



## 面向微博用户的个性化推荐算法研究

周炜翔<sup>1</sup>, 张 雯<sup>1</sup>, 杨 博<sup>2</sup>, 柳 毅<sup>2</sup>, 张 琳<sup>2</sup>, 张仰森<sup>1</sup>

(1. 北京信息科技大学 智能信息处理研究所, 北京 100101; 2. 国家计算机网络应急技术处理协调中心, 北京 100029)

**摘 要:** 微博的个性化推荐对于提升用户体验和帮助用户及时、准确地获取信息具有重要意义。在分析微博用户行为模式的基础上, 提出一种基于情景建模和卷积神经网络的微博个性化推荐模型。从时间和地域两个维度对用户进行情景建模, 提取用户的时间情景模式和地域情景模式, 同时给出情景模式相似度计算方法, 对用户的情景模式进行扩展, 捕捉用户感兴趣的情景模式倾向, 在此基础上建立用户个性化情景模式库, 采用卷积神经网络构建个性化微博推荐模型, 实现微博用户的个性化推荐。实验结果表明, 与 ILCAUSR、RA-CD 算法相比, 该模型具有较好的推荐效果, 相比于时间情景模型和地域情景模型, 其平均绝对误差和平均用户满意度指标均达到最优效果。

**关键词:** 个性化推荐; 情景建模; 卷积神经网络; 情景模式库; 用户满意度

开放科学(资源服务)标志码(OSID):



中文引用格式: 周炜翔, 张雯, 杨博, 等. 面向微博用户的个性化推荐算法研究[J]. 计算机工程, 2020, 46(10): 60-66, 73.

英文引用格式: ZHOU Weixiang, ZHANG Wen, YANG Bo, et al. Research on personalized recommendation algorithm for microblog users[J]. Computer Engineering, 2020, 46(10): 60-66, 73.

## Research on Personalized Recommendation Algorithm for Microblog Users

ZHOU Weixiang<sup>1</sup>, ZHANG Wen<sup>1</sup>, YANG Bo<sup>2</sup>, LIU Yi<sup>2</sup>, ZHANG Lin<sup>2</sup>, ZHANG Yangsen<sup>1</sup>

(1. Institute of Intelligent Information Processing, Beijing Information Science and Technology University, Beijing 100101, China;

2. National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China)

**[Abstract]** Personalized recommendation of microblog is crucial to improving user experience and helping users obtain information accurately in time. Based on the analysis of behavior patterns of microblog users, this paper proposes a personalized recommendation model for the microblog based on scenario modeling and Convolutional Neural Network (CNN). Scenario modeling is implemented for users from the dimensions of time and region, so as to extract the user's temporal scenario pattern and geographical scenario pattern. Then a calculation method of scenario pattern similarity is provided to extend the scenario patterns of users, capturing the scenario pattern tendency that users are interested in. On this basis, a personalized scenario mode library of the user is established, and the CNN is used to construct a personalized recommendation model for microblog users. Experimental results on real data of the microblog show that compared with the ILCAUSR and RA-CD algorithms, the proposed model has better recommendation performance, and achieves the optimal effect in Mean Absolute Error (MAE) and Average User Satisfaction (AUS) indexes compared with the temporal scenario model and geographical scenario model.

**[Key words]** personalized recommendation; scenario modeling; Convolutional Neural Network (CNN); scenario pattern library; user satisfaction

DOI: 10.19678/j.issn.1000-3428.0055979

### 0 概述

微博作为新兴的互联网社交平台, 其以实时性、开放性、互动性和便捷性为人们进行意见表达

和信息交流提供了良好的媒介, 已超越传统媒体成为新的信息聚集地, 并以极快的速度影响着社会的信息传播格局<sup>[1]</sup>。目前, 人们通过微博获取信息的方式主要是通过关注的好友发布的微博信息、通过

**基金项目:** 国家自然科学基金“网络社交媒体中特定社会安全事件的侦测分析与态势评估研究”(61772081)。

**作者简介:** 周炜翔(1993—), 男, 硕士研究生, 主研方向为自然语言处理; 张 雯, 硕士研究生; 杨 博、柳 毅、张 琳, 博士; 张仰森(通信作者), 教授、博士。

收稿日期: 2019-09-11

修回日期: 2019-10-29

E-mail: zhangyangsen@163.com

微博平台的“热门话题”推荐的相关热点微博和通过微博的检索功能检索包含特定关键词的微博。以上信息获取方式都是面向所有微博用户进行推荐的,缺乏一种针对特定用户的个性化推荐功能。同时,由于微博数量巨大,为用户及时、有效地获取自己感兴趣的微博内容带来了极大的困难<sup>[2]</sup>。因此,针对微博的个性化信息服务技术得到了国内外学者的广泛关注,成为当前社交媒体领域研究的热点。

为提升微博个性化推荐的服务效果,本文提出一种基于情景建模和卷积神经网络(Convolutional Neural Network, CNN)的微博个性化推荐模型,在分析微博用户行为模式的基础上,从时间和地域两个维度对用户进行情景建模,构建用户个性化情景模式库,采用卷积神经网络实现微博用户的个性化推荐。

## 1 相关研究

个性化推荐通过采集系统中用户和物品的信息,采用一系列的计算模型对用户的信息选择和决策提供支持。目前的个性化推荐算法主要分为基于协同过滤(Collaborative Filtering, CF)的推荐算法和基于内容的推荐算法。协同过滤的推荐算法<sup>[3]</sup>是最早的推荐模型,主要是从历史数据(如用户以前对物品的评分)中发现用户和物品的联系,构成评分矩阵,通过预测用户对未知物品的评分来进行个性化推荐。文献[4]在协同过滤的推荐算法的基础上提出一种基于概率的共识模型(Consensus Model, COM),通过研究群体活动的生成过程,根据群体中每个成员的行为特征,构建基于组的推荐算法。为解决协同过滤算法中的数据稀疏问题,文献[5]提出一种协同知识库嵌入(Collaborative Knowledge Base Embedding, CKE)的集成框架,采用堆叠的去噪自动编码器和卷积自动编码器提取物品的文本表示和视觉表示,并在两个不同实际运用情景中的实际数据集上验证了算法的适用性。

为解决协同过滤的推荐算法中推荐个性化不足的问题、冷启动的问题以及相似用户群的局限性问题,研究人员提出了基于内容的推荐模型。基于内容的推荐模型<sup>[6-7]</sup>具有较强的可解释性,每个用户的推荐结果都是由其先前的行为决定的,在推荐结果个性化方面具有相当的优势;同时,在形成推荐结果时,直接比较候选推荐对象与用户兴趣模型的相似性,不存在冷启动的问题。但基于内容的推荐方法

也存在着推荐多样性不足与用户兴趣随时间变化的不足<sup>[8]</sup>。

近年来学者们着力探索更有效的推荐技术,并对传统推荐算法进行了融合、改进。如文献[9]把协同过滤算法与LDA主题模型相结合,构建了LDA-MF和LDA\_CF的混合协调过滤方法;同样,文献[10]将协同过滤算法和内容过滤算法相融合,提出一种融合协同过滤和内容过滤的混合推荐方法。

同时,推荐算法并不是独立存在的,需要根据各种平台的特性进行优化,学者们不断探索社交平台数据的特殊性,并研究适合微博数据的推荐算法。如文献[11]通过研究真实的微博数据,分析微博信息和社区信息对推荐结果的影响,验证了社区信息对个性化推荐的重要性,文献[12]通过修改传统协同过滤算法中的各阶段参数,在算法中加入社区信息的影响,验证了利用社区信息相似度修订的SNCF-RM有更好的推荐效率,文献[13]根据标签关联和用户社交关系进行建模,用于识别用户的兴趣,文献[14]设计了基于概率模型的协同过滤算法,分析tweet的文本内容和用户之间的交互关系,用于为用户推荐感兴趣的用户和微博。这些研究在推荐效率上都取得了一定的成功,但是由于微博环境的复杂性和微博数据的独特性,其推荐效果性能还有很大的提升空间。

微博平台是一个错综复杂的社会环境,其信息的产生与交换都存在特定的情景模式,情景模式的有效捕捉,对于提升微博个性化推荐具有重要的意义<sup>[15]</sup>。为此,本文构建了一种基于情景建模和卷积神经网络的微博个性化推荐模型(Scene Modeling and Convolutional Neural Network Mode SM-CNN),其模型结构如图1所示。具体来说,本文从时间维度和地域维度两个方面对用户发布、评论、转发、点赞的微博文本进行情景建模,提取微博用户所关注的特定情景模式,然后采用微博语料库对用户的情景模式进行扩展形成微博用户的个性化情景模式库,在个性化情景模式库的基础上,采用卷积神经网络<sup>[16]</sup>构建用户个性化推荐模型,实现对微博系统中的热点微博进行个性化推荐。微博文本的情景建模改善了用户的数据稀疏问题,基于卷积神经网络的推荐算法有效地提升了推荐的性能。

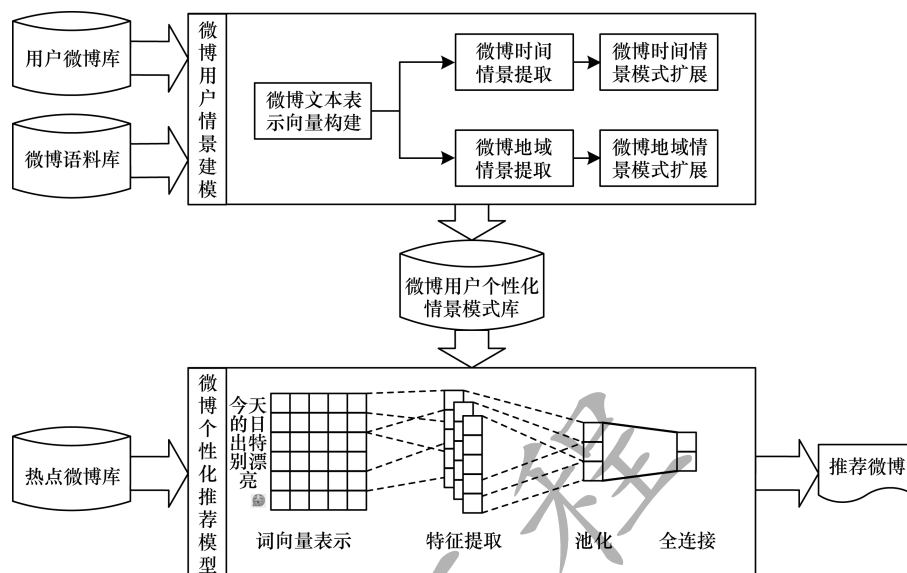


图 1 SM-CNN 模型结构

Fig. 1 SM-CNN model structure

## 2 微博用户的情景建模

在微博系统中,用户的行为偏好与特定的情景模式密不可分,微博用户的偏好随着时间变化而发生变化,在同一天内的不同时间段,其关注的重点是不同的,如用户在早晚的闲暇时浏览微博,则比较着重于实时新闻型的微博或娱乐性微博;而在上班时浏览微博,更倾向于浏览技术类的微博。同时,微博用户的偏好也随着地点的改变而变化,如用户在常住地点比较倾向于关注与本地区日常生活相关的微博,而在一些旅游地点则倾向于关注当地的风土人情、旅游特色之类的微博。因此,本文主要从时间和地域两个维度对微博用户常常关注的情景模式进行建模,通过微博语料库构建用户个性化的模式库,实现微博用户感兴趣情景模式倾向的发现。

### 2.1 情景模式的提取

情景模式获取主要包括以下 3 种方式:1) 显式获取,即通过直接接触相关人士和其他情境信息源,直接问问题或者以引导性的方式显式获取这些信息;2) 隐式获取,即隐式地从数据或环境中获得,例如从微博平台上获得用户的位置变动信息,或可以从一个事务的时间戳隐式地获得时间的情境信息;3) 推断获取,通过统计和数据挖掘方法推断出情景信息。针对微博平台的特点,本文采用以上 3 种方式相结合的方法进行情景模式的提取,具体步骤如下:

**步骤 1** 对微博文本进行分词预处理,若微博文本中存在时间词和地点词,则进行提取。

**步骤 2** 提取微博的主题词及其权重。

**步骤 3** 根据微博文本中的时间词和地点词进行情景模式的提取。

**步骤 4** 提取微博发布的时间和地点,根据微博发布的时间和地点进行情景模式的提取。

#### 2.1.1 微博主题词及其权重的提取

根据主题相关词在微博文本中出现的频次和是否出现在微博的标签中,本文对传统的 TF-IDF 算法进行改进,提出了基于话题标签信息熵的 TIE 算法,用于提取微博的主题词及其权重,其计算方法如式(1)所示:

$$\text{TIE}(w_{ij}) = \frac{\text{TI}(w_{ij})}{\text{TagE}(w_i)} \quad (1)$$

其中,  $\text{TIE}(w_{ij})$  表示词  $w_i$  在微博  $j$  中的 TIE 值,  $\text{TI}(w_{ij})$  表示词语  $w_i$  在微博  $j$  中的 TF-IDF 值,其计算方法如式(2)所示,  $\text{TagE}(w_i)$  表示词  $w_i$  的标签信息熵,其计算方法如式(3)所示:

$$\text{TI}(w_{ij}) = \frac{|w_{ij}|}{\sum_k |w_{kj}|} \cdot \log_a \frac{|D|}{|\{d_j: w_i \in d_j\}|} \quad (2)$$

$$\text{TagE}(w_i) = - \frac{|T(w_i)|}{|\{d_j: w_i \in d_j\}|} \cdot \log_a \frac{|T(w_i)|}{|\{d_j: w_i \in d_j\}|} \quad (3)$$

其中,  $|w_{ij}|$  表示词语  $w_i$  在微博  $j$  内出现的频次,  $\sum_k |w_{kj}|$  表示微博  $j$  中所有词数目之和,  $|D|$  表示微博语料库中所有微博数目,  $|\{d_j: w_i \in d_j\}|$  表示微博语料库中出现词  $w_i$  的微博数目,  $|T(w_i)|$  表示微博

语料库中词  $w_i$  出现在话题标签中的微博数目。

通过计算微博文本中每个词语的 TIE 值,获取微博文本的主题词语及其对应的权重,按照权重的大小,将每条微博表示为主题词及其权重的集合,即  $\text{WeiBo}_i = \{t_1:w_1, t_2:w_2, \dots, t_n:w_n\}$ , 其中,  $t_i$  为主题词,  $w_i$  为其对应的权重。

### 2.1.2 时间情景的提取

按照微博文本中的时间词及微博发布的时间,将时间按照每 3 h 一个时间段进行划分,取值分别为  $\{0, 3, 6, 9, 12, 15, 18, 21\}$ , 构建时间情景提取模型如式(4)所示:

$$\text{TimeScene}_{\text{WeiBo}_i} = \langle \text{WeiBo}_i, \text{time} \rangle \quad (4)$$

其中,  $\text{WeiBo}_i$  为微博的主题词表示,  $\text{time}$  为微博发布的时间值。

### 2.1.3 地域情景的提取

按照微博文本中的地点词及微博发布的地点,将地点按照省份进行划分,其取值为省份的名称,构建地域情景提取模型如式(5)所示:

$$\text{LocationScene}_{\text{WeiBo}_i} = \langle \text{WeiBo}_i, \text{location} \rangle \quad (5)$$

其中,  $\text{WeiBo}_i$  为微博的主题词表示,  $\text{location}$  为微博发布的省份名称。

## 2.2 情景模式的泛化

从微博中提取的用户情景模式一般都过于具体,缺乏一定的代表性,因此,本文在提取情景模式的基础上,制定了一定的泛化规则,从时间、地点、人物 3 个维度,将现有的情境模式泛化为一般的情境模式,实现对用户情景模型的泛化。部分泛化规则如下:

人物:男/女→任意人

地点:地铁/公交→交通工具→任意地点

家→区→市→省→任意地点

办公室→公司→任意地点

时间:周一→工作日→任意时间

那么情境  $s = (\text{男士}, \text{地铁}, \text{周一})$  可以泛化为:  
 $s_1 = (\text{任意人}, \text{地铁}, \text{工作日}); s_2 = (\text{男士}, \text{任意地点}, \text{工作日}); s_3 = (\text{男士}, \text{地铁}, \text{任意时间})$ 。

### 2.3 用户个性化情景模式库的构建

用户个性化情景模式库中包含的微博是与用户相关的微博,反映的是用户感兴趣的情景模式倾向,但是用户发布、评论、转发以及点赞的微博数目相较于微博库中的微博只是占到了很少的比例,如果仅仅采用这部分微博构建用户个性化情景模式库,作为用户个性化推荐的初始数据会存在严重的数据稀疏现象。为进一步获取更多用户感兴趣的微博数

据,需要在现有情景模式的基础上,借助于微博语料库,提取更多用户感兴趣的微博。用户个性化情景模式库构建的具体步骤如下:

**步骤 1** 爬取特定用户发布、评论、转发、点赞的微博,构建用户微博语料库。

**步骤 2** 以特定的数目的微博用户为起点,爬取用户及其关注者和被关注者发布、评论、转发、点赞的微博,构建微博语料库。

**步骤 3** 提取用户微博库和微博语料库中每一条微博的时间情景模式和地域情景模式。

**步骤 4** 按照情景模式的值对微博语料库和用户微博库中的微博进行划分。

**步骤 5** 计算对应情景模式值下的微博语料库中的微博与用户微博库中每一条微博的相似度值,如果最大的相似度大于某一阈值  $\alpha$ , 则将这条微博加入用户个性化情景模式库。相似度计算方法如下:

微博语料库中微博  $W_c$  和用户微博库中微博  $W_u$  的情景模式分别表示为:

$$W_c = \{t_{c1}:w_{c1}, t_{c2}:w_{c2}, \dots, t_{cn}:w_{cn}; s_c\}$$

$$W_u = \{t_{u1}:w_{u1}, t_{u2}:w_{u2}, \dots, t_{un}:w_{un}; s_u\}$$

其中,  $t$  为微博中对应的主题词,  $w$  为主题词的权值,  $s$  为对应情景模式的值, 则  $W_c$  和  $W_u$  的相似度计算方法如下:

首先,任取  $i \in \{1, 2, \dots, n\}$  计算词语  $t_{ci}$  和  $t_{ui}$  间的语义相似度,其语义相似度的计算采用 Word2Vec<sup>[17]</sup> 模型进行计算,其模型的构建将在下文实验部分进行介绍。然后,根据语义相似度的值将  $W_c$  和  $W_u$  的主题词分为  $n$  组,每组为  $W_c$  和  $W_u$  中主题词语义最相近的词。最后,采用式(6)计算每组主题词权重的加权平均和,其结果即为  $W_c$  和  $W_u$  的相似度值。

$$\text{Sim}(W_1, W_2) = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{k=1}^n w_{ck} w_{uk}}{\sqrt{\sum_{k=1}^n w_{ck}^2 \sum_{k=1}^n w_{uk}^2}} \quad (6)$$

## 3 微博个性化推荐模型

在用户个性化微博库的基础上,本文引入卷积神经网络模型,采用情感分类的思想,构建微博个性化推荐模型。具体来说,本文将用户个性化微博库中的微博作为分类模型中的正例,从微博语料库中随机选取与正例数目等量的微博作为负例,组成用户微博个性化推荐模型的训练数据,通过模型的训

练,学习用户感兴趣微博的情景模式倾向,构建用户微博个性化推荐模型。模型参考了文献[18]的情感分类模型,主要采用多通道的卷积神经进行特征提取,运用池化层进行特征采样,并通过全连接层和 Softmax 层进行语义分类。具体的结构如图 1 的微博个性化推荐模型部分所示。

在卷积神经网络中,本文采用多个通道的  $h \times k$  滤波器对输入词向量矩阵进行卷积操作,获取对应窗口内的局部特征以及输入微博文本的特征图,其计算如式(7)所示:

$$c_i = f(w \cdot X_{i:i+h-1} + b) \quad (7)$$

其中,  $c_i$  代表提取的特征图中第  $i$  个特征值,  $f$  为卷积核函数,  $w \in \mathbb{R}^{h \times k}$  为滤波器,  $h$  为滑动窗口大小,  $k$  为词向量的大小,  $b$  为偏置值,  $X_{i:i+h-1}$  表示输入矩阵的第  $i$  行到第  $i+h-1$  行组成的局部特征矩阵。综合各个滑动窗口提取的特征即可得到输入的整体特征图  $C$ , 如式(8)所示:

$$C = [c_1, c_2, \dots, c_{n-h+1}] \quad (8)$$

在池化层采用最大池化的策略,获取特征图中最重要的特征作为当前卷积操作的特征输出,如式(9)所示:

$$\hat{c} = \max(C) \quad (9)$$

由于本文采用的是多通道的卷积神经网络,将提取多种类型的特征,通过对各种类型的特征进行组合,输入到全连接层进行特征融合。在融合特征的基础上,通过 Softmax 输出分类结果,输出的结果中的正例将作为用户个性化推荐的备选微博。

得到备选微博后,采用第 2 节的方法,提取每条微博的时间情景模型和地域情景模型,采用第 2.3 节用户个性化微博库的构建方法,选取 TopN 的微博作为用户个性化推荐的微博。

## 4 实验结果与分析

### 4.1 实验数据

为构建微博语料库,采用开源的爬虫框架 WebCollector 实现一个微博爬虫程序,爬取新浪微博中的用户信息及其相关的微博数据,总共爬取 1 261 967 名用户及其相关信息,182 672 450 条微博数据,用户包括普通用户、知名人士、网络营销号、官方认证机构等,其发布的微博数量及其粉丝数与关注数非常不平衡,用户发布微博的变化区间为[3,17 382],用户的粉丝数的变化区间为[12,12 006 518]。通过进行一系列的微博文本预处理,如繁简转换、URL 替

换、短小无意义微博的剔除等,最终构建了一个包含 104 652 972 条的微博语料库。微博语料库中的微博共包含词语 3 334 763 247 个,本文采用这些语料训练了一个 Word2Vec 模型,在训练的过程中采用 Skip-gram 模型,其他相关参数均采用默认设置。经过训练,最终得到了一个包含 850 599 个词的词向量,每个词的词向量的维数为 200。

### 4.2 实验结果的评测

由于目前还没有微博推荐方面的公共数据集,个性化推荐的结果因人而异,其评测指标很难直接以准确率和召回率来评价,因此本文采用平均绝对误差 (Mean Absolute Error, MAE) 和用户满意度 (Average User Satisfaction, AUS) 进行评价,参与评测的用户为邀请的志愿者,评测者在对模型毫不知情的情况下参与评测,其中 MAE 和 AUS 的计算方法如式(10)和式(11)所示:

$$MAE = 1 - \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \text{Sim}(w_i, w_{ij}) \quad (10)$$

$$AUS = \frac{1}{nm} \sum_{i=1}^m \sum_{j=1}^n \text{feedback}_{ij} \quad (11)$$

其中,  $m$  为参与评测的志愿者的数目,  $n$  为每个用户推荐的微博的数目,  $\text{Sim}(w_i, w_{ij})$  表示为用户  $i$  推荐的第  $j$  条微博与用户个性化微博库的相似度,其计算方法见第 2.3 节,  $\text{feedback}_{ij}$  为用户  $i$  对推荐的第  $j$  条微博的反馈情况,反馈结果分为 3 个等级:“不喜欢,无感,喜欢”,其取值为  $\{-1, 0, 1\}$ 。

在以上数据集和评测指标的基础上,本文设计了如下 3 个实验:1) 微博情景模式相似度阈值  $\alpha$  的确定;2) 本文推荐模型的效果评估;3) 时间情景模型和地域情景模型的对比。

### 4.3 结果分析

#### 4.3.1 微博情景模式相似度阈值的确定

在用户微博个性化微博库构建过程中,通过情景模式的相似度计算,对用户感兴趣的微博进行了扩展,如果采用所有扩展的微博进行相似度阈值的确定,会给志愿者带来大量的反馈工作量,因此,本文仅从扩展的微博中随机挑选 50 条让用户进行评估。在评估过程中采用 AUS 指标对用户的反馈情况进行评估。一般来说,相似度阈值  $\alpha$  取值越大, AUS 的值越大,但是,过大的阈值会导致扩展的微博数目越来越小,这就失去了进行微博扩展的意义,其表现为 MAE 的值越来越小。因此,需要通过 AUS 和 MAE 的变化趋势来寻找扩展数目和相似度阈值之间的平衡点。相似度阈值  $\alpha$  与 MAE、AUS 的关系

如图 2 所示。

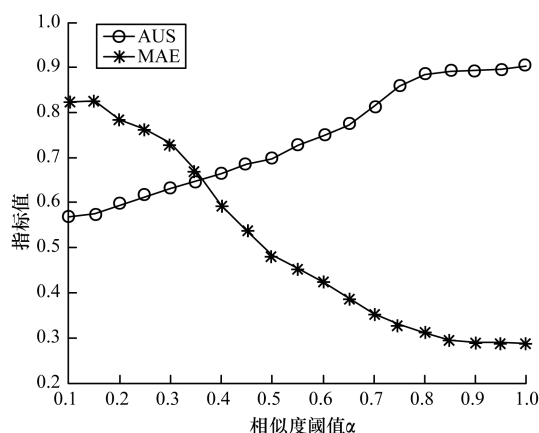


图 2  $\alpha$  与 MAE、AUS 的关系

Fig. 2 Relationship of  $\alpha$  and MAE, AUS

从图 2 可以看出,相似度阈值  $\alpha$  越小,扩增的微博平均绝对误差 (MAE) 值越大,用户满意度 (AUS) 越低。随着相似度阈值的增大,MAE 越来越小,AUS 越来越大,用户对扩增微博的满意程度逐渐增加。在  $\alpha = 0.80$  时,MAE 和 AUS 的变化趋于平缓,说明相似度大于 0.8 的微博作为用户感兴趣的微博扩增到个性化微博库的效果较好,如果再增加  $\alpha$  的取值,用户满意程度的提升有限,反而会带来扩增微博数量的减少。因此,选取 0.80 作为相似度阈值  $\alpha$  的取值。

#### 4.3.2 推荐效果的性能评估

为验证本文模型的有效性和先进性,在选定参数  $\alpha = 0.80$  时,选择融合标签关系与用户关系推荐算法 (ILCAUSR)<sup>[19]</sup>、基于社区发现的微博个性化推荐算法 (RA-CD)<sup>[20]</sup>、用户互动话题的微博推荐算法 (IBCF)<sup>[21]</sup> 与本文基于情景建模和卷积神经网络的推荐模型 (SM-CNN) 进行对比,各个对比模型中的参数均采用论文中选择的最优参数,实验结果如表 1 所示。

表 1 推荐效果对比

Table 1 Comparison of recommend effects

| 算法      | MAE     | AUS     |
|---------|---------|---------|
| ILCAUSR | 0.352 3 | 0.804 3 |
| RA-CD   | 0.332 3 | 0.856 7 |
| IBCF    | 0.324 2 | 0.845 6 |
| SM-CNN  | 0.310 6 | 0.884 2 |

从表 1 可以看出,无论在 MAE 指标还是 AUS 指标,本文的 SM-CNN 模型均达到了最优效果。在平均绝对误差 (MAE) 方面,本文提出的 SM-CNN 模型较最好的 IBCF 模型误差降低了 1.36; 在用户满

意度 (AUS) 方面,本文提出的 SM-CNN 模型较最好的 RA-CD 模型提高了 2.75%,从而证明本文提出的 SM-CNN 模型的有效性。

融合标签间关联关系与用户间社交关系的微博推荐方法 (ILCAUSR),采用标签检索策略和用户关系网构建微博用户-标签矩阵,实现多标签关联的微博推荐,但用户-标签矩阵的稀疏程度对微博推荐的准确性影响较大。基于社区发现的微博个性化推荐算法 (RA-CD),通过改进用户建模方法,融合社区发现算法,构造推荐微博效用函数,实现微博内容的个性化推荐,但该方法在用户建模时没有考虑情境因素的影响。基于用户互动话题的微博推荐算法 IBCF 通过挖掘用户好友的关系话题,实现微博个性化推荐,但该方法只考虑了话题关系,对用户社交关系、情境模式倾向等特征未做充分考虑。与以上 3 种方法相比,本文 SM-CNN 模型的优越性主要体现在以下 2 个方面:

1) 时间情景模型和地域情景模型的提取对于获取用户的兴趣倾向带来了很大的帮助。通过情景建模构建的用户个性化情景模式库充分涵盖了用户感兴趣微博的情景倾向。

2) 基于卷积神经网络的分类模型的引入为推荐模型的性能提升带来了较大的贡献。卷积神经网络的分类模型研究已经比较充分,技术相对比较成熟,对推荐性能的提升帮助很大。

#### 4.3.3 时间情景模型和地域情景模型的对比

为进一步研究时间情景模型和地域情景模型对推荐效果的影响,本文设置了 3 组对照实验,分别为基于时间情景模型的推荐、基于地域情景的推荐、基于时间情景和地域情景相结合的推荐,实验结果如图 3 所示。

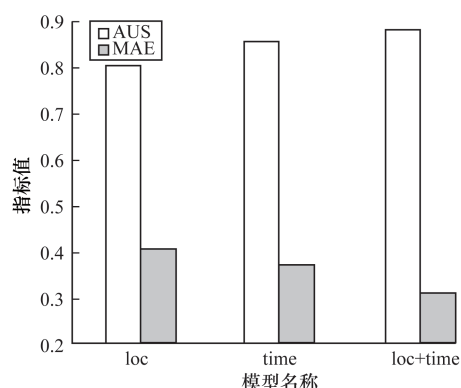


图 3 时间情景模型和地域情景模型的实验结果对比

Fig. 3 Comparison of experimental results of time scenario model and regional scene scenario model



从图 3 可以看出,结合时间情景和地域情景的推荐效果最佳,同时基于时间情景模型的效果要好于基于地域情景模型。本文认为原因主要有以下 2 个方面:1)用户在发布微博时,对于地点定位的标注不是很全面和具体,只是对其中的一部分微博进行了定位标注,很大一部分微博都缺乏具体的地域信息,使得很多微博都无法提取其地域情景模式;2)用户活动的地域一般比较固定,很少有人会频繁地更换地理位置,使得其地域情景模式中的地域值比较单一。以上 2 点为用户地域情景的提取带来了一定的障碍,也导致了基于地域情景的推荐效果不如基于时间情景的推荐效果。

## 5 结束语

本文提出一种基于情景建模和卷积神经网络的微博推荐算法,通过提取用户关注的时间情景模型和地域情景模型构建用户个性化微博库,在此基础上采用卷积神经网络对热点微博进行分类,实现用户的个性化微博推荐。情景模型的引入有助于获取用户的兴趣倾向,同时,基于卷积神经网络的分类模型提高了推荐的性能。但是,本文仅提取了时间情景模型和地域情景模型,比较单一,如何对情景模型进行细化,构建更具有代表性的情景模型,同时利用卷积神经网络在抽象特征抽取方面的优势,挖掘微博字词层面上的语义特征,将是下一步需要研究的重点。

## 参考文献

- [1] ZHONG Zhaoman, GUAN Yan, LI Cunhua, et al. Localized Top-k bursty event detection in microblog [J]. Chinese Journal of Computers, 2018, 41 (7): 1504-1516. (in Chinese)  
仲兆满, 管燕, 李存华, 等. 微博网络地域 Top-k 突发事件检测[J]. 计算机学报, 2018, 41(7): 1504-1516.
- [2] WANG Rongbing, AN Weikai, FENG Yong, et al. Important microblog user recommendation algorithm based on label and PageRank [J]. Computer Science, 2018, 45(2): 276-279. (in Chinese)  
王嵘冰, 安维凯, 冯勇, 等. 基于标签和 PageRank 的重要微博用户推荐算法[J]. 计算机科学, 2018, 45(2): 276-279.
- [3] WANG H, SHI X, YEUNG D. Y. Relational stacked denoising autoencoder for tag recommendation [C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2015: 332-346.
- [4] YUAN Q, CONG G, LIN C Y. COM: a generative model for group recommendation [C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2014: 163-172.
- [5] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems [C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA: ACM Press, 2016: 353-362.
- [6] LOPS P, DE GEMMIS M, SEMERARO G. Content-based recommender systems: state of the art and trends [M]. Berlin, Germany: Springer, 2011: 73-105.
- [7] ZAMANI H, SHAKERY A. A language model-based framework for multi-publisher content-based recommender systems [J]. Information Retrieval Journal, 2018, 21 (5): 369-409.
- [8] DENG Mingtong, LI Xuejun, LI Bin. Diversified recommendation method based on user preference and dynamic interest [J]. Journal of Chinese Computer Systems, 2018, 39(9): 2029-2034. (in Chinese)  
邓明通, 刘学军, 李斌. 基于用户偏好和动态兴趣的多样性推荐方法 [J]. 小型微型计算机系统, 2018, 39(9): 2029-2034.
- [9] HUANG Lu, LIN Chuanjie, HE Jun, et al. Diversified mobile APP recommendation combining topic model and collaborative filtering [J]. Journal of Software, 2017, 28(3): 708-720. (in Chinese)  
黄璐, 林川杰, 何军, 等. 融合主题模型和协同过滤的多样化移动应用推荐 [J]. 软件学报, 2017, 28(3): 708-720.
- [10] LIU Junjie. Research on fusion recommendation method based on content and collaborative filtering [D]. Hohhot: Inner Mongolia Normal University, 2019. (in Chinese)  
刘俊杰. 基于内容和协同过滤的融合推荐方法研究 [D]. 呼和浩特: 内蒙古师范大学, 2019.
- [11] TING I, CHANG P S, WANG S. Understanding microblog users for social recommendation based on social networks analysis [J]. Journal of Universal Computer Science, 2012, 18(4): 554-576.
- [12] CHAI Shuqin, YUAN Qian, ZHOU Peng. Personalized microblog recommendation model based on social network relations [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(5): 520-529. (in Chinese)  
蔡淑琴, 袁乾, 周鹏. 基于社会网络关系的微博个性化推荐模型 [J]. 情报学报, 2014, 33(5): 520-529.
- [13] MA Huifang, JIA Meihuizi, ZHANG Di, et al. Combining tag correlation and user social relation for microblog recommendation [J]. Information Sciences, 2017, 385: 325-337.
- [14] KIM Y, SHIM K. Twitobi: a recommendation system for twitter using probabilistic modeling [C]//Proceedings of the 11th IEEE International Conference on Data Mining. Washington D. C., USA: IEEE Press, 2011: 340-349.
- [15] WANG Zhenhuang, CHEN Siming, YUAN Xiaoru. Visual analysis for microblog topic modeling [J]. Journal of Software, 2018, 29(4): 1115-1130.  
王臻皇, 陈思明, 袁晓如. 面向微博主题的可视分析研究 [J]. 软件学报, 2018, 29(4): 1115-1130.
- [16] FENG Shi, WANG Yaqi, SONG Kaisong, et al. Detecting multiple coexisting emotions in microblogs with convolutional neural networks [J]. Cognitive Computation, 2018, 10(1): 136-155.

(下转第 73 页)

(上接第 66 页)

- [17] SONG Xiaoyu, JIN Liting, LIU Tong, et al. Research on the construction of sentiment dictionary based on Word2vec [C]//Proceedings of 2018 International Conference on Algorithms, Computing and Artificial Intelligence. New York, USA: ACM Press, 2018: 70-87.
- [18] KIM Y. Convolutional neural networks for sentence classification [C]//Proceedings of IEEE EMNLP' 14. Washington D. C., USA: IEEE Press, 2014: 1746-1751.
- [19] MA Huifang, JIA Meihuizi, ZHANG Di, et al. Microblog recommendation based on tag correlation and user social relation [J]. Acta Electronica Sinica, 2017, 45(1): 112-118. (in Chinese)  
马慧芳, 贾美惠子, 张迪, 等. 融合标签关联关系与用户社交关系的微博推荐方法 [J]. 电子学报, 2017, 45(1): 112-118.
- [20] LI Chaoyi, ZHANG Yangsen, TONG Lingling. A microblog personalized recommendation algorithm based on community discovery [J]. Microelectronics & Computer, 2017, 34(6): 40-44. (in Chinese)  
李超逸, 张仰森, 佟玲玲. 一种基于社区发现的微博个性化推荐算法 [J]. 微电子学与计算机, 2017, 34(6): 40-44.
- [21] LU Xiao, LI Peng, WANG Bin, et al. A user interaction topic based microblog recommendation algorithm [J]. Journal of Chinese Information Processing, 2016, 30(3): 187-195. (in Chinese)  
鲁晓, 李鹏, 王斌, 等. 一种基于用户互动话题的微博推荐算法 [J]. 中文信息学报, 2016, 30(3): 187-195.

编辑 索书志