

基于注意力增强的热点感知新闻推荐模型

丁 琪^{1,2}, 田 莹^{1,2}, 孙国栋^{1,2}

(1. 北京林业大学信息学院, 北京 100083; 2. 国家林业草原林业智能信息处理工程技术研究中心, 北京 100083)

摘要: 完全个性化的新闻推荐工作通常只基于用户兴趣,可能会导致推荐结果与点击过的内容过于相似甚至重复. 事实上即使一些热点新闻并不完全符合用户兴趣,用户也可能希望点击类似的新闻. 目前基于热点的新闻推荐方法不能很好挖掘潜在新闻的热点特征、灵活平衡用户兴趣和热点特征. 本文提出一种新颖的注意力增强的热点感知新闻推荐模型(Hotspot-aware Attention enhanced model, HAN),充分利用注意力网络和自注意力网络等深度神经网络的优势,在个性化推荐中将个性化兴趣与新闻热点性进行更好平衡与利用. 该模型包括新闻编码器、热点特征提取器、用户兴趣提取器和点击预测器四个组件. 提出一个热点特征提取器,使用注意力网络动态聚合热点新闻学习热点表示以更好挖掘热点特征;提出一个新颖的点击预测器来灵活融合热点特征、用户兴趣和候选新闻,以提升候选新闻的点击预测准确率. 真实数据集上的实验表明HAN模型在AUC(Area Under the Curve of ROC)和F1两项指标上分别提升了7.51%和8.63%,且能够有效缓解用户冷启动问题.

关键词: 新闻推荐;热点感知;自注意力网络;注意力网络;卷积神经网络

基金项目: 国家重点研发计划(No.2018YFC1603305)

中图分类号: TP391.1

文献标识码: A

文章编号: 0372-2112(2023)01-0093-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210570

HAN: Hotspot-Aware Attention Enhanced News Recommendation

DING Qi^{1,2}, TIAN Xuan^{1,2}, SUN Guo-dong^{1,2}

(1. School of Information Science and Technology, Beijing Forestry University, Beijing 100083, China;

2. Engineering Research Center for Forestry-Oriented Intelligent Information Processing of National Forestry and Grassland Administration, Beijing 100083, China)

Abstract: Personalized news recommendation is usually based on users' interests only, which may cause the recommendation results to be too similar with or even repeat the content that has been clicked. In fact, even if some hot news may not meet the user's interests, users may also want to click on similar news. At present, hotspot-based approaches usually can not well mine the potential news hotspot features and flexibly balance the user interest and hotspot features. In this paper, a hotspot-aware attention enhanced model (HAN) for news recommendation is proposed, which makes full use of the advantages of deep neural networks such as attention networks and self-attention networks to better balance and utilize personalized interests and news hotspot in recommendation algorithm. HAN includes four components: news encoder, hotspot feature extractor, user interests extractor and click predictor. In order to effectively mine hotspot features, a hotspot feature extractor is proposed, which uses an attention network to dynamically aggregate hot news and learn hotspot feature representation; in order to improve the accuracy of predicting the click probability of candidate news, a click predictor is proposed to flexibly fuse hotspot features, user interests feature and candidate news representation. Experiments on a real-world dataset show that the area under the curve of ROC (AUC) and F1 increase by 7.51% and 8.63% respectively. At the same time, the model also helps to alleviate the cold-start problem of users.

Key words: news recommendation; hotspot-aware; self-attention networks; attention networks; convolutional neural networks

Foundation Item(s): National Key R&D Program of China (No.2018YFC1603305)

1 引言

随着互联网的快速发展,越来越多的人习惯上网获取新闻.然而,每天产生的大量新闻使得用户无法通过浏览所有内容来找到真正感兴趣的新闻.新闻推荐可以帮助用户快速找到感兴趣的新闻,成为有效缓解新闻信息过载的重要方式^[1-4].

在当前已有的研究工作中,很多工作利用用户的新闻点击历史等行为数据挖掘用户的个人兴趣,仅基于用户兴趣模型进行推荐^[5,6].然而,这种完全个性化的推荐方式可能会导致推荐结果与点击过的内容过于相似甚至重复.事实上,热点新闻对用户有着重大影响^[7,8].例如,当发生过诸如马航 370 失踪或新冠疫情爆发等热点事件时,对于一个平时只喜爱关注体育新闻和 IT (Information Technology) 新闻的用户来说,他(她)也可能想要点击类似的热点新闻.也就是说,用户想要点击的新闻不仅与个人的兴趣偏好相关,还可能与某些热点新闻相关.目前也一些新闻网站采取了推荐热点新闻的策略,但他们对所有用户提供相同的推荐,无法应对用户兴趣的差异化.因此,如何将个性化兴趣与新闻热点性在推荐算法中更好地平衡与利用,是一个值得深入研究的问题.有一些工作尝试将热点与个性化推荐相结合^[9-11].例如,文献[9]采用了词的流行度、出现度(关键词的出现频率)和关键词之间的共现度来识别流行文章,并通过手工设置权重的方式将文章的流行程度和用户的兴趣得分相加,得到推荐得分.文献[10]提出了一种混合方法,将新闻流行度和用户配置文件结合起来推荐个性化的流行新闻,用余弦相似度来计算一条新闻与流行度或个人兴趣的匹配程度.然而,这些工作通常采用传统的机器学习方法,使用手工设计的权重来平衡用户的兴趣和热点特征.事实上,手工设计的特征通常不是最优的,往往无法深度挖掘新闻的语义信息,无法提取潜在热点特征,也无法灵活平衡用户兴趣和热点特征.

近年来,深度学习技术发展迅速^[12,13],基于深度学习的方法已经成为新闻推荐的主流方法^[5,14-17].它们使用深层神经网络来提取新闻的深层语义特征,并通过用户点击历史学习用户兴趣的精确表示,有效提高了新闻特征和用户兴趣特征的表达能力.然而,目前还没有利用深度学习来提取热点特征、平衡用户兴趣和热点特征的工作.此外,对于点击历史很少的用户,基于兴趣的推荐模型往往无法产生有效的推荐结果^[18,19].

基于以上原因,本文提出一种基于注意力增强的热点感知新闻推荐模型(Hotspot-Aware Attention Enhanced News Recommendation Model, HAN). HAN 使用深度学习技术提取热点特征及用户兴趣特征,并能够将这两种因素灵活地平衡以用于新闻推荐. HAN 包括

新闻编码器、热点特征提取器、用户兴趣提取器和点击预测器四个组件,其中,设计了带有自注意力网络的新闻编码器,提取不连续的文本特征,弥补卷积神经网络(Convolutional Neural Networks, CNN)能提取局部连续特征的限制;提出带有注意力网络的热点特征提取器,动态聚合热点新闻来学习热点表示;用户兴趣提取器源于 DKN (Deep Knowledge-aware Network for news recommendation) 模型^[20],其使用注意力网络获得用户兴趣表示;提出一个新颖的点击预测器,在其中尝试了 3 种不同的融合方法用于融合平衡热点特征与用户兴趣,并预测候选新闻的点击概率.

2 相关工作

2.1 基于深度学习的新闻推荐

近年来,深度学习技术的不断创新推动了新闻推荐领域的蓬勃发展.文献[14]提出使用降噪自编码器(Denoising Auto-Encoder, DAE)从新闻正文中学习新闻表示,并使用门控循环单元(Gated Recurrent Unit, GRU)网络从用户的新闻点击历史中学习用户的兴趣表示.文献[20]设计了一种知识感知的卷积神经网络将单词、实体、实体的上下文作为 CNN 的多个通道输入从新闻标题中学习新闻表示,并使用注意力机制(Attention Mechanism, AM)有轻重地聚合用户的点击新闻表示向量来学习用户兴趣表示.文献[5]提出一种个性化的词级和新闻级的注意力网络分别学习新闻表示和用户兴趣表示,在学习新闻表示时,利用 CNN 处理新闻标题单词序列,通过捕获新闻标题中单词的局部上下文学习单词的语义表示;再基于用户 ID (Identification) 设计出个性化的 AM 来为不同用户区分不同的重要单词;在学习用户表示时,同样基于用户 ID 设计出个性化的 AM 来为不同用户区分不同的重要新闻.文献[21]使用了门控循环网络(Gated Recurrent Unit, GRU)分别从用户点击历史和用户的 ID 嵌入向量中学习长期兴趣表示和短期兴趣表示,并探索了 2 种方法对用户的长期兴趣表示和短期兴趣表示进行融合,以获得最终的用户表示.文献[16]应用了多头自注意力机制(Multi-Head Self-Attention, MHSA)来学习新闻表示,从新闻标题中建模单词之间的交互来学习单词的上下文表示,并使用 AM 选择重要的词语来学习更具有信息量的新闻表示;同时使用 MHSA 捕捉新闻之间的长距离依赖关系学习用户兴趣表示,以增强用户兴趣特征表达.

然而,这些研究中的大多数都是关注新闻建模或用户的个人兴趣建模,忽略了用户的个人点击历史以外的信息,即忽略了除个人兴趣外对用户有重大影响的热点这一因素,这可能会限制推荐效果.

2.2 融合热点的新闻推荐

文献[8]指出热点新闻对新闻推荐具有重要意义. 尽管一个热点新闻可能并不符合用户平时的兴趣, 但用户想要点击类似热点新闻的可能性依然存在甚至很大(正如在引言中所举的例子). 因此, 热点新闻可以被作为新闻推荐模型中的辅助信息.

在前人的工作中, 有一些研究已经提出了将热点新闻(或称流行新闻)与用户兴趣相结合的方法进行新闻推荐. 文献[10]提出了一种将基于流行度的新闻推荐器和基于用户配置文件(profile)的新闻推荐器相结合的混合推荐方法, 其流行新闻采集自推特网站, 并采用余弦相似度来计算一个候选新闻和流行新闻或者个人兴趣的匹配程度. 与之类似, 文献[22]同样从推特网站收集众多流行新闻, 分别计算候选新闻与所有流行新闻的相似度, 并将相似度值累加作为该篇候选新闻的流行度权重以预测其流行性. 文献[9]将用户兴趣计算和事件流行程度预测相结合来获得用户对热门目标文章的个性化偏好. 其中, 该文献采用了词的流行度、出现度(关键词的出现频率)和关键词之间的共现度来识别流行文章, 并通过手工设置权重的方式将文章的流行程度和用户的兴趣得分相加, 得到推荐得分. 文献[23]考虑将用户兴趣与其当前位置的热门新闻相结合, 获取用户移动位置信息并通过推特提供的 API(Application Program Interface)直接检索该位置附近的热点新闻话题, 根据这些热点新闻话题对原来仅由用户个性化兴趣决定的新闻推荐列表做出重新排序. 文献[11]综合考虑了文章流行度、流行趋势、用户个人配置文件和位置偏好, 通过计算新闻文章和流行推文(tweets)之间的余弦相似度来获得流行新闻, 同时计算新闻文章和具有流行趋势的推文之间的 TF-IDF(Term Frequency-Inverse Document Frequency)值来获得具有流行趋势的新闻, 最后设计了一个手动调节的参数来帮助用户选择被推荐的流行新闻和具有流行趋势新闻的比重.

可见, 以上的大部分工作使用传统的机器学习方法提取新闻特征和热点特征, 手工设计权重聚合用户兴趣和热点特征. 这种手工的特征通常不是最优的, 它无法有效利用新闻文本中的词语顺序等潜在的语义信息来精确建模新闻, 也无法有效挖掘用户兴趣和热点新闻之间的潜在关联, 此外, 它也无法将用户兴趣和热点灵活地融合平衡以得到更好的推荐结果.

综上所述, 目前基于深度学习的新闻推荐工作大多只关注用户点击和新闻文本, 忽略了热点等其他因素, 同时, 大多数的融合了热点的新闻推荐工作使用的是传统的机器学习方法, 无法提取新闻及用户的深层次特征, 无法灵活融合用户兴趣和热点, 限制了新闻推

荐的效果. 基于以上 2 点原因, 本文的方法将使用深度学习技术提取热点特征及用户兴趣特征, 并将这两种因素灵活地融合以用于新闻推荐.

3 问题定义

3.1 热点度量

热点(或流行度)度量在之前的研究中经常被定义为一个具体数值, 例如视频的观看数、微博的评论数或者转发数等^[24, 25], 文献[26]使用新闻文章的分享数作为流行度的度量指标. 文献[27]使用新闻的浏览数作为流行度的度量. 另外有研究^[28]表明, 文章的点击数和文章的评论数或其他指标之间有较强的相关性. 也就是说, 使用何种指标来度量新闻的热点性并无明显差别. 因此, 在本论文中, 使用用户对新闻的总点击量作为其热点性的度量.

3.2 问题描述

为了更准确地描述新闻推荐问题, 本节对论文中的符号阐述如下. 对于一个用户 u , 其新闻点击历史记为 $\{t_1^u, t_2^u, \dots, t_{N_u}^u\}$, 其中, $t_i^u (i=1, \dots, N_u)$ 是用户 u 点击历史中第 i 篇新闻的标题, N_u 是用户 u 点击历史中新闻总数量. 对于过往一段时间 v 内的热点新闻记为 $\{p_1, p_2, \dots, p_M\}$, 其中, $p_j (j=1, \dots, M)$ 是该段时间内第 j 篇热点新闻的标题, M 是该段时间内热点新闻的总数量. 每个点击历史新闻标题 t 或热点新闻标题 p 由一系列字 $[w_1, w_2, \dots, w_n]$ 组成, 其中, n 为标题字数. 本文使用新闻标题作为模型的输入, 是由于新闻标题是影响用户点击选择的决定性因素. 本模型的目标是, 根据给定的用户点击历史和过往一段时间内的热点新闻, 来预测用户 u 是否会点击之前他从未见过的候选新闻 x (即待预测推荐得分的新闻). 表 1 总结了本文出现的重要变量符号.

4 模型与方法

本节将首先介绍所提出的 HAN 模型的整体架构, 然后逐一详细介绍 HAN 模型中的新闻编码器、热点特征提取器、用户兴趣提取器和点击预测器四个关键组件, 最后再介绍模型的损失函数及惩罚项等训练细节.

4.1 总体框架

HAN 包含四个主要组件: 新闻编码器、热点特征提取器、用户兴趣提取器、点击预测器. HAN 的输入包括过去一段时间内的一些热点新闻、某个用户的点击历史和一条候选新闻. 对于每一条新闻, 通过新闻编码器可以获得其新闻表示, 为提升对文本特征的提取能力, 本模型应用了一个字级的自注意力网络来提取不连续的文本特征, 弥补 CNN 只能提取局部连续特征的限制. 在获得了每个新闻表示的基础上, 为了基于热点新闻

表 1 本文常用符号

符号	描述	符号	描述
t_i^u	第 i 篇新闻的标题	$o_{i_u, h, x}$	$o_{i_u, x}$ 与 $o_{h, x}$ 融合后的表示
N_u	用户 u 点击历史中新闻总数量	d	字嵌入维度
p_j	第 j 篇热点新闻的标题	N_f	卷积核数量
M	热点新闻的总数量	F	卷积核尺寸
w_n	标题中第 n 个字	$2k+1$	滑动窗口尺寸
e	新闻标题表示	s	窗口滑动步长
x	候选新闻表示	d_a	自注意力网络的隐藏单元数量
A	自注意力网络的权重矩阵	r	自注意力网络中关注标题的 r 个不同部分
h	热点新闻表示	l_c	L_2 正则项系数
i_u	用户 u 的兴趣表示	p_c	p 惩罚项的系数
$o_{i_u, x}$	h 与新闻 x 拼接后的表示	l	学习率
$o_{i_u, h, x}$	i_u 与新闻 x 拼接后的表示	B	批尺寸

表示提取热点特征,设计了一个热点特征提取器,它使用了注意力网络来关注和候选新闻相关的热点新闻,从而能够使用不同的权重来动态聚合热点新闻表示.为了提取用户兴趣特征,使用用户兴趣提取器来聚合用户的点击历史^[20].在获得了热点特征表示和用户兴趣表示的基础上,为了预测用户对于候选新闻的点击

概率,提出了一个点击预测器,它可以融合热点特征表示、用户兴趣表示和候选新闻表示,以平衡用户兴趣和热点特征,并最终预测候选新闻的点击概率.图 1 展示了 HAN 模型架构,四个组件的具体设计将在接下来的小节中一一介绍.

4.2 新闻编码器

通常,用户根据新闻标题来决定是否点击该新闻,故 HAN 中的新闻编码器基于新闻标题来学习新闻表示.新闻编码器由三层架构组成.第一层和第二层沿用传统的词嵌入和 CNN 设置. CNN 擅长提取局部连续文本特征,却无法有效提取不连续的文本特征,为了弥补 CNN 的局限性,本模型增设了自注意力模块,以提升新闻表示的准确性.

第一层是词嵌入.它将新闻标题中的单词序列转换成低维密集向量序列,学习新闻标题的分布式表示.按照第 3.2 节中的符号定义,表示长度为 n 的点击历史新闻和热点新闻标题单词序列分别记为 $t = w_{1:n} = [w_1, w_2, \dots, w_n]$ 和 $p = w_{1:n} = [w_1, w_2, \dots, w_n]$,通过在大型语料库的预训练嵌入,它被转换成一个向量序列 $w_{1:n} = [w_1, w_2, \dots, w_n] \in \mathbf{R}^{d \times n}$,其中, d 是词嵌入向量的维度.

第二层是 CNN^[29]. CNN 通过其卷积操作可以捕捉一个新闻标题中的局部连续特征.局部连续特征指的是从相邻的几个字中提取的特征,字的数量一般不会

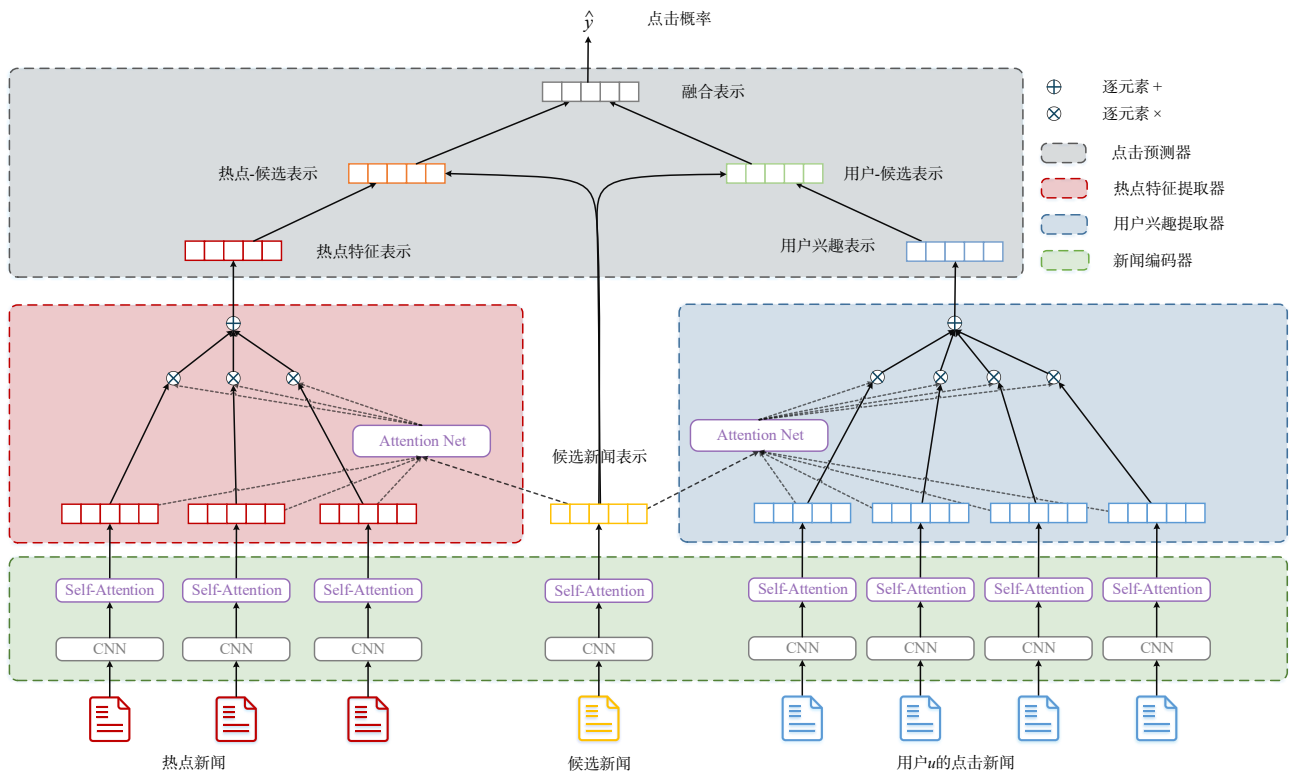


图 1 HAN 模型的架构

大于卷积核的尺寸. 通常, 新闻标题字词信息丰富而集中, 新闻标题中的局部上下文对表示新闻很关键. 例如, 在“迎新会精彩瞬间”这一新闻标题中, 局部连续的字“迎”和“新”对于揭示新闻主题很重要. 因此, 使用 CNN 并设置滑动窗口来捕捉标题中的局部上下文进而学习字的上下文表示. 该滑动窗口可通过滑动步长来调节. 第 i 个字的表示记为 c_i , 其计算方式如式(1)所示.

$$c_i = f(W_w * w_{(i-k):(i+k)} + b_w) \quad (1)$$

其中, $w_{(i-k):(i+k)}$ 表示从标题位置 $(i-k)$ 到位置 $(i+k)$ 的字嵌入向量的拼接, 同时, 滑动窗口的大小为 $2k+1$. f 代表一种非线性变换, $*$ 为卷积操作. $W_w \in \mathbf{R}^{N_f \times (2k+1)d}$ 和 $b_w \in \mathbf{R}^{N_f}$ 是 CNN 卷积核的可训练参数, 其中 N_f 为 CNN 卷积核的数量, 并且采用多种尺寸的卷积核提取位置 $(i-k)$ 到位置 $(i+k)$ 中的局部上下文特征. CNN 层的输出为字的上下文表示向量序列, 记为 $[c_1, c_2, \dots, c_m]$, 其中 m 是滑动窗口数, 它取决于滑动步长 s .

第三层是自注意力网络. 由于 CNN 关注于局部连续特征, 因此新闻标题中一些距离较远的关键信息可能被遗漏. 例如, 在“点亮朋友圈的迎新现场精彩热图集锦”这一新闻标题中, “迎新”和“热图”是具有重要信息的关键字词, 它们之间有四个字的距离, CNN 对于此类有距离不连续的特征的探查能力并不明显(不连续特征是指从几个不相邻的字词中提取的特征, 其关键字词之间的距离一般大于卷积核的大小). 文献[30]提出在双向长短期记忆网络(Bidirectional Long Short-Term Memory, Bi-LSTM)的基础上引入自注意力网络来学习句子的表示, 获得了良好的效果. 受该工作启发, 本文使用自注意力网络来增强 CNN, 以捕捉新闻标题中不连续特征. 此外, 一些长标题句是由多个成分组成, 为了更好地表示标题的整体语义, 需要关注标题的 r 个不同部分. 为了简单表示, 使用 C 的定义如式(2)所示, 其中 $1 \leq i \leq m$. C 是自注意力网络的输入, 输出是其权值矩阵 A ^[30], 其计算方法如式(3)所示.

$$C = [c_1, c_2, \dots, c_m] \quad (2)$$

$$A = \text{softmax}(W_{s_2} \tanh(W_{s_1} C^T)) \quad (3)$$

其中, $W_{s_1} \in \mathbf{R}^{N_f \times d_a}$ 和 $W_{s_2} \in \mathbf{R}^{d_a \times r}$ 表示自注意力网络中的参数, d_a 和 r 是可以被设置的超参数. 接下来, 将矩阵 A 和整体词表示 C 相乘得到 r 部分的加权和, 即为每个新闻标题的最终表示向量 e . 其计算方法如式(4)和(5)所示.

$$E = AC \quad (4)$$

$$e = \text{flatten}(E) \quad (5)$$

这里的 $\text{flatten}(\cdot)$ 表示将矩阵 $E \in \mathbf{R}^{r \times N_f}$ 转变形状为向量 $e \in \mathbf{R}^\delta$ 的拉平操作, $\delta = rN_f$. 最终的新闻表示向量 e

将作为接下来热点特征提取器和用户兴趣提取器的输入.

4.3 热点特征提取器

为了判断候选新闻的热点潜质, 本文将热点因素引入深度神经网络架构中, 首先利用热点新闻提取热点特征, 再利用热点特征计算候选新闻的热点性. 本节的热点信息提取器旨在根据以往一段时间内的热点新闻来学习热点特征表示. 对于给定的热点新闻 $\{p_1, p_2, \dots, p_M\}$, 其向量表示被记为 $\{e(p_1), e(p_2), \dots, e(p_M)\}$, 为了获得热点特征表示, 一种直接的方式是将热点新闻表示做平均处理, 如式(6)所示.

$$h = \frac{1}{M} \sum_{j=1}^M e(p_j) \quad (6)$$

然而, 热点新闻的主题可能是多方面的, 在考虑一篇候选新闻是否具有潜在热点性时, 更希望关注和当前候选新闻相关的热点信息. 为了表征不同方面的热点信息, 本文使用注意力机制来选择和当前候选新闻相关的热点新闻. 具体来说, 对于一个热点新闻 p_j 和一个候选新闻 x , 将候选新闻表示 $e(x)$ 作为热点查询向量(query vector), 热点新闻表示记为 $e(p_j)$. 热点特征提取器通过计算热点查询向量和热点新闻表示 $e(p_j)$ 之间的相似度来学习第 j 个热点新闻的注意力权重 α_j , 其计算方法如式(7)所示.

$$\alpha_j = \text{softmax}(e(p_j)^T e(x)) = \frac{\exp(e(p_j)^T e(x))}{\sum_{j=1}^M \exp(e(p_j)^T e(x))} \quad (7)$$

对于候选新闻 x , 其最终热点特征表示 h 为所有热点新闻表示的注意力加权和, 计算方法如式(8)所示.

$$h = \sum_{j=1}^M \alpha_j e(p_j) \quad (8)$$

在第 5.3.3 节中, 将在实验中对比基于注意力网络的热点特征提取方法与普通的做平均方法的效果, 以证明本模型中热点特征提取器的优越性.

4.4 用户兴趣提取器

用户兴趣提取器旨在根据用户的点击历史来学习用户兴趣表示. 和热点信息相似, 用户的兴趣通常也是多方面的, 在考虑一个用户是否会点击一篇候选新闻时, 更希望关注和当前候选新闻相关的兴趣. 为了表征用户不同的兴趣, 本模型使用注意力机制来选择和当前候选新闻相关的点击历史. 具体而言, 对于用户 u 的点击新闻 t_i^u 和当前候选新闻 x , 将候选新闻表示 $e(x)$ 作为兴趣查询向量, 点击新闻表示记为 $e(t_i^u)$. 用户兴趣提取器通过计算兴趣查询向量和点击新闻表示 $e(t_i^u)$ 之间的相似度来学习第 i 个点击新闻的注意力权重 β_i , 其计算方法如式(9)所示.

$$\beta_i = \text{softmax}(\mathbf{e}(t_i^u)^T \mathbf{e}(x)) = \frac{\exp(\mathbf{e}(t_i^u)^T \mathbf{e}(x))}{\sum_{i=1}^{N_u} \exp(\mathbf{e}(t_i^u)^T \mathbf{e}(x))} \quad (9)$$

对于候选新闻 x , 其最终的用户兴趣表示 \mathbf{i}_u 为所有点击新闻表示的注意力加权和, 计算方法如式(10)所示.

$$\mathbf{i}_u = \sum_{i=1}^{N_u} \beta_i \mathbf{e}(t_i^u) \quad (10)$$

4.5 点击预测器

点击预测器旨在预测一个用户在每个候选新闻上的点击概率得分. 与其他工作^[5, 15, 16, 20, 21]不同的是, 本文需要考虑候选新闻表示、用户兴趣表示和热点特征表示3个向量, 而非只有前2个. 本模型采用了如下策略: 首先, 将候选新闻表示分别与热点特征表示和用户兴趣表示相拼接, 得到热点-候选向量 $\mathbf{o}_{h,x}$ 和用户-候选向量 $\mathbf{o}_{i_u,x}$, 计算方法分别如式(11)、(12)所示(拼接方法用 \oplus 表示).

$$\mathbf{o}_{h,x} = \tanh(\mathbf{w}_1(\mathbf{h} \oplus \mathbf{e}(x)) + b_1) \quad (11)$$

$$\mathbf{o}_{i_u,x} = \tanh(\mathbf{w}_2(\mathbf{i}_u \oplus \mathbf{e}(x)) + b_2) \quad (12)$$

然后, 通过打分模块计算用户点击该候选新闻的概率得分, 计算方法如式(13)、(14)所示.

$$\mathbf{o}_{i_u,h,x} = \text{fuse}(\mathbf{o}_{i_u,x}, \mathbf{o}_{h,x}) \quad (13)$$

$$\hat{y}_{u,x} = \text{sigmoid}(\mathbf{w} \mathbf{o}_{i_u,h,x} + b) \quad (14)$$

其中, $\text{fuse}(\cdot, \cdot)$ 函数表示通过融合方法将两个向量相融合, 最后, 采用 sigmoid 非线性变化作为激活函数用来预测用户 u 对于当前候选新闻 x 的点击概率 $\hat{y}_{u,x}$, 其中的 \mathbf{w} 和 b 都是可学习的参数. 对于 $\text{fuse}(\cdot, \cdot)$ 函数, 在第 5.3.4 节实验中, 尝试了 3 种不同的融合方法, 具体计算方式如式(15)~(17)所示.

(1) 逐元素加 (element-wise summation).

$$\mathbf{o}_{i_u,h,x} = \tanh(\mathbf{w}_3 \mathbf{o}_{i_u,x} + \mathbf{w}_4 \mathbf{o}_{h,x}) \quad (15)$$

该方法首先将热点-候选向量 $\mathbf{o}_{h,x}$ 和用户-候选向量 $\mathbf{o}_{i_u,x}$ 输入进一个线性层, 其中 \mathbf{w}_3 和 \mathbf{w}_4 是可学习的参数, 然后将这两个向量做逐元素加法, 最后使用 tanh 非线性变换作为激活函数.

(2) 带权重的逐元素加 (weighted element-wise summation).

$$\mathbf{o}_{i_u,h,x} = \tanh(\mathbf{w}_5(\lambda \mathbf{o}_{i_u,x} + (\mathbf{1} - \lambda) \mathbf{o}_{h,x}) + b_5) \quad (16)$$

该方法中, λ 是一个用来衡量 $\mathbf{o}_{i_u,x}$ 和 $\mathbf{o}_{h,x}$ 重要程度的参数, 它随着训练而自动更新. $\mathbf{1}$ 代表元素全为 1 的矩阵, \mathbf{w}_5 和 b_5 是可学习的参数, 最后将 tanh 非线性变换作为激活函数.

(3) 拼接 (concatenation).

$$\mathbf{o}_{i_u,h,x} = \tanh(\mathbf{w}_6(\mathbf{o}_{i_u,x} \oplus \mathbf{o}_{h,x}) + b_6) \quad (17)$$

该方法将热点-候选向量 $\mathbf{o}_{h,x}$ 和用户-候选向量 $\mathbf{o}_{i_u,x}$ 做拼接, 之后将其结果依次送入线性层与 tanh 非线性层, 其中 \mathbf{w}_6 和 b_6 是可学习的参数.

4.6 损失函数及惩罚项

在训练时, 本文采用用户最近的点击新闻作为正样本, 同一时段出现但未点击的新闻作为负样本. 将一条训练样本记为 $s = (\{t_1^u, t_2^u, \dots, t_{N_u}^u\}, \{p_1, p_2, \dots, p_M\}, x, y_{u,x})$, 其中 $\{t_1^u, t_2^u, \dots, t_{N_u}^u\}$ 为用户 u 的点击历史集, $\{p_1, p_2, \dots, p_M\}$ 为同时段的热点新闻集, x 为当前的候选新闻. 每一个正样本的标签为 1, 即 $y_{u,x} = 1$; 而每一个负样本的标签为 0, 即 $y_{u,x} = 0$. 经过最后的输出层, 每个样本都会得到一个对其候选新闻 x 的概率估计 $\hat{y}_{u,x}$. 本文采用最小化负对数似然函数的方法来训练模型, 如式(18)所示.

$$L = - \left(\sum_{s \in S^+} y_{u,x} \log \hat{y}_{u,x} + \sum_{s \in S^-} (1 - y_{u,x}) \log (1 - \hat{y}_{u,x}) \right) \quad (18)$$

其中 S^+ 和 S^- 分别表示正样本集和负样本集.

为了避免过拟合情况的发生以及提高模型的泛化能力, 本文采用 L_2 正则化来降低模型复杂度. 另外, 为了避免新闻编码器中的自注意力网络总是赋予新闻标题的 r 个不同部分相同权重, 本文借鉴了文献[30]的做法, 引入了惩罚项 P , 具体计算方式如式(19)所示.

$$P = (\|\mathbf{A}\mathbf{A}^T - \mathbf{I}\|_F)^2 \quad (19)$$

其中, \mathbf{A} 是自注意力网络中的权重矩阵, \mathbf{I} 表示单位矩阵, $\|\cdot\|_F$ 表示给定矩阵的弗罗贝尼乌斯范数 (Frobenius Norm). 本文还设置 L_2 正则项的系数和 P 惩罚项的系数分别为 l_c, p_c , 并将它们与原损失函数 L 一起最小化.

5 实验

5.1 数据集描述

现有常见的新闻推荐公开数据集包括 Adressa、MIND (Microsoft News Dataset) 等, 这些数据集大都是非中文数据, 更重要的是, 没有或无法准确统计某一时间段内新闻的总点击量, 即无法确定热点新闻集, 所以本文构建了一个基于真实世界的新闻数据集, 其新闻和用户点击记录等数据来源于绿色新闻网. 作为北京林业大学重要的对内和对外窗口, 绿色新闻网旨在及时准确地报道学校的最新发展与成果, 其内容包括教学科研、校园文化、党建思政、学术交流、国内外林业等领域最新发展等. 本文选取了 2019 年 3 月~10 月的新闻和用户日志构建该数据集, 主要包括数据预处理和数据集构建两大步骤.

(1) 数据预处理. 首先对原始用户日志等数据进行预处理, 主要包括剔除冗余数据和数据规整化. 原始用户日志中存在很多重复的用户与新闻的交互数据, 这些重复的数据需要去除. 数据规整化是指用户、新闻

以及用户新闻之间的交互需要有完整的映射关系,因此,如果用户新闻交互文件中有某条交互记录,但是新闻文件中并没有该新闻的信息,此类映射不完整的数据也需要去除。

(2) 数据集构建. 由于绿色新闻网为学校新闻网站,相较于社会新闻网站,日更新量较少且用户量有限,因此,本文按月划分时间窗口,并按照以下规则选取用户点击历史、热点新闻和候选新闻:

①用户和候选新闻来自某双数月的用户及其点击记录. 另外,由于在点击历史非常少的用户上模型无法学习到有用的信息,本文摒除了点击历史数低于3的用户。

②用户的点击历史出自上一个单数月的该用户点击历史记录。

③热点新闻集同样出自上一个单数月发布的新闻,根据新闻的总点击量确定新闻的热点程度,按照总点击量从大到小排序形成热点新闻列表.(新闻总点击量中已扣除点击历史数低于3的用户所对应的点击数.)

一条样本的构成如表2所示. 它包含一个用户的 n 条点击新闻($n \geq 0$),与点击新闻同时段的前 m 条热点新闻($m \geq 0$,热点新闻按照总点击量从大到小排名),一条候选新闻,一个非1即0的标签. 标签代表了现实情境中用户对这条候选新闻的点击与否,该标签为1代表此样本为正样本,即用户真实点击了候选新闻,反之,标签为0代表此样本为负样本,即真实数据中用户未点击该候选新闻. 其中, n 和 m 的取值将在5.3.5小节中探究。

表2 一条样本的构成

点击新闻	热点新闻	候选新闻	标签
n 条	m 条	1条	1/0

本文采用用户最近的点击新闻作为正样本,同一时段出现但未点击的新闻作为负样本;且为使得正负样本数量平衡,本文在用户未点击的新闻中随机挑选了与正样本数相近的新闻来构造最终负样本. 该数据集的详细统计信息如表3所示. 其中,训练集和测试集按5:3的比例在数据集中划分。

表3 数据集统计

类别	数量
用户	7 020
新闻	1 846
标题平均字数	20.77
样本	184 616
正样本	92 660
负样本	91 956

另外,为了在实验设置中截取合适的新闻标题长度,本文对新闻标题长度分布进行了统计,统计结果如图2所示。

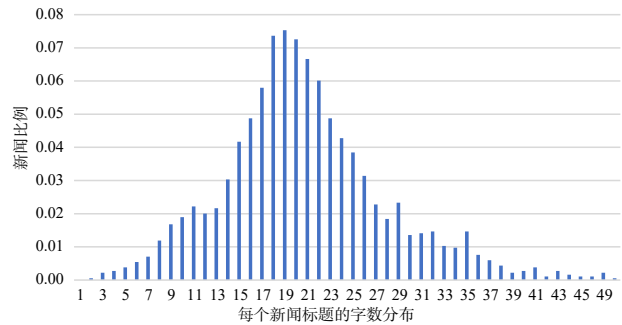


图2 新闻标题长度分布

5.2 实验设置

在实验中,根据对新闻标题字数分布的统计,如图2所示,确定了最大标题长度为29,该范围可以包含约85%的新闻标题的全部信息. 避免因输入模型的最大新闻标题长度过大而加剧计算时长,或者过小而损失新闻标题中的有用信息. 另外,本文采用经过预训练的Word2Vec嵌入^[31, 32]来初始化嵌入矩阵,并采用Adam^[33]算法作为梯度下降的优化器。

HAN模型中涉及众多超参数,每个超参数值的选择都经过了调优过程. 其中,词嵌入维度 d 的值从{20, 50, 100}中调优选择. 在CNN中,应用了多种尺寸的卷积核来提取特征,并将其拼接在一起作为字词的上下文表示. 具体来说,卷积核数量 N_f 从{4, 8, 16, 32, 64}中调节,卷积核 F 的多种尺寸从{[1], [1, 2], [1, 2, 3], [2, 3]}中调节. 在滑动窗口中,从{3, 5, 7}中挑选窗口尺寸 $(2k+1)$ 的值,滑动步长 s 的值同样从{3, 5, 7}中调节. 在自注意力网络中,隐藏单元数 d_a 从{10, 20, 30, 50}中挑选, r 则尝试了{1, 2, 3, 4, 5}5种取值. 另外,对于 L_2 正则项的系数 l_c ,尝试了{0.01, 0.1, 0.5}3种取值, P 惩罚项的系数 p_c 尝试了{0.1, 0.3, 0.5, 0.7, 0.9}5种取值,初始学习率 l 尝试了{0.005, 0.01, 0.05},批大小 B 尝试了{32, 64, 128, 256}.

本文采用新闻推荐领域同类型工作经常采用的ROC (Receiver Operating Characteristic) 曲线下的面积 (Area Under the Curve of ROC, AUC) 和调和平均(F1)作为模型的评价指标,并设置数据集最大训练轮次(epoch)为100,同时采用早停(early stopping)技巧避免出现拟合. 为保证结果的公正性,对于同一组超参数,通过随机初始化参数将模型独立训练10次,然后在测试集上测试它们,计算平均结果. 经过多次实验,本模型的最佳超参数配置如下: $d=50, N_f=8, F=[1, 2, 3]; 2k+$

$l=3, s=3, d_a=20, r=3; l_c=0.01, B=128$.

5.3 实验结果

5.3.1 模型性能测试

据本文所知,在目前基于深度学习的新闻推荐工作中还没有融合热点特征的模型,因此,本文采用不带知识图谱的DKN模型^[20]作为基线算法. DKN使用了卷积神经网络来学习新闻标题表示,并使用注意力网络从用户的点击新闻中学习用户兴趣表示,但DKN中并没有使用自注意力网络,也无热点新闻. 本文将DKN在本数据集上训练,将其视为表4中第4行的Hwo模型(不带自注意力网络和热点新闻的HAN模型). 为了测试HAN模型的推荐效果,在数据集上对HAN模型及其变体做了对比实验,主要包含以下三个方面:有无热点新闻、有无自注意力网络、有无点击新闻. 实验结果如表4所示,具体分析如下.

表4 不同模型的性能表现

模型	有无自注意力网络	有无热点新闻	有无点击新闻	AUC/%	F1/%
HAN	√	√	√	68.08	67.34
Hwoh (w/o hot news)	√	×	√	64.58	64.88
Hwos (w/o self-attention)	×	√	√	67.37	66.35
Hwo (originates from DKN, w/o self-attention nor hot news)	×	×	√	63.32	61.99
Hwoc (w/o clicked news)	√	√	×	64.20	63.59

(1) 热点特征的有效性

本文首先分别对改进前无热点特征的模型和改进后包含热点特征的模型进行训练,来检验热点特征提取器对推荐效果的改善情况. 为排除自注意力模块的影响,分别在有自注意力网络和无自注意力网络的情况下,对模型进行增加热点特征模块的优化. 实验结果如表4所示,在无自注意力网络的情况下,增加热点特征模块后(Hwos)其AUC和F1分别达到0.6737和0.6635,相较于不含热点特征模块(Hwo)的模型其AUC和F1分别提高了6.40%和7.03%;在有自注意力网络的情况下,增加热点特征模块后(HAN)其AUC和F1分别达到0.6808和0.6734,相较于不含热点特征模块(Hwoh)的模型其AUC和F1分别提高了5.42%和3.79%. 由于用户的点击行为不仅基于自身的兴趣,还受热点新闻的影响,因此,与之前的热点新闻相似的候选新闻可能更容易赢得用户的点击青睐.

尽管热点特征对于推荐准确性起着重要作用,但只含有热点特征不含有点击新闻的模型(Hwoc)的结果并不令人满意,相较于既含热点特征又含点击新闻的模型(HAN),其AUC和F1分别降低5.70%与5.57%. 该

结果证明,只含有热点特征并不能为新闻推荐模型提供完整足够的信息,因此点击新闻对于推荐模型来说是必不可少的. 换言之,基于点击新闻的用户兴趣和基于热点新闻的热点特征的同时利用,相较于单一使用的情况(仅含有用户兴趣或仅含有热点特征),更能够提升新闻推荐效果.

(2) 新闻编码器中自注意力网络的有效性

本文分别对改进前无自注意力网络的模型和改进后包含自注意力网络的模型进行训练,来检验自注意力网络对推荐结果的改善. 为排除热点特征提取器的影响,该实验分别在有热点新闻和无热点新闻的情况下,对模型进行增加自注意力网络的优化. 实验结果如表4所示,在无热点新闻的情况下,增加自注意力网络(Hwoh)其AUC和F1分别达到0.6458和0.6488,相较于不含自注意力网络(Hwo)的模型其AUC和F1分别提高了1.99%和4.66%;在有热点新闻的情况下,增加自注意力网络(HAN)其AUC和F1分别达到0.6808和0.6734,相较于不含自注意力网络(Hwos)的模型其AUC和F1分别提高了1.05%和1.49%. 实验结果表明,增加自注意力网络的模型比不含自注意力网络的模型表现得更好,说明在新闻特征提取器的卷积神经网络基础上增加自注意力网络的有效性,自注意力网络能够捕获标题中不连续的特征,作为能够捕获局部连续特征的卷积神经网络的补充,提高了新闻特征表达能力.

5.3.2 冷启动问题

为了验证HAN应对用户冷启动问题的有效性,本文对比了不同最大点击新闻数下增加热点新闻前后AUC的增长,如图3所示. 本实验从用户最大点击新闻数不少于4的情况下开始观察结果,舍弃了最大点击新闻数为1~3的情况,因为在过少的点击新闻情况下,可能无法从中找到与候选新闻相关的历史,无法看出其结果的差异. 实验结果显示,当用户最大点击新闻数较少(点击新闻数为4)时,引入热点新闻特征后AUC值提升了5.67%,远大于用户最大点击新闻数较多(点击新闻数为10)时,引入热点新闻特征后所带来的AUC的提升量2.59%. 该实验结果说明,热点信息对用户有着较大影响,特别是对点击新闻数量较少的新用户,融入热点信息可以有效提升推荐效果. 综上,HAN模型在较少历史的用户上也可以表现出良好的推荐效果,能够缓解用户冷启动问题.

5.3.3 不同融合方法的性能

为了更好地融合用户兴趣与热点特征,文本在实验中对点击预测器中的融合方法fuse(·, ·)进行了3种不同计算方式的尝试,分别如第4.5节式(15)~(17)所示. 其中,逐元素加的基本思想是将兴趣-候选向量与

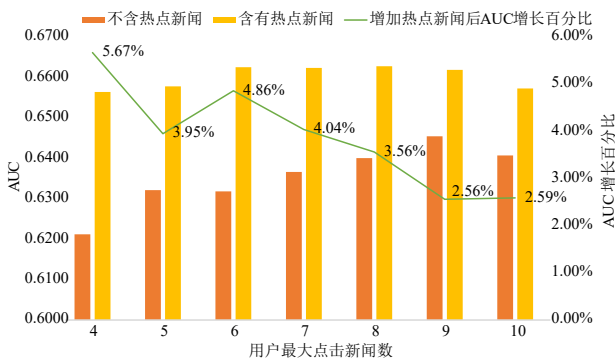


图3 不同最大点击新闻数下增加热点新闻前后 AUC 的值

热点-候选向量按照元素位置相加;带权重的逐元素加的思想是将兴趣-候选向量与热点-候选向量逐元素相加的基础上再引入一个衡量二者重要性程度的权重参数;拼接则是将兴趣-候选向量与热点-候选向量做拼接操作。如表5所示,实验结果表明拼接方法在AUC和F1指标上的表现均超过了前两种方法,这可能是由于拼接方法保留了匹配信息,并让神经网络学会动态地融合它们。

表5 点击预测器中3种不同融合方法的对比

方法	AUC/%	F1/%
逐元素加	64.68	65.17
带权重的逐元素加	67.01	66.59
拼接	68.08	67.34

5.3.4 热点注意力的效果

为了验证热点特征提取器中注意力网络对模型性能的提升,本文分别对改进前将热点新闻表示直接平均化来提取热点特征的模型,与改进后利用注意力网络根据当前候选新闻来聚合热点新闻表示的模型进行训练,分别如式(6)~(8)所示。如表6所示,实验结果表明,含有注意力网络的热点特征提取器使得模型在AUC和F1指标上均有所提升。注意力网络能够根据当前候选新闻动态地聚合热点新闻表示,使得热点特征表示更具表达力,从而使得新闻推荐效果得以提升。

表6 热点提取器中注意力网络的效果

方法	AUC/%	F1/%
不含注意力网络的直接平均法	66.61	65.54
含有注意力网络	68.08	67.34

5.3.5 超参数敏感性

HAN模型涉及了一系列超参数,本文在此讨论两个重要超参数的影响:最大点击新闻数和最大热点新闻数。

首先探究最大点击新闻数对模型性能的影响。如图4(a)所示,当最大点击新闻数从1到4时,模型的性

能几乎没有提高,这可能是因为最大新闻点击数太小,无法提供充足的有用信息。接着,随着最大点击新闻数量的增大(从4到9),AUC会随之提升。然而,当点击新闻量很大时(从10到20),AUC值会波动并不再有明显的上升趋势,这可能是因为引入太多的点击新闻会产生信息冗余,对提高模型性能不再有帮助。鉴于以上实验结果,本文选取15为最大点击新闻数的最优值。

再来探究最大热点新闻数对模型性能的影响。为消除最大热点新闻数这一变量的影响,本实验中固定最大热点新闻数为15;同时由于GPU内存的限制,本实验探究最大热点新闻数为0~10时的情况。如图4(b)所示,当最大热点新闻数从0增加到1时,AUC值有显著提升,其可能的原因为与没有热点新闻的情况相比,热点新闻的引入创造性地发现了候选新闻的潜在热度,扩展了用户兴趣。接下来,随着热点新闻数量的增加(从1到3),模型的性能不断提高,这可能是由于引入多个热点新闻比引入一个热点新闻带来了更多的有用信息。但是,随着最大热点新闻数的越来越多(大于4),AUC值不再有明显提升趋势,这可能是由于热点新闻是按照热门程度由高到低添加到训练集中,顺序越往后的热点新闻所含信息量越少。鉴于以上实验结果,本文选取3为最大热点新闻数的最优值。

5.3.6 案例研究

为更加直观地展示热点特征的有效性,本文进行了一个案例研究。首先提取了某用户的相关样本数据,包含一部分点击新闻、热点新闻和候选新闻,如表7所示,其中第1条至第3条点击新闻和就业相关,第4条与第5条点击新闻和活动相关(这些新闻类别是人工标注的,并不包含在原始数据集中,为了更方便分析该案例)。在热点新闻中,排名第1至3位的分别是与人物榜样、活动和政治相关的新闻。将这些样本分别输入进已经训练好的不带热点特征模块的HAN和带有热点特征模块的HAN,并分别得到候选新闻的预测概率。

候选新闻的预测结果如表8所示。在不含有热点特征模块的HAN中,第1条候选新闻获得了最高的预测点击概率,这是由于它与第1至3条点击新闻有同样的“学院”和“就业”等相关词。然而,第3条候选新闻由于和所有点击新闻并不相似,因此获得了较低的预测概率。相反,在含有热点特征模块的HAN中,第3条候选新闻获得了较高的预测点击概率,这是由于它与第1条热点新闻共享“学生”和“抗洪救灾”等相关词,这意味着第3条候选新闻具有一定的潜在热点性。因此,热点特征的引入提高了该候选新闻的预测点击概率,将“人物榜样”带入用户视野。综上所述,热点特征影响了最终的预测结果:含热点特征的HAN准确预测了所有候选新闻,而不含热点特征模块的HAN在第3个候选

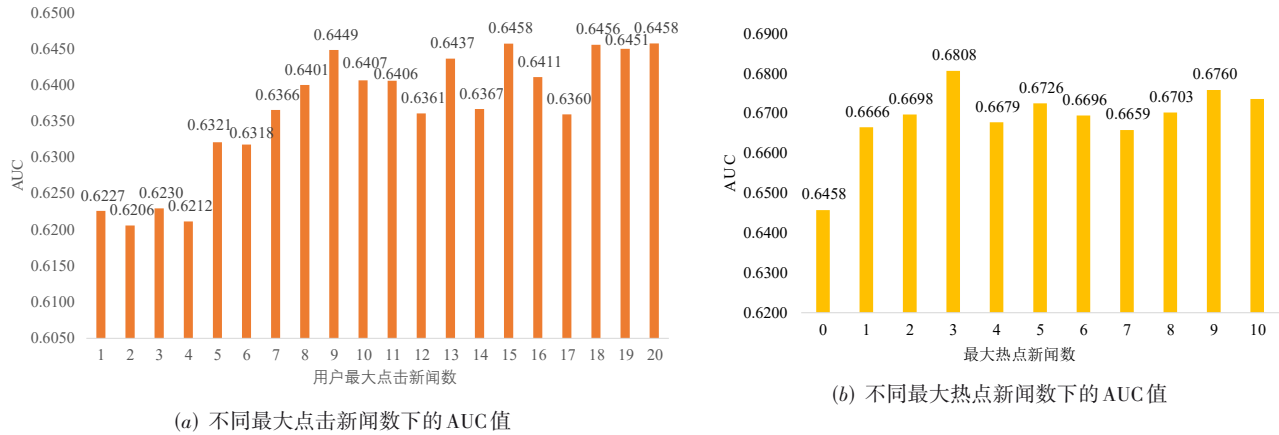


图4 最大点击新闻数和最大热点新闻数的超参数敏感性

表7 某用户点击新闻与热点新闻说明

	序号	日期	新闻标题	类别
点击新闻	1	2019/06/17	林学院新闻一个就业实习实践基地	就业
	2	2019/07/12	信息学院新建凌波创元就业实习实践基地	就业
	3	2019/09/25	材料学院依托实习实践基地助力学生就业成长	就业
	4	2019/09/29	2019年世界避孕日主题宣传系列活动在我校开展	活动
	5	2019/09/30	材料学院组织师生开展“我和国旗同框”活动	活动
	6	2019/09/01	迎新:你好,新绿!你好,明天!	其它
	7	2019/09/24	打造智慧型录播教室 助力精品在线开放课程建设	其它
热点新闻	1	2019/09/27	林学院学生参加抗洪救灾获当地好评	人物榜样
	2	2019/09/02	2019级新生开学典礼举行 师生同唱我和我的祖国	活动
	3	2019/09/19	“不忘初心,牢记使命”主题教育工作全覆盖	政治

表8 该用户候选新闻及其预测结果说明

	序号	日期	新闻标题	类别	标签	预测点击概率	
						不含热点时	含热点时
候候选新闻	1	2019/10/29	林学院召开2020届毕业生就业动员会	就业	1	0.76	0.81
	2	2019/10/18	材料学院开展八步沙“六老汉”主题分享活动	活动	1	0.59	0.63
	3	2019/10/09	林学院发布《全体师生向抗洪救灾学生学习》的倡议	人物榜样	1	0.39	0.86
	4	2019/10/24	我校学生在北京高校大学生优秀创业团队评选中喜获佳绩	其它	0	0.46	0.43

新闻上预测失败。

6 结论

本文提出了一种新颖的基于注意力增强的热点感知新闻推荐模型HAN,充分利用注意力网络和自注意力网络等深度神经网络的优势,提高热点特征提取和新闻特征提取能力,并将个性化兴趣与新闻热点性在算法中更好地平衡与利用,进一步提高推荐性能。HAN模型的贡献总结如下:(1)将注意力网络应用于热点特

征提取中,根据候选新闻动态聚合热点新闻,增强热点特征的表达力;(2)在新闻编码器中采用一种基于CNN的自注意力网络,捕获新闻标题中不连续的特征,弥补CNN对新闻建模的局限性(即只能从新闻中提取局部连续的特征);(3)提出一个新颖的点击预测器来灵活融合热点特征与用户兴趣,以提升候选新闻的点击预测准确率。在实际数据集上的实验表明,该模型在AUC和F1两项指标上有明显提升,且有助于缓解用户的冷启动问题。

参考文献

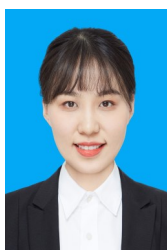
- [1] PRAWESH S, PADMANABHAN B. The “top N” news recommender: Count distortion and manipulation resistance [C]//Proceedings of the Fifth ACM Conference on Recommender Systems. New York: ACM, 2011: 237-244.
- [2] GULLA J A, YU B, ÖZGÖBEK Ö, et al. 3rd international workshop on news recommendation and analytics (INRA 2015)[C]//Proceedings of the 9th ACM Conference on Recommender Systems. New York: ACM, 2015: 345-346.
- [3] CHEN C, MENG X W, XU Z H, et al. Location-aware personalized news recommendation with deep semantic analysis[J]. IEEE Access, 2017, 5: 1624-1638.
- [4] 田莹, 丁琪, 廖子慧, 等. 基于深度学习的新闻推荐算法研究综述[J]. 计算机科学与探索, 2021, 15(6): 971-998.
TIAN X, DING Q, LIAO Z H, et al. Survey on deep learning based news recommendation algorithm[J]. Journal of Frontiers of Computer Science and Technology, 2021, 15(6): 971-998. (in Chinese)
- [5] 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. New York: ACM, 2019: 2576-2584.
- [6] WU C H, WU F Z, AN M X, et al. Neural news recommendation with attentive multi-view learning[EB/OL]. (2019-07-12). <https://arxiv.org/abs/1907.05576>.
- [7] WANG X K, SHENG Y Q, DENG H J. Joint deep network with auxiliary semantic learning for popular recommendation[J]. IEEE Access, 8: 41254-41261.
- [8] KARIMI M, JANNACH D, JUGOVAC M. News recommender systems-Survey and roads ahead[J]. Information Processing & Management, 2018, 54(6): 1203-1227.
- [9] LIU D R, OMAR H, LIOU C H, et al. Recommending blog articles based on popular event trend analysis[J]. Information Sciences, 2015, 305: 302-319.
- [10] JONNALAGEDDA N, GAUCH S. Personalized news recommendation using twitter[C]//2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT). Piscataway: IEEE, 2013: 21-25.
- [11] NATARAJAN S, MOH M. Recommending news based on hybrid user profile, popularity, trends, and location [C]//2016 International Conference on Collaboration Technologies and Systems (CTS). Piscataway: IEEE, 2016: 204-211.
- [12] ZHANG S, YAO L N, SUN A X, et al. Deep learning based recommender system[J]. ACM Computing Surveys, 2020, 52(1): 1-38.
- [13] 黄立威, 江碧涛, 吕守业, 等. 基于深度学习的推荐系统研究综述[J]. 计算机学报, 2018, 41(7): 1619-1647.
HUANG L W, JIANG B T, LV S Y, et al. Survey on deep learning based recommender systems[J]. Chinese Journal of Computers, 2018, 41(7): 1619-1647. (in Chinese)
- [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. New York: ACM, 2017: 1933-1942.
- [15] GAO J, XIN X, LIU J S, et al. Fine-grained deep knowledge-aware network for news recommendation with self-attention[C]//2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI). Piscataway: IEEE, 2018: 81-88.
- [16] 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). Stroudsburg: Association for Computational Linguistics, 2019: 6390-6395.
- [17] 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: 5973-5980.
- [18] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms [C]//WWW'01: Proceedings of the 10th International Conference on World Wide Web. New York: ACM, 2001: 285-295.
- [19] HE X N, LIAO L Z, ZHANG H W, et al. Neural collaborative filtering[C]//Proceedings of the 26th International Conference on World Wide Web. Geneva: International World Wide Web Conferences Steering Committee, 2017: 173-182.
- [20] WANG H W, ZHANG F Z, XIE X, et al. DKN: Deep knowledge-aware network for news recommendation[C]//WWW'18: Proceedings of the 2018 World Wide Web Conference. Geneva: International World Wide Web Conferences Steering Committee, 2018: 1835-1844.
- [21] 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. Stroudsburg: Association for Computational Linguistics, 2019: 336 – 345.
- [22] JONNALAGEDDA N, GAUCH S, LABILLE K, et al. Incorporating popularity in a personalized news recommender system[J]. PeerJ Computer Science, 2016, 2: e63.
- [23] TIWARI S, PANGTEY M S, KUMAR S. Location aware personalized news recommender system based on Twitter popularity[C]//International Conference on Computational Science and Its Applications. Cham: Springer, 2018: 650-658.
- [24] LIU T Y, ZHONG Y, CHEN K. Interdisciplinary study on popularity prediction of social classified hot online events in China[J]. Telematics and Informatics, 2017, 34 (3): 755-764.
- [25] KHOSLA A, SARMA A D, HAMID R. What makes an image popular? [C]//WWW' 14: Proceedings of the 23rd International Conference on World Wide Web. New York: ACM, 2014: 867-876.
- [26] REN H, YANG Q. Predicting and evaluating the popularity of online news[EB/OL]. (2021-07-01). http://cs229.stanford.edu/proj2015/328_report.pdf.
- [27] 李博. 基于项目特征和排序学习的新闻推荐系统设计与实现[D]. 北京: 北京邮电大学, 2019.
- LI B. The Design and Implementation of News Recommendation System Based on Item Feature and Learning to Rank[D]. Beijing: Beijing University of Posts and Telecommunications, 2019. (in Chinese)
- [28] CHATZOPOULOU G, SHENG C, FALOUTSOS M. A first step towards understanding popularity in YouTube [C]//2010 INFOCOM IEEE Conference on Computer Communications Workshops. Piscataway: IEEE, 2010: 1-6.
- [29] YOON K. Convolutional neural networks for sentence classification[EB/OL]. (2014-08-25). <https://arxiv.org/abs/1408.5882>.
- [30] LIN Z H, FENG M W, SANTOS C N D, et al. A structured self-attentive sentence embedding[EB/OL]. (2017-03-09). <https://arxiv.org/abs/1703.03130>.
- [31] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//NIPS' 13: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. Red Hook: Curran Associates Inc, 2013: 3111-3119.
- [32] REHUREK R, SOJKA P. Software framework for topic

modelling with large corpora[C]//Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks. Marrakech: European Language Resources Association, 2010: 46-50.

- [33] KINGMA D P, BA J. Adam: A method for stochastic optimization[EB/OL]. (2014-12-22). <https://arxiv.org/abs/1412.6980>.

作者简介



丁 琪 女, 1996年生, 山东邹城人. 北京林业大学信息学院硕士研究生. 主要研究方向为智能信息处理、个性化推荐等.
E-mail: 947515733@qq.com



田 萱(通讯作者) 女, 1976年生, 山东济宁人. 2008年于中国人民大学获得博士学位, 现为北京林业大学副教授, CCF高级会员. 主要研究方向为智能信息处理、文本挖掘等.
E-mail: tianxuan@bjfu.edu.cn



孙国栋 男, 1981年生, 黑龙江哈尔滨人. 2009年与哈尔滨工业大学获得博士学位, 现为北京林业大学副教授. 主要研究方向为无线传感器网络、移动计算、数据挖掘等.
E-mail: sungd@bjfu.edu.cn