

# 微博用户推荐模型的研究

徐雅斌, 石伟杰

(北京信息科技大学计算机学院 北京 朝阳区 100101; 北京信息科技大学网络文化与数字传播北京市重点实验室 北京 朝阳区 100101)

**【摘要】** 微博用户推荐对改善用户体验、促进社交网络长远发展具有重要意义。该文提取了能够充分反映微博用户之间相关性的多个特征, 并通过逻辑回归模型对潜在的用户进行评分排序, 为目标用户推荐前 $N$ 个潜在用户。基于新浪微博数据集的实验结果表明, 基于逻辑回归方法的用户推荐模型是切实有效的, 可以为微博用户提供高性能的个性化用户推荐。

**关键词** 逻辑回归; 微博; 个性化推荐; 用户推荐

**中图分类号** TP301

**文献标志码** A

doi:10.3969/j.issn.1001-0548.2015.02.016

## Research on Micro-Blog User Recommendation Model

XU Ya-bin and SHI Wei-jie

(Computer School of Beijing Information Science & Technology University Chaoyang Beijing 100101;

Beijing Key Laboratory of Internet Culture and Digital Dissemination Research, Beijing Information Science & Technology University

Chaoyang Beijing 100101)

**Abstract** Micro-blog user recommendation has great significance and value for improving the user experience and promoting the long-term development of the social network. In this paper, multiple features reflecting the correlation between micro-blog users are extracted. Combining the user features and ranks scores for potential users, top- $n$  potential users are recommended for the target user. The experimental results based on the dataset gained from Sina Micro-Blog shows that the proposed method is feasible and effective, and it can provide personalized user recommendation with high performance for micro-blog users.

**Key words** logic regression; micro-blog; personalized recommendation; user recommendation

微博作为社交网络的一种表现形式, 是新一代社交、实时网络服务的典范, 以其新颖的表现形式、独特的运营方式, 快速吸引了大量的用户。在微博平台上, 用户可以通过关注同学、朋友等组成自己的社交圈, 也可以通过关注其他用户来丰富自己的社交圈, 这就为社交网络的推荐服务提供了丰富的土壤。

用户推荐服务可以快速有效地帮助用户扩大社交范围, 因而成为社交网络的一个研究热点, 个性化推荐服务更是成为社交网络应用的发展方向。为微博用户提供个性化用户推荐服务对发展用户社交圈, 加强用户间的社交联系, 完善社交网络生态系统具有重要意义, 蕴含着巨大的商业价值。因此, 研究个性化微博用户推荐具有重要的现实意义。

## 1 相关工作

微博社交网络拥有数以亿计的用户, 其中的每

个用户都可能有自己的兴趣点和特性, 如何高效地为特定用户从海量用户中选择出其感兴趣的用户, 已成为当下社交网络研究领域的热点。

文献[1]进行的一个扩展分析证明: 在社交网络的历史数据是可以获取的情况下, 使用基于内容的方法进行用户推荐是有效的。但单纯使用基于内容的方法较为片面, 不能完整的表现用户推荐行为。文献[2]综合考虑了用户兴趣和用户社交网络方法, 并将它们合并到一个基于PageRank算法的混合模型, 通过实验证明了该混合模型的有效性。但该模型未考虑到用户个人信息、标签等相关信息, 达不到个性化用户推荐的要求。文献[3]和文献[4]展示了社交网络中两种基于标签进行用户推荐的方法, 并通过实验证明, 基于标签进行用户推荐是有效的。文献[5]采用基于内容的相似和协同过滤方法来推荐twitter用户, 取得了一定的推荐效果, 但该方法未能

收稿日期: 2014-10-22; 修回日期: 2015-02-06

基金项目: 国家自然科学基金(61370139); 网络文化与数字传播北京市重点实验室项目(ICDD201309); 北京市属高等学校创新团队建设与教师职业发展计划(IDHT20130519)

作者简介: 徐雅斌(1964-), 男, 教授, 主要从事社交网络、云计算和未来网络方面的研究。

充分考虑其他用户信息, 个性化推荐效果不理想。文献[6]使用LDA(latent dirichlet allocation)模型对twitter文本进行主题建模, 并与传统的向量空间模型方法进行比较, 实验结果发现LDA的效果优于向量空间模型。文献[7]根据关联规则算法得到用户间的共同好友关系进行Top-N用户选取, 再对选取出来的Top-N用户进行标签协同过滤, 从而得到推荐用户列表。但该方法忽略了用户内容、用户个人信息等相关要素, 仍不能满足个性化微博用户推荐的要求。

总而言之, 现有的方法在进行微博用户推荐时使用的方法单一, 性能不足; 虽然有些方法使用混合模型综合了几种不同的推荐方法, 取得了一定的推荐效果, 但采用的信息不够全面, 没有贴切地考虑与微博用户密切相关的其他信息; 同时其使用的混合模型在多参数时常需使用经验方法进行参数调节, 不能获得理想的推荐结果。针对这些情况, 本文从微博实际出发, 综合考虑了基于内容、标签、社交圈的推荐方法和用户个人信息特征, 并使用逻辑回归模型将以上几种方法有机融合到一起。逻辑回归模型通过挖掘用户信息中内在支配推荐排序的信息, 很好地解决了使用经验调节参数的问题。

## 2 微博用户的特征提取与推荐模型

### 2.1 基于内容的用户推荐(UC)

内容中往往包含用户兴趣, 而兴趣的相似性加大了用户成为好友的可能性。在微博研究领域, “内容”一般指的是用户微博, 包括用户发布、评论、转发的内容。

然而现有的基于内容的方法在进行微博用户推荐时, 只考虑了用户自身的微博, 忽略了与用户有关联的其他用户的微博。研究表明, 关注用户的微博中包含了用户的某些兴趣<sup>[8]</sup>。针对以上问题, 本文统一将用户与其关注用户的微博作为微博内容, 以改善其性能。

本文使用潜在狄利克雷主题模型<sup>[9]</sup>(LDA)获取用户微博的主题分布, 并计算用户间微博主题的相似度。LDA是一种文档主题生成模型, 是一种非监督学习技术, 其原理如图1所示。

将每个用户及其关注用户的微博作为输入文档 $d$ , 所有用户及其关注用户的微博作为文档集合 $D(d \in D)$ 。 $\alpha$ 和 $\eta$ 表示狄利克雷先验参数;  $\beta_k$ 表示词汇的主题分布;  $\theta_d$ 为文档 $d$ 的主题概率 $\theta_{dk}$ 表示文档 $d$ 中第 $k$ 个主题的主题概率;  $Z_d$ 表示文档 $d$ 的主题分配,  $Z_{d,n}$ 表示文档 $d$ 中第 $n$ 个单词的主题分配;  $W_d$ 为文档 $d$

中观察到的单词,  $W_{d,n}$ 表示文档 $d$ 中的第 $n$ 个单词。

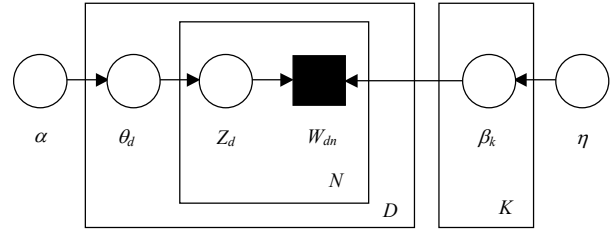


图1 LDA模型

对于文档集中的每篇文档, LDA定义了如下的生成过程: 1) 对于每一篇文档, 从主题分布中抽取一个主题; 2) 从上述被抽到的主题所对应的单词分布中抽取一个单词; 3) 重复上述过程直至遍历文档中的每一个单词。

进行用户推荐需计算用户 $u$ 及其关注用户和用户 $v$ 及其关注用户微博内容间的相似度 $\text{Sim}_{UC}(u, v)$ , 取最高的前 $K$ 个用户作为推荐对象。将用户 $u$ 及其关注用户的微博和用户 $v$ 及其关注用户的微博分别记为文档 $d_u$ 和 $d_v$ 。LDA主题个数参数统一取为20, 迭代次数参数统一取为150, 并在统一的LDA计算过程中得到主题分布向量 $T_u$ 和 $T_v$ 。本文采用余弦相似度表征主题分布向量 $T_u$ 和 $T_v$ 间的相似度, 其计算方法为:

$$\text{Sim}_{UC}(u, v) = \text{sim}(T_u, T_v) = \frac{\sum_{i=1}^n u_i \times v_i}{\sqrt{\sum_{i=1}^n u_i^2} \times \sqrt{\sum_{i=1}^n v_i^2}} \quad (1)$$

式中,  $u_i$ 和 $v_i$ 分别表示向量 $T_u$ 和 $T_v$ 的元素。

### 2.2 基于标签的用户推荐(UT)

标签是一组关键字, 是用户对自己既客观又主观的精简描述, 直接反映了用户的兴趣, 又具有个性化的特点。对于用户 $u$ 和用户 $v$ , 相同或相似的标签越多, 表明他们越相似。

因为标签是一组关键字, 故需要对用户间的标签进行同义词判断, 本文利用《知网》进行标签同义词判断。标签同义词判断需计算标签 $t_1$ 、 $t_2$ 间的相似度 $\text{sim}(t_1, t_2)$ ; 当相似度 $\text{sim}(t_1, t_2)$ 大于阈值 $\delta$ 时, 则认为标签 $t_1$ 、 $t_2$ 为同义词。本文使用文献[10]的方法计算标签间的相似度。标签关键字为实词, 一个实词概念的语义表达式分为第一基本义原描述式、其他基本义原描述式、关系义原描述式、符号义原描述式4个部分, 故两个实词间相似度是这4个义原相似度 $\text{sim}_1$ 、 $\text{sim}_2$ 、 $\text{sim}_3$ 、 $\text{sim}_4$ 的加权和, 其计算方法为:

$$\text{Sim}(t_1, t_2) = \sum_{i=1}^4 \beta \prod_{j=1}^i \text{sim}_j \quad (2)$$

计算出词语间相似度后, 便可以得到判断标签

是否为同义词的方法为:

$$\text{Sim}(t_1, t_2) = \begin{cases} 1 & \text{Sim}(t_1, t_2) \geq \delta \\ 0 & \text{Sim}(t_1, t_2) < \delta \end{cases} \quad (3)$$

使用标签特征计算用户 $u$ 与用户 $v$ 之间相似性的计算方法为:

$$\text{Sim}_{\text{UT}}(u, v) = \frac{|R_u \cap R_v|}{|R_u| + |R_v|} \quad (4)$$

式中,  $R_u$ 表示用户 $u$ 的标签集;  $R_v$ 表示用户 $v$ 的标签集。  $|R_u \cap R_v|$ 表示用户 $u$ 和用户 $v$ 相同或相似的标签数量;  $|R_u| + |R_v|$ 表示用户 $u$ 和用户 $v$ 的标签数量和。

### 2.3 基于社交圈的用户推荐(USN)

对于用户 $u$ 来说, 社交圈的成员包括用户 $u$ 的粉丝以及用户 $u$ 的关注用户。粉丝是关注了用户 $u$ 的人群, 反映了用户 $u$ 的影响力; 关注用户是用户 $u$ 关注的人群, 表征了用户 $u$ 的偏好倾向。社交网络本质上描述的是用户间的关系, 共同的粉丝或关注用户正是用户间关系的体现。因此, 对于两个微博用户来说(向目标用户 $u$ 推荐用户 $v$ ), 社交圈的相似程度也是影响推荐的一个重要因素。用户社交圈的相似性由用户粉丝和关注用户两个方面决定。

#### 2.3.1 用户粉丝(Follower)

将用户 $u$ 的粉丝集记为  $\text{Follower}(u) = \{\text{fer}_{u_1}, \text{fer}_{u_2}, \dots, \text{fer}_{u_m}\}$ , 用户 $v$ 的粉丝集记为  $\text{Follower}(v) = \{\text{fer}_{v_1}, \text{fer}_{v_2}, \dots, \text{fer}_{v_n}\}$ 。用户 $u$ 与用户 $v$ 的共同粉丝越多, 用户间的相似性越高。

**定义 1** 用户粉丝贡献度为:

$$\text{Stren\_er}(u, v) = \frac{|\text{Follower}(u) \cap \text{Follower}(v)|}{|\text{Follower}(u)| + |\text{Follower}(v)|} \quad (5)$$

式中,  $|\text{Follower}(u) \cap \text{Follower}(v)|$ 表示用户 $u$ 和用户 $v$ 共同的粉丝数;  $|\text{Follower}(u)| + |\text{Follower}(v)|$ 为用户 $u$ 和用户 $v$ 的总粉丝数。

#### 2.3.2 关注用户(Following)

将用户 $u$ 的关注用户集记为  $\text{Following}(u) = \{\text{fer}_{u_1}, \text{fer}_{u_2}, \dots, \text{fer}_{u_m}\}$ , 用户 $v$ 的关注用户集记为  $\text{Following}(v) = \{\text{fing}_{v_1}, \text{fing}_{v_2}, \dots, \text{fing}_{v_n}\}$ 。用户 $u$ 与用户 $v$ 的共同关注用户越多, 用户间越相似。

**定义 2** 关注用户贡献度为:

$$\text{Stren\_ing}(u, v) = \frac{|\text{Following}(u) \cap \text{Following}(v)|}{|\text{Following}(u)| + |\text{Following}(v)|} \quad (6)$$

式中,  $|\text{Following}(u) \cap \text{Following}(v)|$ 表示用户 $u$ 和用户 $v$ 共同的关注用户数;  $|\text{Following}(u)| + |\text{Following}(v)|$ 为用户 $u$ 和用户 $v$ 的总关注用户数。

#### 2.3.3 用户社交圈相似性( $\text{Sim}_{\text{USN}}(u, v)$ )

根据  $\text{Stren\_er}(u, v)$ 和  $\text{Stren\_ing}(u, v)$ , 可得到用户社交圈相似性 $\text{Sim}_{\text{USN}}(u, v)$ 的计算方法。

**定义 3** 用户社交圈的相似性为:

$$\text{Sim}_{\text{USN}}(u, v) = \theta * \text{Stren\_er}(u, v) + (1 - \theta) * \text{Stren\_ing}(u, v) \quad (7)$$

式中,  $\theta$ 为相似权重, 用来平衡用户粉丝和关注用户的相似度。在本文中, 作简化处理, 取 $\theta=0.5$ 。

### 2.4 基于用户个人信息特征(PI)的推荐

用户个人信息是表征用户的一组属性, 是用户最为直观反映, 可以明显区分用户间的异同。用户个人信息包括年龄、地区、性别等。现有的方法在使用用户个人信息特征时, 并未考虑到用户对推荐的影响, 实际上, 微博社交网络汇聚了大量来自不同地区的用户, 同一地区或家乡相同的用户由于地理位置上的靠近或天然存在的亲近感, 使得用户往往对与自己处于同一地区或家乡相同的用户感兴趣, 故本文选取用户地区作为用户的一个属性。对于性别, 异性用户间往往更容易互相吸引, 微博用户也是如此, 所以本文将性别也作为用户的一个属性。

#### 2.4.1 用户地区(UD)

本文建立了一个行政地域数据库, 数据库包含全国各省份名称及其所辖的地市名称。对于向目标用户 $u$ 推荐用户 $v$ , 计算用户地区相似度的步骤如下:

1) 根据数据库对用户地区进行过滤, 只保留省级或地市级地区;

2) 查找数据库, 判断用户 $u$ 的地区 $\text{UD}_u$ 与用户 $v$ 的地区 $\text{UD}_v$ 是否为同一省份。若不同, 则 $\text{sim}_{\text{UD}}(u, v)=0$ ; 若相同, 进行下一步;

3) 查询数据库用户地区是否为同一地市。若是, 则 $\text{sim}_{\text{UD}}(u, v)=1$ ; 否则,  $\text{sim}_{\text{UD}}(u, v)=0.5$ 。

#### 2.4.2 用户性别(US)

对于目标用户 $u$ , 若用户 $v$ 为异性, 用户 $w$ 为同性, 则同等条件下用户 $u$ 往往更偏向于接受用户 $v$ , 故本文将用户性别作为用户的一个特征。用户性别US可表示为:

$$\text{US} = \begin{cases} 1 & \text{用户为男性} \\ 0 & \text{用户为女性} \end{cases} \quad (8)$$

对于目标用户 $u$ 和待推荐用户 $v$ , 用户在性别维度上的相似度 $\text{sim}_{\text{US}}(u, v)$ 可表示为:

$$\text{sim}_{\text{US}}(u, v) = \begin{cases} 0.5 & \text{US}_u = \text{US}_v \\ 1 & \text{US}_u \neq \text{US}_v \end{cases} \quad (9)$$

2.4.3 用户个人信息相似度(PI)

由 $sim_{UD}(u,v)$ 和 $sim_{US}(u,v)$ , 可得到用户个人信息相似度 $Sim_{PI}(u,v)$ 。

定义 4 用户个人信息相似为:

$$Sim_{PI}(u,v) = \alpha * sim_{UD}(u,v) + (1-\alpha) * sim_{US}(u,v) \quad (10)$$

式中,  $\alpha$ 为权重因子, 由于 $sim_{UD}(u,v)$ 和 $sim_{US}(u,v)$ 对用户个人信息相似度 $Sim_{PI}(u,v)$ 都是正相关, 故本文取 $\alpha=0.5$ 。

3 基于逻辑回归的融合推荐模型

根据不同方法或特征进行用户推荐产生的排名不同, 如何将这方法或特征进行有效融合是准确和有效推荐微博用户的关键。逻辑回归(logistic regression)模型是一种多元统计方法, 和支持向量机(SVM)一样, 是一种判别学习模型, 具有良好的性能, 但逻辑回归模型的时间复杂度和空间复杂度都低于支持向量机。更重要的是, 逻辑回归模型可用于概率预测, 将用户推荐转化为对一个特征向量进行概率预测过程。通过挖掘用户信息中内在支配推荐排序的信息, 很好的解决了使用经验调节参数的问题, 可以很容易地调整模型的参数, 使模型能够适应所应用的环境。故本文采用逻辑回归模型进行微博用户推荐。

推荐过程(向用户 $u$ 推荐用户 $v$ )中, 逻辑回归模型会基于与用户 $u$ 和用户 $v$ 都相关的特征向量 $\mathbf{X}$ , 得到用户 $u$ 关注用户 $v$ 的概率 $P(u \rightarrow v | \mathbf{X})$ , 最后基于概率 $P$ 产生一个最终的推荐用户列表。逻辑回归模型为:

$$P(u \rightarrow v | \mathbf{X}) = \frac{1}{1 + e^{-\theta^T \mathbf{X}}} \quad (11)$$

式中,  $\mathbf{X}$ 表示与用户 $u$ 和用户 $v$ 都相关的特征向量, 由归一化处理后的各种用户推荐方法的推荐得分(用户间相似度)和用户特征相似度组成, 即基于内容的推荐方法的推荐得分、基于用户标签的推荐方法的推荐得分、基于用户社交圈推荐方法的推荐得分和用户个人信息特征相似度;  $\theta$ 表示特征向量权重, 可通过训练集对逻辑回归模型进行训练得到。

取用户 $u$ 及其已关注用户 $v_i$ 的数据作训练集 $\{u, v_i, \mathbf{X}_i\}_{i=1}^n$ , 其中 $n$ 表示用户 $u$ 的关注用户数,  $v_i$ 表示用户 $u$ 的第 $i$ 个关注用户。 $\mathbf{X}_i$ 表示与用户 $u$ 和用户 $v_i$ 相关的特征向量。

特征向量 $\mathbf{X}_i$ 由以下特征或它们的子集组成:

- 1) 基于用户及其关注用户的微博内容推荐方法计算得到的用户 $u$ 和用户 $v_i$ 之间的相似得分;
- 2) 基于用户标签推荐方法计算得到的用户 $u$ 和用户 $v_i$ 之间的相

似得分; 3) 基于用户社交网络推荐方法计算得到的用户 $u$ 和用户 $v_i$ 之间的相似得分; 4) 基于用户个人信息特征计算得到的用户 $u$ 和用户 $v_i$ 之间的相似度。

其中, 特征1)~特征3)是由3种用户推荐方法计算得到的, 特征4)是用户自身的特征。由于各个特征值范围不一, 故需要对这4个特征进行归一化处理。

使用训练数据集对逻辑回归模型进行训练, 得到可使用的推荐模型。测试数据集类型与训练数据集类型相同, 即具有相同用户特征。对于目标用户 $u$ , 根据推荐模型即可计算出待推荐用户的推荐得分, 按降序排列生成Top- $N$ 推荐列表。

4 实验及结果分析

4.1 数据集

数据集中每个用户的微博信息应包括用户及其关注用户的微博、用户标签、用户社交圈、用户个人信息(地区、性别)。同时为方便进行实验, 本文要求数据集选取的每个用户的关注数不少于15, 粉丝数不大于2 000, 且用户微博数不少于20。

本文依照数据集的要求, 使用python语言编写爬虫程序, 从新浪微博爬取了387个用户及其关注用户的微博信息, 同时爬取了其他4 152个用户及其关注用户的微博信息。将爬取的387个用户作为目标用户, 并将这387个目标用户的关注用户以及其他1 152个用户作为待推荐用户。

将爬取的微博用户数据集分为训练数据集和测试数据集。训练数据集由287个目标用户及其关注用户的微博信息组成, 测试数据集由剩余的100个目标用户及其部分关注用户的微博信息和其他4 000个待推荐用户及其关注用户的微博信息组成。对于测试集中的每个目标用户, 随机选取其40个待推荐用户并隐藏其10个关注用户(共50个用户)作为测试集。模型训练结束后, 对测试数据集进行实验, 得到按降序排列的Top- $N$ 推荐结果。推荐过程如图2所示。

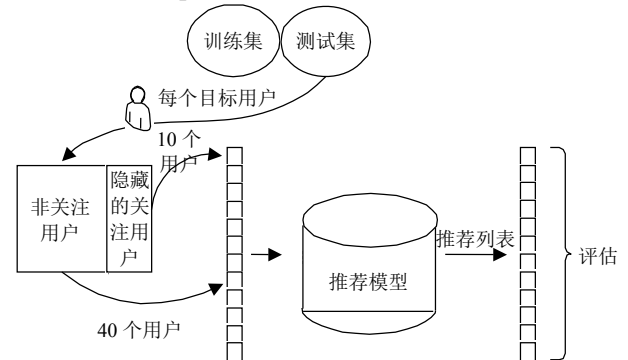


图2 用户推荐过程

## 4.2 评价指标

根据微博实际,用户往往更关注排名靠前的用户,故研究高排名的用户推荐性能具有重要意义。本文重点测试Top-1、Top-3、Top-5、Top-6、Top-8、Top-10六种情况下本文微博用户推荐模型的性能。

Top- $N$ 推荐列表中包含隐藏关注用户的个数是评价用户推荐方法有效性的重要指标<sup>[11]</sup>。本文使用查准率(Precision)、召回率(Recall)两个指标进行推荐方法的评价,有:

$$\text{Precision} = \frac{\text{推荐用户是隐藏的关注用户的个数}}{\text{推荐的总用户个数}N} \quad (12)$$

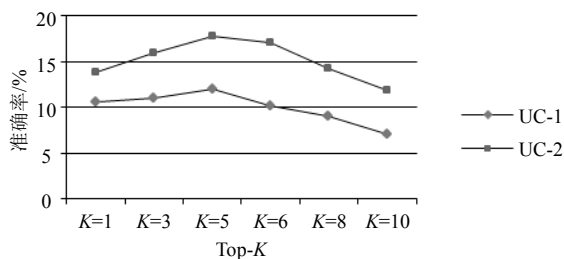
$$\text{Recall} = \frac{\text{推荐用户是隐藏的关注用户的个数}}{\text{隐藏的关注用户总个数}} \quad (13)$$

## 4.3 结果与分析

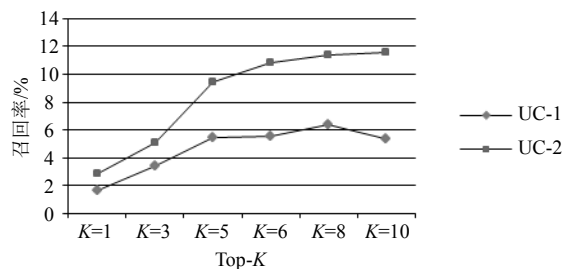
基于测试数据集,本文首先进行了基于用户自身微博进行推荐和基于用户及其关注用户微博进行推荐的实验,并进行了对比分析。本文分别取Top-1、Top-3、Top-5、Top-6、Top-8、Top-10六种情况下的实验结果,即推荐结果的前1个、前3个、前5个、前6个、前8个、前10个用户中,隐藏的关注用户占推荐结果的平均比例。实验结果如图3和图4所示。

在实验结果图3a、3b中,UC-1为基于用户自身微博内容推荐方法的推荐效果,UC-2为基于用户及其关注用户微博推荐方法的推荐效果。可以看出UC-2比UC-1在推荐效果上有了较大提高。这是由于UC-1忽略了关注用户对微博用户的潜在影响,进而忽略了关注用户微博中包含的有用信息,同时用户自身微博数据也不充足;UC-2综合利用用户及其关注用户微博进行推荐,在一定程度上也缓解了微博数据量不足的情况,这使得UC-2的推荐效果优于UC-1。

为验证文本方法(LR)的准确性和有效性,实验对基于协同过滤的用户推荐方法(BPR)、基于兴趣和社交网络的用户推荐方法(BIS)在本文数据集上推荐结果的准确率(Precision)和召回率(Recall)两个指标进行对比,从而进行用户推荐效果的比较。分别取Top-1、Top-3、Top-5、Top-6、Top-8、Top-10六种情况下的实验结果,实验结果如图4所示。



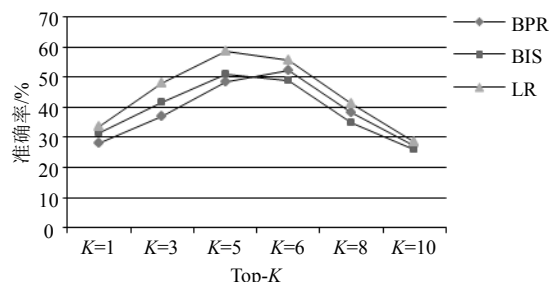
a. 两种方法的准确率



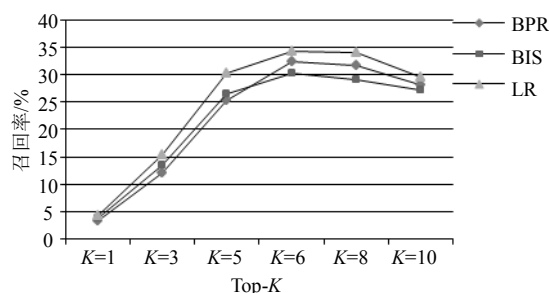
b. 两种方法的召回率

图3 UC-1和UC-2推荐效果比较

由实验结果可以看出:1) LR方法的整体准确率和召回率均优于BIS和BPR方法。2)  $K$ 取值在1~6之间时,在保证召回率不低于BIS和BPR方法的基础上,本文提出的LR方法的准确率明显优于BIS和BPR方法;当 $K$ 值继续增大时,3种方法性能趋于平稳。而用户推荐关注的是排名靠前用户的推荐效果,推荐准确率越高,效果越好。这说明本文LR方法的推荐效果明显优于BIS和BPR方法。



a. 三种方法的准确率



b. 三种方法的召回率

图4 BPR、BIS和LR的推荐效果比较

BPR方法虽然考虑了用户多个特征间的相似性,并取得了一定的推荐效果,但由于该方法存在一定的数据稀疏性,使得靠前排名的推荐并不理想;BIS方法从用户兴趣和社交网络两个方面出发,统筹考虑用户兴趣相似性和社交关系相似性,取得了较好的推荐效果,但其考虑的用户信息仍不够全面,同时BIS靠经验调节相关参数存在一定误差,故性能有待进一步提高;LR方法更全面的从用户兴趣、社交圈、个人信息多个角度出发,使得数据信息较为充分,并通过充分挖掘用户信息中内在支配推荐排

序信息的逻辑回归模型, 综合利用这些特征进行微博用户推荐, 使得采用LR方法进行微博用户推荐的效果有较大提高。

## 5 结束语

本文基于新浪微博进行个性化用户推荐研究。针对微博用户的实际特点, 深入挖掘并广为利用能够反映微博用户兴趣的特征, 并对相关方法进行了效果对比实验, 最后采用逻辑回归方法将多种用户方法和特征融合到一起, 进行个性化用户推荐。通过对爬取的新浪微博数据集进行实验, 发现综合考虑用户及其关注用户的基于内容推荐方法要好于原方法, 取得了较好的推荐效果; 本文提出的逻辑回归模型融合多方法和特征的推荐效果优于现在的推荐方法, 并取得了很好的推荐效果, 从而证明了该方法的有效性。

### 参考文献

- [1] CHEN J, GEYER W, DUGAN C, et al. Make new friends, but keep the old: Recommending people on social networking sites[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM, 2009: 201-210.
- [2] SHANG Y, ZHANG P, CAO Y. A new interest-sensitive and network-sensitive method for user recommendation [C]//2013 IEEE Eighth International Conference on Networking, Architecture and Storage (NAS). Washington: IEEE Computer Society, 2013: 242-246.
- [3] YAN Z, ZHOU J. User recommendation with tensor factorization in social networks[C]//2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Washington: IEEE Computer Society, 2012: 3853-3856.
- [4] GUY I, ZWERDLING N, RONEN I, et al. Social media recommendation based on people and tags[C]//Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2010: 194-201.
- [5] HANNON J, BENNETT M, SMYTH B. Recommending twitter users to follow using content and collaborative filtering approaches[C]//Proceedings of the ACM Conference on Recommender Systems. New York: ACM, 2010: 199-206.
- [6] PENNACCHIOTTI M, GURUMURTHY S. Investigating topic models for social media user recommendation[C]//Proceedings of the 20th International World Wide Web Conferences. New York: ACM, 2011: 101-102.
- [7] 胡大伟. 基于标签协同过滤算法在微博推荐中的研究[D]. 包头: 内蒙古科技大学, 2012.  
HU Da-wei. Tag-based collaborative filtering algorithm microblog recommended[D]. Baotou: Inner Mongolia University of Science and Technology, 2012.
- [8] 孙威. 微博用户兴趣挖掘与建模研究[D]. 大连: 大连理工大学, 2012.  
SUN Wei. Interest mining and modeling for Micro-bloggers of Micro-blog[D]. Dalian: Dalian University of Technology, 2012.
- [10] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. The Journal of Machine Learning Research 2003, 3: 993-1022.
- [11] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[J]. 中文计算语言学, 2002, 7(2): 59-76.  
LIU Qun, LI Su-jian. Word similarity computing based on how-net[J]. Computational Linguistics & Chinese Language Processing, 2002, 7(2): 59-76.
- [12] GOLBECK J, HENDLER J, et al. Film trust: Movie recommendations using trust in Web-based social networks[C]//CCNC'06: Proceedings of the third Consumer Communications and Networking Conference. Washington: IEEE, 2006: 282-286.

编辑 蒋晓