

文章编号:1674-2869(2018)04-0455-07

# 一种基于社交网络友情度的个性化推荐算法

董辉,盛魁,张继美

亳州职业技术学院,安徽 亳州 236800

**摘要:**针对社交网络服务中传统个性化推荐系统的推荐性能和满意度低的问题,在分析社交网络服务中影响个性化推荐各种因素的基础上,引入社交网络用户关系亲密度度量方式——友情度及其三要素,并给出它们的计算方法。研究大型社交数据的用户主题兴趣和各类相似度,设计了一种基于友情度的个性化推荐系统,以提高社交大数据复杂环境下推荐精度及质量,提高用户对推荐结果的满意度。通过实验分析证实了所提出个性化推荐系统比基于PCC和JMSD算法等传统推荐系统的性能更优越,且推荐结果质量要高。最后给出了未来的研究方向。

**关键词:**社交网络;协同过滤;相似度;个性化推荐;友情度

中图分类号:TP391 文献标识码:A doi:10.3969/j.issn.1674-2869.2018.04.021

## A Personalized Recommender Algorithm Based on Friendship Degree in Social Network

DONG Hui, SHENG Kui, ZHANG Jimei

Bozhou Vocational and Technical College, Bozhou 236000, China

**Abstract:** In view of the low recommendation performance and satisfaction of the traditional personalized recommendation system (PRS) in social network services (SNS), this paper introduces metrics of SNS users relationship intimacy, including the degree of friendship and its three factors, and gives their algorithms based on the analysis of various factors that influence personalized recommendation in SNS. A PRS based on the degree of friendship was designed to improve the accuracy and quality of recommendations in the complex environment of social big data, based on investigating the user's interest topics of social big data and calculating various similarities. Through experimental analysis, it was verified that the proposed method was superior to the traditional PRS based on Pearson correlation coefficient and Jaccard mean squared difference.

**Keywords:** social network; collaborative filtering; similarity; personalized recommendation; degree of friendship

随着网络技术发展和社交网络服务的推广,人类交往全面进入在线社交网络服务(social networking services, SNS)时代<sup>[1]</sup>,国内外知名SNS如微博、微信、FaceBook和Twitter等。SNS用户规模

持续扩大,信息交互更加快捷,同时在SNS的基础上又延伸出许多功能强大的应用,如微信在线支付、小程序及微信公众号等,使得SNS的服务内容越来越丰富。在SNS中,用户的自发参与产生了

收稿日期:2018-06-13

**基金项目:**安徽省高校自然科学研究重点项目(KJ2018A0887, KJ2016A493, KJ2015A417, KJ2014A171);安徽省高校振兴计划优秀青年人才支持计划项目(gxyqZD2016529);安徽省亳州市科技创新团队项目(亳组[2014]21号);安徽省亳州市产业创新团队项目(亳组[2015]20号-2)

**作者简介:**董辉,男,硕士,副教授。E-mail:dhstars@163.com。

**引文格式:**董辉,盛魁,张继美.一种基于社交网络友情度的个性化推荐算法[J].武汉工程大学学报,2018,40(4):455-461.

具有一些明显特征属性的社交大数据,主要的特征属性是“用户创建的内容”、“交互信息”及“关系信息”。“用户创建的内容”是指用户自己在SNS上创建的内容及其相关信息,如博客上的帖子、信息发布的时间和标签信息等;“交互信息”是指用户间交互的内容,例如微信上用户之间的信息传输、朋友圈的点赞、评论等;“关系信息”构成一个社交圈的信息,代表SNS用户之间的直接或间接关联关系,如微信上朋友圈、微博的好友圈等。这类社交大数据已经被应用于个性化推荐领域的研究,挖掘用户的需求和偏好,及时向用户推荐满足个性化需求的信息<sup>[2]</sup>。

传统的个性化推荐系统主要基于相似度的协同过滤(collaborative filtering, CF)算法向用户提供推荐。相似度的算法主要有皮尔逊相关系数(pearson correlation coefficient, PCC)、余弦相似度(cosine similarity, CS)及杰卡德均方差(jaccard mean squared difference, JMSD)等方法<sup>[3]</sup>。但是以上算法只适合于利用明确的数据信息(如用户的评分)计算相似度,而大多数社交数据都含有许多隐式信息,这使得传统的推荐系统不能完全适合基于SNS的个性化推荐需求。此外,用户在社交圈信任关系密切的用户信息已被许多研究证实对于提升个性化推荐服务非常有用<sup>[4-6]</sup>,然而,现有的相似度算法不适合于在线社交用户之间亲密关系的度量,因为除了用户创建的内容之外,它们很少关注社交大数据的其它方面的特征,更没有考虑用户之间共享的主题数量。

本文提出了一种描述SNS用户之间亲密度的新方法——友情度,并通过交互、群体和个人三种相似度来计算友情度。该方法考虑了社交大数据的各种特征,提出的个性化推荐系统充分利用用户关心的主题和兴趣及紧密关联的用户的信息,为SNS用户推荐满足个性化需求的兴趣和主题。并使用各种评估指标如平均绝对误差(mean absolute error, MAE)、精确度 Precision、召回率 Recall 和 F1 值等,评估所提出的推荐系统的性能和推荐质量,验证友情度在个性化推荐中的重要作用,证实了基于友情度的个性化推荐系统比基于PCC和JMSD算法的CF推荐系统的性能要优越,且推荐结果的质量要高。

## 1 个性化推荐系统

### 1.1 基于社交网络大数据的个性化推荐系统

个性化推荐系统收集关于用户对项目的偏好

的信息,为用户推荐希望获取的项目。而社交大数据中包含了显式或隐式信息,如博文、朋友圈信息、关注、博文时间和标签信息等,如能合理的利用这些信息,则可以增强个性化推荐。近年来,国内外一些学者在不断努力,试图利用隐式数据信息来改进个性化推荐系统。如潘一腾等<sup>[5]</sup>提出一种基于信任关系隐含相似度的社会化推荐算法,提高了社会化推荐算法的精度;刘静等<sup>[6]</sup>提出了一种个性化的标签推荐系统,使用用户的标记历史和地理信息来生成基于学习方法的推荐系统;赵亚辉和刘瑞<sup>[7]</sup>提出了基于评论的隐式社交关系进行个性化推荐。但是,与考虑SNS用户关联信息以获得推荐的方法相比,上述方法难以充分反映个人倾向并且推荐结果的准确度较低。

### 1.2 基于人际关系的个性化推荐系统

基于人际关系的个性化推荐方式可以分为基于影响力的推荐和基于朋友圈的推荐。在SNS上彼此相关的用户的偏好相似度要高于不相关用户的偏好相似度,他们之间的关联信息显然比不相关用户的信息更具有利用价值<sup>[8]</sup>。

**1.2.1 基于社交影响力的个性化推荐系统** 基于影响力的推荐系统主要用于需要专业领域知识的推荐,如学术或新闻领域的推荐系统。例如陈林等<sup>[9]</sup>定义了对新闻界影响专家模型,提出一种新颖的个性化新闻推荐算法;文献<sup>[10]</sup>基于学术搜索平台 ArnetMiner,融合深度学习方法、社会网络分析和语义搜索等相关技术,进行跨区域协作推荐与用户相关的专家和论文。

**1.2.2 基于朋友圈的个性化推荐系统** 对于SNS用户,人们更倾向于关注熟人和朋友的意见,基于朋友之间信任进行的推荐更容易被接受,因此基于朋友的信息提供的推荐效率更高。但是在SNS中,基于朋友的个性化推荐对于用户关系是均等看待的,没有考虑到用户之间关系远近的亲密程度,大多数研究都是针对一组具有相似兴趣和偏好的社交网络服务用户提供推荐<sup>[11]</sup>。如刘海峰<sup>[12]</sup>提出了基于信任关系的商品推荐方法,改进了推荐的性能,并解决多样性推荐的问题。

本文综合分析诸多学者研究成果的基础上,研究影响用户间亲密度的各种因素,提出基于友情度的个性化推荐系统,并对实验过程和实验结果进行分析,评估所提出系统的性能和推荐质量。

## 2 基于友情度的个性化推荐系统

近年来,国内外一些学者对基于用户之间亲

密度的个性化推荐方法进行了一定的研究。如文献<sup>[13]</sup>考虑了用户互动和社交圈信息来计算他们之间的亲密度,提出一种基于联系强度的个性化模型来增强社交服务,给出了更高质量的推荐信息;文献<sup>[14]</sup>提出一种综合信任模型,构建了基于社会网络关系分析的综合推荐模型,应用矩阵分解技术推断微博的主题分布和用户的兴趣取向,实现微博的个性化推荐。但是,先前的研究没有考虑到影响友情度的各种因素,更没有对影响因素进行多维度分析,导致推荐信息的准确性和精度不高,因此需要一种新的方法分析影响友情度的各种因素,利用它们计算友情度,设计一种高效的个性化推荐系统。

2.1 推荐算法基本流程

本文提出的个性化推荐系统的流程主要分为以下3个阶段9个步骤。

阶段一:处理所需的数据集,此阶段分为5个步骤。步骤1:在社交服务网站上抓取推荐领域相关意见及评论数据构建数据集;步骤2:在数据集中添加与推荐领域相关的正面、中性或负面的意见词词语,构建意见词词典;步骤3:根据意见词典,分析用户生成的内容如微博上的博文、大众点评网的贴文、FaceBook上的帖子等,给出内容的极性,如正面、负面或中性;步骤4:挖掘出与推荐领域及其评论信息相关的用户,提取用户的数据如用户ID值等,创建用户社交圈的用户矩阵;步骤5:利用前面内容的极性分析得分及用户的参考项目的次数得分,计算友情度及推荐系统中用到的隐式偏好值和主题分布值。

阶段二:计算SNS用户间友情度,分为3个步骤。步骤6:利用SNS用户间的交互信息计算交互相似度,交互信息如微信朋友圈或微博上的评论、点赞或转发等信息;步骤7:使用主题分布值及隐

式偏好值,计算个人相似度;步骤8:使用SNS社交圈用户信息计算用户组的相似度。融合以上3个相似度,计算出用户间的友情度。

阶段三:构建推荐系统进行推荐阶段。步骤9:利用用户间友情度、隐式偏好值及前面所计算的相似度值,构建推荐系统,为用户提供高质量推荐结果。

2.2 用户创建内容的极性分类

用户在SNS上发布的内容中表达了自己的看法,代表了自己情绪或感受。根据原先建立的意见词典,可以把用户生成的内容的极性( $O_D$ )分类为正面、中性或负面的内容,分别数字化为1,0,-1。

2.3 用户偏好值的计算

从用户生成的信息中提取关键字(即项目)和包含关键字的主题,用户偏好分为用户项目偏好和用户主题偏好。根据内容的极性和用户生成内容的总数计算项目偏好 $r_{u,i}$ :

$$r_{u,i} = \left( \sum_{D \in D_{u,i}} O_D \right) / |D_{u,i}| \tag{1}$$

式(1)中, $D_{u,i}$ 表示用户 $u$ 关于项目 $i$ 所创建的内容的集合。用户的主题偏好是用户高级偏好,在这里以主题分布值来表示,本文通过使用与主题 $t$ 相关联的用户生成的内容的总数的分布 $\theta_{u,t}$ 值来计算。算法如式(2):

$$\theta_{u,t} = |D_{u,t}| / |D_u| \tag{2}$$

式(2)中, $D_u$ 是关于用户 $u$ 内容的集合、 $D_{u,t}$ 与用户 $u$ 相关的主题 $t$ 的内容的集合。

2.4 友情度的计算

在SNS中,综合考虑影响用户友情度的多种因素,用3种相似度来计算用户之间的友情度。这3种相似度分别是个人相似度 $P$ 、交互相似度 $I$ 和群体相似度 $G$ ,如表1所示。

表1 友情度的三种相似度及其因素  
Tab. 1 Three kinds of similarity and their factors of degree of friendship

三种相似度	各种因素	权重值	含义
个人相似度( $P$ )	兴趣信任度	$Q_r(u,f)$	用户 $u$ 和 $f$ 兴趣的相似度
	主题相似度	$Q_s(u,f)$	用户 $u$ 和 $f$ 之间讨论过多少个主题
	交互频率	$Q_l(u,f)$	用户 $u$ 与他的朋友 $f$ 通讯多少次
交互相似度( $I$ )	交互近因	$Q_R(u,f)$	用户 $u$ 最后与朋友 $f$ 交互的最近因子
	交互时长因子	$Q_L(u,f)$	用户 $u$ 最后与朋友 $f$ 交互生存比
群体相似度( $G$ )	用户间亲密关系	$Q_A(u,f)$	用户 $u$ 的社交圈与朋友 $f$ 的相似度

**2.4.1 个人相似度** 个人相似度表示 SNS 上用户生成内容之间的相似度,可以通过项目的偏好值和主题分布来计算。用户生成的内容会涉及到各种喜好及兴趣,特别是在类似于大众点评、微博等社交网站上,在计算用户对项目的偏好时,共同兴趣的比例比仅计算用户偏好相似度更为重要。因此,本文利用 JMSD 算法,来计算用户之间的偏好相似度。

$\sigma_{\text{MSD}}$  为均方差(mean square difference, MSD)和 Jaccard 系数的乘积。本文中 MSD 系数表示用户  $u$  和  $f$  的项目偏好值之间的差异的平均值,用  $\sigma_{\text{MSD}}$  表示, Jaccard 系数  $J(u, f)$  是 SNS 上用来计算用户共同兴趣的比率。计算方法分别如下:

$$\sigma_{\text{MSD}}(u, f) = \frac{\sum_{i \in I} d_{u, f}^i}{|d_{u, f}|}$$

$$J(u, f) = \frac{|r_u \cap r_f|}{|r_u \cup r_f|} = \frac{|d_{u, f}|}{|r_u| + |r_f| - |d_{u, f}|}$$

其中  $d_{u, f}^i$  表示用户  $u$  和  $f$  关于项目  $i$  的偏好值之差的平方,  $d_{u, f}$  为一组  $d_{u, f}^i$ 。如果  $r_{u, i}$  和  $r_{f, i}$  皆不 null,  $d_{u, f}^i = (r_{u, i} - r_{f, i})^2$ , 否则  $d_{u, f}^i = \text{null}$ ;  $r_u$  和  $r_f$  分别表示用户  $u$  和  $f$  的一组偏好值。则 JMSD 的算法如式(4)所示:

$$\sigma_{\text{JMSD}}(u, f) = \sigma_{\text{MSD}}(u, f) \times J(u, f) \quad (4)$$

SNS 用户创建大量与推荐领域主题相关的内容,用户间的关系越亲密,用户之间交流的主题范围就越广,而且 SNS 用户之间的主题的相似度是实现个性化检索的主要元素之一。因此,个人相似度的计算不仅考虑项目信任,还要考虑用户共享的主题数量及主题的相似度,以便在推荐领域根据个人倾向提供高质量的推荐结果。用户间的主题相似度可以通过使用散度(kullback-leibler divergence, KL)计算<sup>[15]</sup>。算法如式(5)所示:

$$Q_A(u, f) = \frac{1}{\varepsilon_{\text{KL}}(\theta_{u, t} || \theta_{f, t} + 1)} \quad (5)$$

则个人相似度计算式如式(6),取项目信任度和主题相似度的加权之和,表示项目与用户提及的主题的相似度。

$$\text{Sim}_p(u, f) = \alpha_p \cdot Q_1(u, f) + \beta_p \cdot Q_A(u, f) \quad (6)$$

式(6)中  $\alpha_p + \beta_p = 1$ 。

**2.4.2 交互相似度** 在 SNS 中,用户通过信息的交互来分享彼此的好恶及情感,若用户的好恶或偏好相似,则他们会积极分享信息。因此交互相似度可以通过用户之间“互动的广度和深度”来衡

量,用户之间的交互相似度又是计算友情度的重要因素。在许多社交网站中,用户都可以通过转发、评论、关注、回复等功能进行相互交互,因此可以使用这些信息,与交互频率、交互近因等属性结合起来共同计算交互相似度。

首先,计算 SNS 用户的交互频率。用户交互频率反映了用户之间的关系远近,关系近的用户会频繁的积极交流,交流的信息也更有价值。用户交互频率算法如式(7)所示,其中  $q(u, f)$  表示用户  $u$  和  $f$  联系的次数。

$$Q_1(u, f) = \begin{cases} \lg\{1 + q(u, f)\} & \text{if } \{q(u, f) < 10\} \\ 1 & \text{if } \{q(u, f) \geq 10\} \end{cases} \quad (7)$$

其次计算交互近因和交互时长因子。交互近因和交互时长因子是与用户之间的联系时间密切相关的权重。SNS 大数据中,最新的信息非常重要,最近的联系意味着用户分享当前的兴趣,因此用户最近是否接触过是衡量亲密关系的关键因素。用户  $u$  和  $f$  的交互近因可表示为式(8)所示:

$$Q_R(u, f) = R(u, f) / L(u) \quad (8)$$

$R(u, f)$  表示用户  $u$  和  $f$  最近一次联系到当前时间所经过的时间,  $L(u)$  表示收集用户  $u$  的数据的时间。

一般地,在 SNS 中与他人交互时间较长的人的信息比交互时间较短甚至没有沟通的人的信息更重要,如果用  $L(u, f)$  表示用户  $u$  和  $f$  之间的一次交互时长,则交互时长因子算法如式(9)所示:

$$Q_L(u, f) = L(u, f) / L(u) \quad (9)$$

本文交互相似度是通过  $Q_1(u, f)$ ,  $Q_R(u, f)$  和  $Q_L(u, f)$  3 个权重值进行加权来计算的,如式(10)所示:

$$\text{Sim}_I(u, f) = \alpha_I \cdot Q_1(u, f) + \beta_I \cdot Q_R(u, f) + \chi_I \cdot Q_L(u, f) \quad (10)$$

其中  $\alpha_I + \beta_I + \chi_I = 1$ 。

**2.4.3 群体相似度** 群体相似度是指不同用户在 SNS 上所属社交圈的相似度,可以用来衡量用户与他的朋友之间的亲密关系。在 SNS 的社交圈里,用户倾向于将与他们密切接触或具有类似兴趣的用户作为他们的朋友,因此拥有类似朋友列表的不同用户所在社交圈相似度较高,即群体相似度较高。所以可以通过用户  $u$  和  $f$  共享的朋友的比率来计算  $u, f$  所在社交圈的相似度。在 SNS 社交圈里,用户朋友关系可分为直接和间接朋友关系。通过间接朋友关系,也可以获取用户社交圈中有用的信息和资源。本文用朋友级别的概念来描述

群体相似度。如图1所示,将用户 $u$ 的直接关联朋友的级别定义为一 $n$ 级,表示为 $F_u^1$ ,用户间接关联(第 $n$ 层级)朋友的级别定义为 $n$ 级,以 $F_u^n$ 表示,例如 $F_u^2$ 表示用户 $u$ 的两个边缘内连接的一组用户。

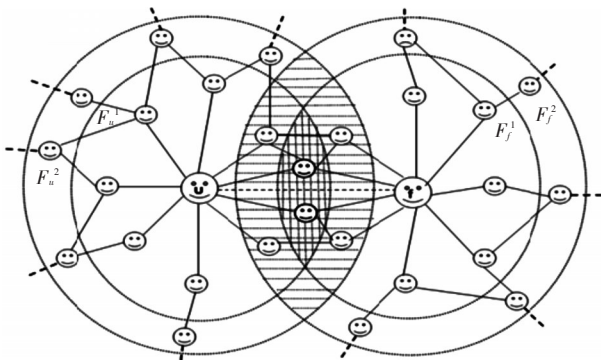


图1 群体相似度  
Fig. 1 Swarm similarity

在大多SNS中,几乎所有的用户的朋友用户都在6层以内,即 $n=6$ 。结合Jaccard系数,则用户群体相似度的算法表示式(11)所示:

$$\text{Sim}_G(u, f) = \alpha_G \frac{|F_u^1 \cap F_f^1|}{|F_u^1 \cup F_f^1|} + \beta_G \frac{|F_u^2 \cap F_f^2|}{|F_u^2 \cup F_f^2|} + \dots + \chi_G \frac{|F_u^n \cap F_f^n|}{|F_u^n \cup F_f^n|} \quad (11)$$

其中 $(\alpha_G + \beta_G + \dots + \chi_G = 1)$ 。

2.4.4 友情度的计算 在SNS中,本文用友情度来度量用户之间的亲密关系,根据前文分析友情度主要体现在个人相似度、交互相似度和群体相似度几个重要的方面,综合式(6)、式(10)和式(11)3种相似度算式,本文SNS中友情度 $Q_F(u, f)$ 算法如式(12)所示:

$$Q_F(u, f) = \delta [\text{Sim}_P(u, f), \text{Sim}_I(u, f), \text{Sim}_G(u, f)] \quad (12)$$

其中, $\delta$ 表示根据每个元素的重要性提供权重值差异的加权矢量。

在社交网络中,直接或间接关联的用户的信息比未关联的用户的信息更为重要,根据这些用户的信息结合用户友情度和社交圈朋友的级别的定义,给出对用户 $u$ 关于项目 $i$ 的偏好值预测式,如式(13)所示:

$$p_{u, f} = \bar{r}_u + \frac{\left[ \sum_{f \in F_u^1 \cup F_u^n} (r_{f, i} - \bar{r}_f) * QF(u, f) \right]}{\left[ \sum_{f \in F_u^1 \cup F_u^n} QF(u, f) \right]} \quad (13)$$

式(13)中 $\bar{r}_u$ 和 $\bar{r}_f$ 是用户 $u$ 和 $f$ 对所有项目的平均偏好值, $r_{f, i}$ 是用户 $f$ 对项目 $i$ 的偏好值。

3 结果与讨论

3.1 实验环境与数据集的选取

实验环境为 Intel Xeon E5-2630\*2CPU、32 GB DDR4 内存、3 TB Sata3 硬盘、Windwos Server 2008 系统服务器,实验数据集来源于新浪微博音乐及电影 2018 年 1 月份乐评及影评博文数据的抓取,共 28 部在国内上映的电影和亚洲音乐排行榜 60 首音乐、8 138 315 条评论博文和 2 368 372 个微博会员相关的信息。经过整理,实验中把该数据集分为评论博文信息、微博会员关系信息和评论意见词典。博文信息有博文 ID、博文内容、发表时间、转发量、回复数、关注量、点赞量、博主 ID、对博文的评论及关注者的微博会员 ID;微博会员关系信息包括有微博会员的 ID、会员名及会员关注者的 ID 及会员名;意见词典主要包括电影的正负面词汇列表,用来标志电影评论博文的极性情况。

3.2 实验过程分析

实验对数据集进行了交叉实验,目的是分析评估本文所提出的个性化推荐系统。实验中,把微博会员数据集分成两部分,其 20% 用于测试用户集、80% 视为训练用户集,并建立了用户对用户的友情度矩阵,基于交互相似度的考虑,20% 与其他用户积极互动的活跃用户是测试用户。在数据集中,实验中使用了 28 个电影和 60 个音乐共 88 个项目,但电影评论活跃用户提及的项目数量平均值大约为 10 个、音乐评论活跃用户大约为 16 个。对测试集合使用 Leave-One-Out 交叉验证<sup>[16]</sup>。首先基于测试集,计算用户之间的友情度,建立用户-用户友情度矩阵,基于此矩阵计算用户间交互和群体相似度;其次利用测试用户与训练用户之间的训练项目的偏好值及主题分部值,来计算个人的相似度;最后利用友情度 3 个重要元素:交互相似度( $I$ ),群体相似度( $G$ )和个人相似度( $P$ )的组合来计算用户之间的相似度。

3.3 实验结果分析

协同过滤个性化推荐系统是最受欢迎的推荐方式,而基于 PCC 和 JMSD 是在 CF 推荐系统中常用的系数,它的有效性已经在经典的的推荐系统数据集(如 Movielens)中得到证实。本文所提出的系统利用直接或间接连接到 SNS 的用户的信息进行推荐,把本文推荐系统的性能与基于 PCC 和 JMSD 的 CF 推荐系统进行比较。使用衡量个性化推荐系统性能的几个主要指标 MAE、精确度、召回率、 $F_1$  值验证了推荐系统的推荐质量。

3.3.1 IGP 实验结果分析 本文针对精确度、召回率和  $F_1$  值对友情度 3 个重要元素  $I$ 、 $G$ 、 $P$  及其组合进行实验。实验结果如图 2 所示。

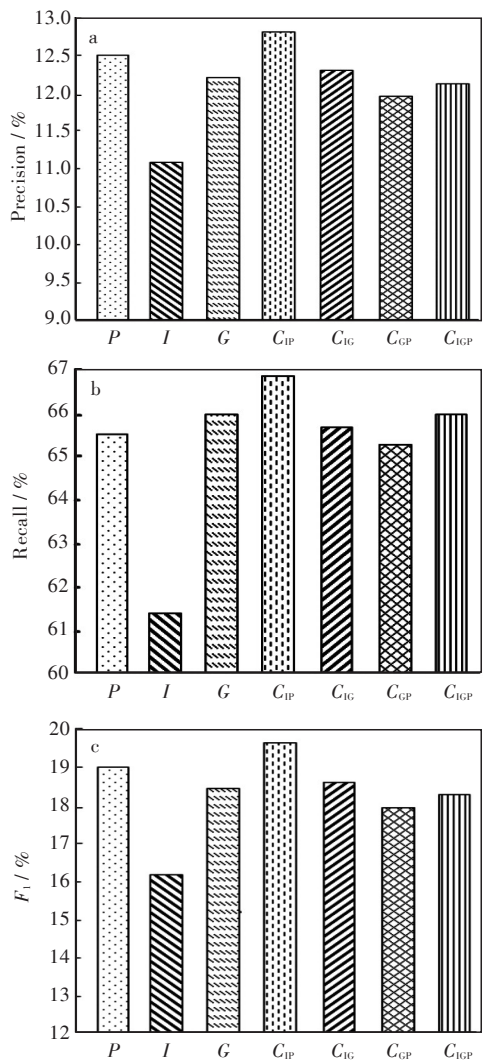


图 2 友情度三要素及其组合实验结果：  
(a)精确度，(b)召回率，(c) $F_1$  值

Fig. 2 Experimental results of three elements of degree of friendship and their combination:  
(a)precision, (b)recall, (c) $F_1$  value

从图 2 中可以发现，在所有的组合中， $I$  和  $P$  的组合 ( $C_{IP}$ ) 表现出最好的性能。因此，结合式 (12)，选择  $C_{IP}$  作为友情度。

3.3.2 MAE 实验结果分析 MAE 主要用于预测值的误差测量，定义为公式 (14)：

$$\delta_{MAE} = \frac{1}{|I|} \sum_{i \in I} \left( \frac{1}{|O_{u,i}|} \sum_{u \in O_{u,i}} |p_{u,i} - r_{u,i}| \right) \quad (14)$$

式 (14) 中  $O_{u,i}$  表示用户集合， $p_{u,i}$  和  $r_{u,i}$  在测试用户集  $U$  中都不为 null 值。MAE 评估实验的结果如图 3 所示，实验中  $K$  近邻 ( $K$ ) 的范围是 100 到 2 000 之间。

从图 3 中可以分析，本文所提出的推荐算法比基于 PCC 和 JMSD 的 CF 推荐具有更低的错误率。基于友情度推荐预测偏好值的准确性在整体范围内分比 PCC 提高 20% 左右，与 JMSD 相比，在  $K < 1\,000$  时，它的性能略高，但  $K \geq 1\,000$  时，性能提高了大约有 8%， $K$  的数量越多，性能改善越高。

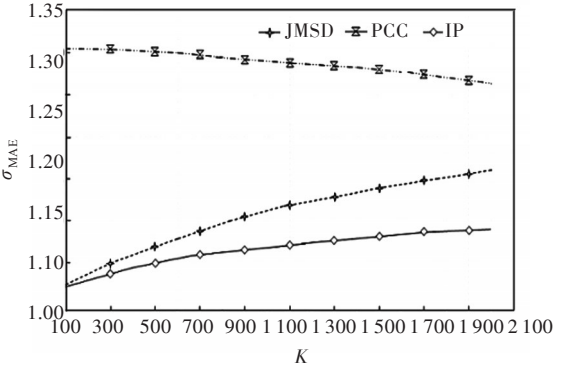


图 3 推荐质量实验结果

Fig. 3 Experimental results of quality of recommendations

3.3.3 针对推荐有效性的实验分析 MAE 可能是衡量推荐算法性能的一个很好的指标，但是它不能准确衡量用户对推荐结果的满意度，因此有必要使用新的算法测量用户对推荐结果的满意度即评估推荐的有效性。通常精确度  $p$ 、召回率  $r$  和  $F_1$  值被用于衡量个性化推荐结果的有效性，本文中精确度、召回率和  $F_1$  值表示如式 (15)~式 (17)，实验结果如图 4 所示。

$$p = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i \in Z_u | r_{u,i} \geq \theta\}|}{|N|} \quad (15)$$

$$r = \frac{1}{|U|} \sum_{u \in U} \frac{|\{i \in Z_u | r_{u,i} \geq \theta\}|}{|\{i \in I | r_{u,i} \geq \theta\}|} \quad (16)$$

$$F_1 = 2 \times (p \times r) / (p + r) \quad (17)$$

其中  $Z_u$  表示用户提供的  $N$  个推荐项的集合， $r_{u,i} \in [0, 1]$ ， $\theta$  表示阈值。为了衡量精确度、召回率和  $F_1$  值，不仅要根据预测偏好值对推荐项目进行排序，而且还要确定推荐项目相关性判断的阈值。根据式 (1) 进行多次计算发现，式 (15) 和式 (16) 中如果  $r_{u,i} > 0.5$ ，则表示用户在他所有的博文对于项目  $i$  的正面评论都多于 1 次，因此可以判断推荐结果与  $i$  是相关的。因此，将阈值设置为 0.5。

从实验结果可以看出，就精度而言，所提出的系统与 PCC 和 JMSD 相比分别显示提高约 8% 和 5%；在召回率方面，与 PCC 相比，所提议的系统性能提高约 8%，与 JMSD 相比约为 6%；对于  $F_1$  值，与 PCC 相比，所提出的系统性能提高约 10%，与

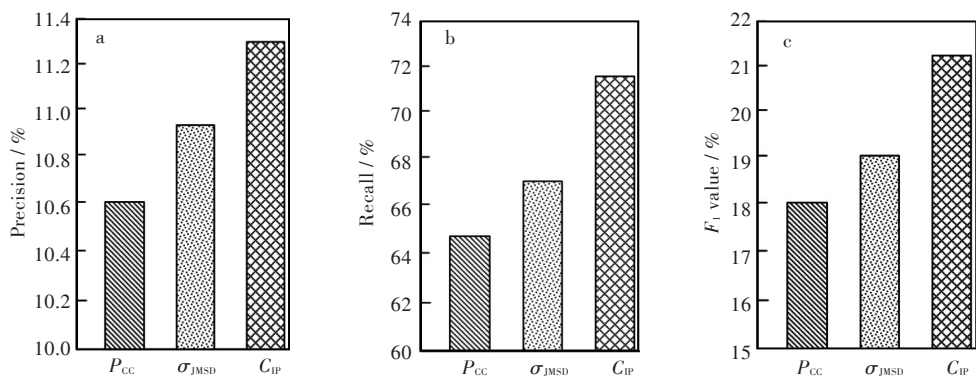


图4 推荐结果有效性的比较:(a)精确度,(b)召回率,(c) $F_1$ 值

Fig. 4 Comparisons of recommendation results:(a)precision,(b)recall,(c) $F_1$  value

JMSD 相比约为 9%。可见,本文所提出的推荐系统推荐质量上比基于 PCC 和 JMSD 的 CF 推荐系统整体上分别提高 8.5% 和 7% 左右。

4 结 语

基于 SNS 提出的友情度对于推荐有很大帮助,只要它用户之间的交互数据、好友列表信息或具有多个级别的项目,如微信、陌陌、微博等社交网络拥有用户之间的联系信息和用户之间评论的信息,利用这些信息就可以计算出用户之间友情度,而大多数基于 SNS 的推荐系统都可以使用友情度进行推荐。

使用 1 个月的微博电影、音乐评论数据,针对 MAE、精度、召回率、 $F_1$  值进行了比较实验。结果表明,综合考虑 SNS 上的用户间的友情度,对于推荐结果的有效性有很大的影响,所提出的系统在所有度量方面比基于 JMSD 和 PCC 的推荐表现出较好的性能和推荐质量,对于提高推荐的准确性是有用的。虽然推荐质量的差异并不十分显著,但所提出的系统在所有度量指标中值都是最高的。

参考文献:

[1] 胡长军,许文文,胡颖,等. 在线社交网络信息传播研究综述[J]. 电子与信息学报,2017,39(4):794-803.  
[2] 陈婉,朱世杰. 基于数据挖掘的社交网络信息推荐与预测方法研究[J]. 内蒙古师范大学学报(自然科学汉文版),2018,47(2):128-130.  
[3] 冷亚军,陆青,梁昌勇. 协同过滤推荐技术研究综述[J]. 模式识别与人工智能,2014,27(8):720-734.  
[4] 叶卫根,宋威. 融合信任用户间接影响的个性化推荐算法[J]. 计算机工程与科学,2016,38(12):2579-2586.  
[5] 陈婷,朱青,周梦溪,等. 社交网络环境下基于信任的

推荐算法[J]. 软件学报,2017,28(3):721-731.

[6] 潘一腾,何发智,于海平. 一种基于信任关系隐含相似度的社会化推荐算法[J]. 计算机学报,2018,41(1):65-79.  
[7] JING L, ZECHAO L, JINH T, et al. Personalized geo-specific tag recommendation for photos on social websites[J]. IEEE Transactions on Multimedia, 2014, 16(3): 588-600.  
[8] 赵亚辉,刘瑞. 基于评论的隐式社交关系在推荐系统中的应用[J]. 计算机应用研究,2016,33(6):1628-1632.  
[9] CHEN L, XIE R N, LI L, et al. PRemISE: personalized news recommendation via implicit social experts[J]. Acm International Conference on Information & Knowledge Management,2012,254(1):1607- 1611.  
[10] 王妍,唐杰. 基于深度学习的论文个性化推荐算法[J]. 中文信息学报,2018,32(4):119.  
[11] 张朝恒,何小卫,陈勇兵. 基于社交网络信息的协同过滤推荐算法[J]. 计算机技术与发展,2017,27(12):28-34.  
[12] 刘海峰. 电子商务中基于信任关系的商品推荐方法[D]. 大连:大连理工大学,2015.  
[13] SANDRASERVIA-RODRÍGUEZ, REBECA P. DÍAZ-REDONDO, ANAFERNÁNDEZ-VILAS, et al. A tie strength based to socially enhance applications and its enabling implementation: My socialSphere[J]. Expert Systems with Applications, 2014,41(5):2582-2594.  
[14] 李慧,马小平,施珺,等. 结合信任度与社交网络关系分析的微博推荐方法研究[J]. 中文信息学报,2017,31(2):146-153.  
[15] 王永,邓江洲. 基于 KL 散度的用户相似性协同过滤算法[J]. 北京邮电大学学报,2017,40(2):110-114.  
[16] 刘学艺,李平,邵传厚. 极限学习机的快速留一交叉验证算法[J]. 上海交通大学学报,2011,45(8):1140-1145.