向量表示,

$$\text{即 } r_c = \sum_{j=1}^{(N_f-1)/\text{stride}+1} \alpha_j c_j$$

(2) 双向门控循环单元层 BiGRU

卷积神经网络存在一定的缺陷,不能捕获长距离特征,而长短期记忆网络(LSTM)可从词向量中提取上下文信息,能有效捕捉长序列依赖关系. 门控循环单元网络(GRU)是 LSTM 的改进,既能有效缓解循环神经网络(RNN)梯度消失的问题,又具备参数量少、降低过拟合的特点.

本文所采用的 BiGRU 网络由前向 GRU 和后向 GRU 网络组成. GRU 主要由更新门和重置门组成,更新门主要控制将过去信息传入给当前状态,重置门主要控制前一状态信息写入当前候选集中. 设当前时刻为 t ,GRU 计算公式如式(5)所示:

$$\begin{aligned} r_t &= \sigma(\omega_r \cdot [h_{t-1}, W_t] + b_r) \\ z_t &= \sigma(\omega_z \cdot [h_{t-1}, W_t] + b_z) \quad t \in [1, m] \\ \tilde{h}_t &= \tanh(\omega_h \cdot [r_t * h_{t-1}, W_t] + b_h) \\ h_t &= z_t * h_{t-1} + (1 - z_t) * \tilde{h}_t \end{aligned} \quad (5)$$

式中, σ 是 sigmoid 函数, $\cdot$ 是点积, $\omega_r, \omega_z, \omega_h$ 是权重矩阵, b_r, b_z, b_h 是偏置值. W_t 是 t 时刻输入的词向量, r_t 是重置门, z_t 是更新门, $\tilde{h}_t$ 是当前状态的候选集, h_t 是隐藏状态,也是最后的输出.

双向门控循环单元网络(BiGRU)由两个不同

方向的 GRU 构成. t 时刻, BiGRU 具体计算公式如式(6)所示:

$$\begin{aligned}\vec{h}_t &= \text{GRU}(W_t, \vec{h}_{t-1}) \\ \overleftarrow{h}_t &= \text{GRU}(W_t, \overleftarrow{h}_{t-1}) \quad t \in [1, m] \\ h_t &= [\vec{h}_t, \overleftarrow{h}_t]\end{aligned}\quad (6)$$

最终双向门控循环单元网络 (BiGRU) 输出记为 $H_t = [h_1, h_2, \dots, h_t]$

为了压缩数据,减少参数的数量,在 BiGRU 网络之后,加入池化层,选择最大池化方法,具体计算公式如式(7)所示:

$$d = \max(H_t) \quad (7)$$

池化之后,输出向量 $d = [d_1, \dots, d_{(T \times 2 - 1)/\text{stride} + 1}]$,其中 T 是 BiGRU 的隐藏层的特征维度, stride 是最大池化的步长.

同理,为了避免冗余,引入注意力网络捕捉语句中的关键词,形成新的句子向量,计算公式如式(8):

$$\begin{aligned}a_i &= q^T \tanh(V \times d_i + v) \\ \alpha_i &= \frac{\exp(a_i)}{\sum_{j=1}^M \exp(a_j)} \\ r_g &= \sum_{j=1}^{(T \times 2 - 1)/\text{stride} + 1} \alpha_j d_j\end{aligned}\quad (8)$$

式中, V 和 v 是参数, q^T 表示注意力查询向量, α_i 是第 i 个单词的注意力权重值,得到句子向量表示为 r_g .

(3) 组合

最后,将词向量经过卷积神经网络 CNN 和双向门控循环单元 BiGRU 作用之后的向量,进行拼接,得到该特征的表示向量 r ,具体计算公式如式(9):

$$r = \text{concat}(r_c, r_g) \quad (9)$$

因此,标题和摘要经多通道 CNN-BiGRU 模型后,得到的特征表示向量分别为 r^t 和 r^b .

3.3 多特征融合

多特征融合层主要利用注意力网络将新闻的标题、主类别、子类别和摘要四个特征进行融合. 由于各个特征所包含的信息量也不同,比如标题往往是简短且点明主旨的,而摘要内容多且详细,显然标题的重要性要比摘要重要,因此,基于注意力机制的多特征融合促使模型重视关键特征信息,忽略无关紧要的内容,进一步提高模型的特征提取能力.

假设标题、摘要、主类别和子类别的注意力权重

表示为 $\alpha_t, \alpha_b, \alpha_c$ and α_s , 以标题的注意力权重 α_t 为例,计算公式如式(10):

$$\begin{aligned}a_t &= q_v^T \tanh(U_v \times r^t + u_v) \\ \alpha_t &= \frac{\exp(a_t)}{\exp(a_t) + \exp(a_b) + \exp(a_c) + \exp(a_s)}\end{aligned}\quad (10)$$

式中, U_v 和 u_v 是参数, q_v^T 是注意力查询向量.

新闻主分类、子分类和摘要计算注意力权重方法与标题的注意力权重计算相似. 最后,新闻表征由多特征加权求和得到,计算公式如式(11):

$$r = \alpha_c r^c + \alpha_s r^s + \alpha_t r^t + \alpha_b r^b \quad (11)$$

式中, r^c 与 r^s 是主分类和子分类经词嵌入后得到的特征表示向量.

3.4 用户表征提取

用户表征提取层主要从用户历史会话中学习用户表征,主要分成以下两部分:

第一部分,多头自注意力网络. 由于用户浏览过的新闻集为 $[D_1, D_2, \dots, D_N]$, 它们在时间上具有顺序性,且单条新闻往往与多条新闻间存在交互关系,因此,按用户浏览新闻的先后顺序,利用多头自注意力机制挖掘新闻间的交互关系进行用户表征的构建. 在第 k 个自注意力网络上第 i 条新闻的特征表示 $h_{i,k}^n$ 的计算公式如式(12):

$$\begin{aligned}\beta_{i,j}^k &= \frac{\exp(r_i^T Q_k^n r_j)}{\sum_{m=1}^N \exp(r_i^T Q_k^n r_m)} \\ h_{i,k}^n &= V_k^n \left(\sum_{j=1}^N \beta_{i,j}^k r_j \right)\end{aligned}\quad (12)$$

式中, n 是自注意力网络的个数, N 是新闻的个数, Q_k^n 和 V_k^n 是第 k 个自注意力网络的参数, $\beta_{i,j}^k$ 代表着第 i 条与第 j 条新闻之间的相关性. 经过多头自注意力网络后,第 i 条新闻的表示为 $h_i^n = [h_{i,1}^n, h_{i,2}^n, \dots, h_{i,h}^n]$.

第二部分,新闻层面的注意力网络. 因为用户经常浏览的新闻比用户偶尔浏览的新闻更能准确地对用户进行表示,显然,用户经常浏览的新闻在用户表征中的重要性要高于用户偶尔浏览的新闻. 因此,引入新闻层面的注意力机制,以选取重要的新闻进行用户表征的学习. 第 i 条新闻的注意力权重 α_i^n 的计算公式如式(13):

$$\begin{aligned}a_i^n &= q_n^T \tanh(V_n \times h_i^n + v_n) \\ \alpha_i^n &= \frac{\exp(a_i^n)}{\sum_{j=1}^N \exp(a_j^n)}\end{aligned}\quad (13)$$

式中, V_n 和 v_n 是参数, q_n^T 是注意力的查询向量. 最后, 输出用户表征向量为 $u = \sum_{i=1}^N \alpha_i^n h_i^n$.

3.5 推荐

设候选新闻 D_c 的新闻表征为 r_c , 用户的向量表示为 u . 根据 Okura 等人^[14] 提出的方法, 将候选新闻 r_c 和用户特征向量 u 进行点积, 获得点击率预测值 $\hat{y}$, 即 $\hat{y} = u^T r_c$.

4 实验

4.1 数据集

为验证本文模型 CBMR 的有效性, 选取了微软公开的新闻数据集 MIND^[15] 进行实验. 数据集收集了匿名用户在微软新闻网站上最近一周的行为数据, 即从 2019 年 11 月 9 日到 2019 年 11 月 15 日的用户行为. 每条用户历史行为包含序列号、用户 ID、时间、点击的新闻和留有印象的新闻. 其中, 留有印象的新闻是指在某个时间点展现给用户看的新闻界面里的新闻, 包含新闻 ID 以及标签, 表示用户是否有点击过该新闻, 若点击过标签为 1, 未点击过标签为 0. 选取最后一天的用户行为作为测试集, 其余时间的作为训练集. 实验数据集的统计信息如表 1 所示.

表 1 MIND 数据集统计信息
Table. 1 MIND Dataset Statistics

信息	数量
用户	94 057
新闻	65 238
用例	675 603
实体	12 956
每个标题的平均单词数	15. 48
每条摘要的平均单词数	48. 55
主分类	18
子分类	260

4.2 参数设置

使用 Glove 预训练模型^[16] 获取词嵌入向量, 词嵌入维度设为 300. 实体嵌入向量是通过 TransE 方法从维基百科的知识图谱中获取的, 具体矢量值在数据集中已提供, 维度设为 100. CNN 过滤器的数量设为 60, 窗口大小设为 [2, 3, 4]. 注意力机制维度设置 200, 多头自注意力网络中设置 10 个头, 根据 transformer 模型^[17], 在多头自注意力网络之中, 增加残差连接和层标准化 (Layer Normalization), 以防止梯度消失, 同时控制当前层的输出结果对下一层的影响范围. 损失函数选择交叉熵 (Cross En-

tropy Loss), 优化算法采用 Adam^[18], 批大小设为 64, 独立地重复每个实验 10 次.

4.3 基准方法和评价指标

通过与四种基线方法的比较, 对基于多通道 CNN-BiGRU 与多特征融合模型 CBMR 的性能进行评估. 比较方法包括: (1) CNN^[9], 将卷积神经网络 (CNN) 应用于新闻标题中, 学习新闻表征; (2) DKN^[10], 一种联合知识图谱和卷积神经网络 (CNN) 的新闻推荐方法; (3) TANR^[11], 将新闻标题和辅助新闻分类进行结合的推荐方法; (4) NRMS^[12], 一种基于自注意力机制的新闻推荐方法;

为评估 CBMR 模型的有效性, 选取了 AUC、MRR、nDCG@K 三种作为评价指标:

(1) AUC. 指 ROC 曲线下与横轴围成的面积, 表示预测得到正样本的概率大于负样本概率的概率.

$$AUC = \sum_{i \in P} \frac{\text{rank}_i - \frac{M \times (M + 1)}{2}}{M \times N}$$

式中, P 是样本数, M 代表正样本数, N 是负样本数, rank 是样本预测值的排名.

(2) MRR. 平均倒数排名, 把正确检索结果值在检索结果中的排序取倒数作为其准确率值, 再对所有的问题取平均, 以评估推荐系统的性能.

$$MRR = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{\text{rank}_i}$$

式中, Q 是查询个数, rank_i 是第 i 个查询, 与它第一个相匹配的结果所在的排序位置.

(3) nDCG@K. 归一化折损累积增益, 定义为

$$\text{nDCG@K} = \frac{\text{DCG}_K}{\text{IDCG}}$$

$$\text{DCG}_K = \sum_{i=1}^K \frac{2^{\text{rel}_i} - 1}{\log_2(i + 1)}$$

式中, K 是参数表示列表长度. rel_i 表示第 i 个推荐的相关程度, 值为 1 时表明项目实际被选中, 0 表示未被选中, IDCG 是归一化的 DCG, 即根据理想排序, 获得的最大 DCG 值^[19].

4.4 实验结果与分析

(1) 模型结果对比

通过与四种基线方法的比较, 对本文所提模型 CBMR 的性能进行评估, 结果如表 2 所示. 其中, TANR-和 TANR 模型构造一致, 但 TANR-模型仅考虑了新闻标题未考虑类别, 而 TANR 模型同时考虑了新闻标题和类别, 是一种基于主题分类意识的新闻推荐方法.

表 2 模型对比				
Table. 2 Model comparison				
模型	AUC	MRR	nDCG@5	nDCG@10
CNN	0.550 9	0.238 6	0.249 2	0.315 3
DKN	0.585 0	0.249 2	0.266 0	0.335 0
TANR-	0.624 1	0.278 8	0.306 7	0.372 5
TANR	0.622 8	0.283 4	0.311 5	0.375 3
NRMS	0.624 9	0.274 9	0.303 3	0.368 3
CBMR	0.673 6	0.326 0	0.362 5	0.421 4

根据表 2,可以观察到 CBMR 模型明显优于其他模型.原因如下:(1)本文模型是基于多特征的,不同与其他方法(例如:CNN,DKN 和 NRMS)仅对标题特征进行学习,CBMR 模型不仅利用了标题,还利用了摘要和类别多个特征;(2)CBMR 模型充分发挥卷积神经网络(CNN)和双向门控循环单元(BiGRU)的优势,分别利用 CNN 和 BiGRU 网络对实体嵌入向量和词向量两个通道内容进行学习,可以更好地捕获词汇的语义信息和上下文信息;(3)CBMR 模型利用多头自注意力机制挖掘新闻间的交互信息,有助于构建出准确的用户表征.

(2)不同因素对比

为研究多通道 CNN-BiGRU 模型与多头自注意力机制对本文模型性能的影响,对 CBMR 模型进行变体验证:(1)CBMR-O-MA 模型,单独仅考虑使用多头自注意力网络捕捉到的用户兴趣.(2)CBMR-O-CNN 模型,单独仅考虑使用卷积神经网络学习局部特征.(3)CBMR-O-BiGRU 模型,单独仅考虑使用双向门控循环单元提取词汇的长距离依赖关系.(4)CBMR 模型,则使用多通道 CNN-BiGRU 模型捕获局部特征和长序列特征,使用多头自注意力机制考虑历史记录中用户的兴趣.结果如表 3 所示.

表 3 不同因素对比分析			
Table. 3 Comparative analysis of different factors			
模型	AUC	MRR	nDCG@5
CBMR-O-MA	0.585 641	0.252 225	0.269 064
CBMR-O-CNN	0.651 182	0.310 585	0.341 553
CBMR-O-BiGRU	0.659 554	0.316 509	0.347 849
CBMR	0.673 604	0.325 952	0.362 454

从表 3 中可以看出,仅使用多头自注意力网络,且不使用卷积神经网络和双向门控循环单元时,模型性能显著下降,而同时使用和多头自注意力机制、CNN 和 BiGRU 时,模型性能得到显著提升.这是因为使用多头自注意力机制可以有效捕捉用户兴趣,而通过 CNN 网络中的多个卷积核,可以学习到多种特征,同时,使用 BiGRU 网络有助于捕捉到词

与词间的长距离依赖关系,这三者均有助于模型性能的提升.

同时,考虑不同特征的影响,将 CBMR 模型中的多个特征进行单独实验并对比,进一步探讨本文模型中多特征融合的有效性,结果如图 3 所示.

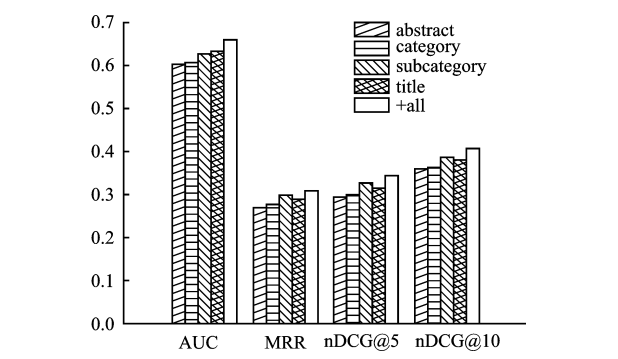


图 3 不同特征因素的对比分析

Fig. 3 Comparative analysis of different features

从图 3 中,可以发现标题和子分类的学习效果要明显高于摘要和主分类的学习.这可能是因为标题能点明主旨的,在新闻和用户表示中,往往起着决定性作用,而子分类标签是简短且具体的,在新闻主题中也扮演着必不可少的角色.因此,标题和子分类信息都可以捕捉更准确的新闻表征.本文模型将标题、摘要、主分类和子分类多种特征进行融合的效果要明显优于单个特征的学习,进一步说明了 CBMR 模型的有效性.

(3)参数分析

主要分析卷积神经网络的卷积核数量 N_f 对于 CBMR 模型性能的影响.分别设置卷积核数量为 $\{30,60,90\}$,选择窗口大小为 $\{2,3,4\}$,观察评价指标值的变化,结果如表 4 所示.

表 4 卷积核数量 N_f 的影响			
Table. 4 Influence of the number of convolution kernels N_f			
N_f	AUC	MRR	nDCG@5
30	0.663 474	0.311 944	0.345 838
60	0.671 523	0.320 318	0.356 986
90	0.657 351	0.310 036	0.343 402

从表 4 中可以看出,在同一窗口下,随着卷积核数量的增加,AUC 等评价指标的值也随之上升,说明多卷积的卷积神经网络有利于挖掘语句中的隐性信息,提高 CBMR 模型的准确性.然而,当过滤器数量达到 90 时,AUC、MRR、nDCG@5 的值随之下落,因此卷积核的数量 N_f 取值不能过大,否则会导致模型性能下降.

同时,分析双向门控循环单元(BiGRU)中隐藏层特征维度 T 对本文模型的影响,结果如表 5 所示.

表 5 BiGRU 隐藏层维度 T 的影响

Table5 Influence of BiGRU hidden layer dimension T			
T	AUC	MRR	nDCG@5
100	0.671 523	0.320 318	0.356 986
200	0.673 604	0.325 952	0.362 454
300	0.661 130	0.311 697	0.340 066

从表 5 中可以发现,随着 BiGRU 中隐藏层数的增加,模型性能也上升,但当隐藏层数为 300 时,模型性能下降,这可能是因为过拟合造成的,因此要选择适中的隐藏层维度.

5 结束语

本文提出了一种基于多通道 CNN-BiGRU 与多特征融合方法(CBMR 模型)用于新闻推荐中,与之前方法相比,本文模型一方面是基于多特征学习的,另一方面充分发挥卷积神经网络(CNN)、双向门控循环单元(BiGRU)和注意力机制的优势提取特征.模型先在词向量中融入了实体嵌入向量,利用 CNN 提取局部特征,使用 BiGRU 提取词语间的长距离依赖关系,获取丰富的语句特征.然后,借助注意力网络融合多个特征,构建精确的新闻表征.在用户表征中,采用多头自注意力机制提取用户历史浏览记录中新闻间的交互关系,捕捉用户兴趣,完成推荐.

通过真实数据集上的实验,证明了 CBMR 模型的有效性,提升了新闻推荐的准确性.然而,本文未考虑用户其他属性信息,下一步将探索对用户个性化特征的建模,继续提升推荐的性能.

参考文献:

[1] 吴方照,武楚涵,安鸣霄,等.基于深度学习的个性化新闻推荐[J].南京信息工程大学学报(自然科学版),2019,11(3):278-285. DOI: 10.13878/j.cnki.jnuist.2019.03.005.
WU F Z, WU C H, AN M X, et al. Personalized news recommendation based on deep learning[J]. Journal of Nanjing University of Information Science and Technology(Natural Science Edition), 2019, 11(3): 278-285. DOI: 10.13878/j.cnki.jnuist.2019.03.005.

[2] WU C H, WU F Z, AN M X, et al. NPA: neural news recommendation with personalized attention[C]// Proceedings of the 25th ACM SIGKDD International

Conference on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019: 2576-2584. DOI: 10.1145/3292500.3330665.

[3] AN M X, WU F Z, WU C H, et al. Neural news recommendation with long-and short-term user representations[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy: ACL, 2019: 336-345. DOI: 10.18653/v1/P19-1033.

[4] ZHU Q N, ZHOU X F, SONG Z L, et al. DAN: deep attention neural network for news recommendation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 5973-5980. DOI: 10.1609/aaai.v33i01.33015973.

[5] RICHARDSON M, DOMINOWSKA E, RAGNO R. Predicting clicks: estimating the click-through rate for new ads [C]//Proceedings of the 16th International Conference on World Wide Web. Banff, Alberta: ACM, 2007: 521-530. DOI: 10.1145/1242572.1242643.

[6] RENDLE S. Factorization machines[C]//Proceedings of 2010 IEEE International Conference on Data Mining. Sydney, NSW, Australia: IEEE, 2010: 995-1000. DOI: 10.1109/icdm.2010.127.

[7] CHENG H T, KOC L, HARMSSEN J, et al. Wide & deep learning for recommender systems[C]//Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston: ACM, 2016: 7-10. DOI: 10.1145/2988450.2988454.

[8] GUO H F, TANG R M, YE Y M, 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, Australia: ACM, 2017. DOI: 10.24963/ijcai.2017/239.

[9] KIM Y. Convolutional neural networks for sentence classification[C]//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics, 2014. DOI: 10.3115/v1/d14-1181.

[10] WANG H W, ZHANG F Z, XIE X, et al. DKN: deep knowledge-aware network for news recommendation[C]//Proceedings of 2018 World Wide Web Conference. Lyon, France: ACM, 2018: 1835-1844. DOI: 10.1145/3178876.3186175.

[11] WU C H, WU F Z, AN M X, et al. Neural news recommendation with topic-aware news representation [C]//Proceedings of the 57th Annual Meeting of the

- Association for Computational Linguistics. Florence, Italy: Association for Computational Linguistics, 2019: 1154-1159. DOI: 10.18653/v1/p19-1110.
- [12] WU C H, WU F Z, GE S Y, et al. Neural news recommendation with multi-head self-attention[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China: Association for Computational Linguistics, 2019: 6389-6394. DOI: 10.18653/v1/d19-1671.
- [13] YANG Z C, YANG D Y, DYER C, et al. Hierarchical attention networks for document classification [C]//Proceedings of the 2016 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California: Association for Computational Linguistics, 2016: 1480-1489. DOI: 10.18653/v1/n16-1174.
- [14] OKURA S, TAGAMI Y, ONO S, et al. Embedding-based news recommendation for millions of users[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax: ACM, 2017: 1933-1942. DOI: 10.1145/3097983.3098108.
- [15] WU F Z, QIAO Y, CHEN J H, et al. MIND: a large-scale dataset for news recommendation[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2020: 3597-3606. DOI: 10.18653/v1/2020.acl-main.331.
- [16] 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, Qatar: Association for Computational Linguistics, 2014: 1532-1543. DOI: 10.3115/v1/d14-1162.
- [17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. arXiv preprint arXiv:1706.03762, 2017.
- [18] KINGMA D, BA J. Adam: a method for stochastic optimization [C]//3rd International Conference on Learning Representations. San Diego:ICLR, 2015.
- [19] 王海艳, 胡阳. 基于会话的双层注意力机制新闻推荐方法[J]. 南京邮电大学学报(自然科学版), 2021, 41(1): 99-106. DOI: 10.14132/j.cnki.1673-5439.2021.01.012.
- WANG H Y, HU Y. News recommendation method for session-based neural neuralnetwork with two-layer attention mechanism[J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2021, 41(1): 99-106. DOI: 10.14132/j.cnki.1673-5439.2021.01.012.

作者简介:

王凯丽 女,(1995-),硕士研究生.研究方向为自然语言处理、推荐系统. E-mail: karry_eva@163.com

李海明 男,(1964-),博士,教授.研究方向为智能信号处理的基础和应用研究.