

基于微博的智能数字图书馆个性化推荐

黄婷

(上海理工大学图书馆, 上海 200093)

摘要:针对传统图书馆在为用户提供准确多样的个性化服务方面的不足,文章提出基于用户微博信息的个性化推荐模型。通过挖掘用户的微博文本短信息,引入本体描述用户对本体领域概念的偏好,建立基于微博用户的多层语义社区网络模型,并将该模型应用于混合协同过滤推荐系统中,实现智能数字图书馆的个性化推荐。

关键词:微博;本体;多层语义社区网络;智能图书馆;个性化推荐

中图分类号:G250.76

DOI: 10.3772/j.issn.1673-2286.2015.11.009

1 引言

数字图书馆最早于1994年提出,作为知识仓储和信息交流的渠道,成为未来社会的公共信息中心和枢纽^[1]。现今信息资源空前膨胀,其形式由单一化发展到如今的海量化、复杂化和多样化,用户的信息需求也不断提升。为了更好地为用户服务,数字图书馆需要进入新的知识阶段,自动发现知识和帮助用户获取知识,满足用户更加个性化、多样化和智能化的服务需求。

目前,多数高校图书馆只是依据借阅次数进行热门读物的推荐,且其图书标签只有学科主题等信息,不能很好地与用户兴趣相对应。微博作为新兴的信息媒介,近些年来取得了迅猛发展,用户发布的帖子内容真实反应了其兴趣取向。其中,豆瓣读书是国内信息最全、用户数量最大且最为活跃的读书网站,它的图书几乎涵盖了高校图书馆的馆藏资源,而且标签更加丰富。

信息推荐中,社会关系往往比推荐内容与用户喜好的匹配程度更加重要。文献^[2-3]将信任机制融入到个性化推荐过程中,提出基于社会网络信任的多样推荐算法;A. Bellogin^[4]等提出基于社会推荐和协同过滤的混合推荐算法。基于微博的推荐研究,大部分文献^[5-6]是对微博用户进行好友推荐、散列标签推荐或新闻推荐,将微博作为信息资源推荐的工具;曾琦^[7]将图书馆资源进行整合,把图书馆微博作为一个信息共享的平台,开展各种资源的推荐服务;蔡淑琴^[8]设计了基于社会化网

络修订的协同过滤推荐,有效的实现了个性化推荐。

我们将人工智能原理运用到数字图书馆中,重点在于智能推荐。通过算法分析用户的微博文本信息,挖掘用户的行为数据,从关键词入手,引入本体概念,描述用户对本体领域概念群的偏好,建立基于微博用户的多层语义社区网络模型,最大程度地自动分析获取用户的兴趣,进而针对用户兴趣从豆瓣网信息源中抽取相应的信息并过滤,给用户推送其最感兴趣、更多样的阅读材料。这样不仅省去了用户在面对海量资源时的手工检索和浏览时间,还可以为用户提供个性化和多样化的服务。

2 个性化推荐系统设计

为了将微博用户信息集成进协同过滤推荐系统,实现智能数字图书馆的个性化推荐,本文提出了基于微博的智能数字图书馆个性化推荐模型框架,如图1所示。

微博用户登录到图书馆系统,基于领域本体和获取到的用户微博信息内容建立个性化本体用户兴趣模型,随着用户微博内容的不断更新,该模型能够自动识别和适应用户行为的变化。最终,通过个性化本体用户兴趣模型获得多层语义社区,并参考用户的兴趣偏好和豆瓣数据集产生推荐列表。该框架最重要的特征有:

(1) 用户利用微博账号登陆图书馆系统,推荐框架通过用户的微博信息获取用户的兴趣模型,对用户的兴

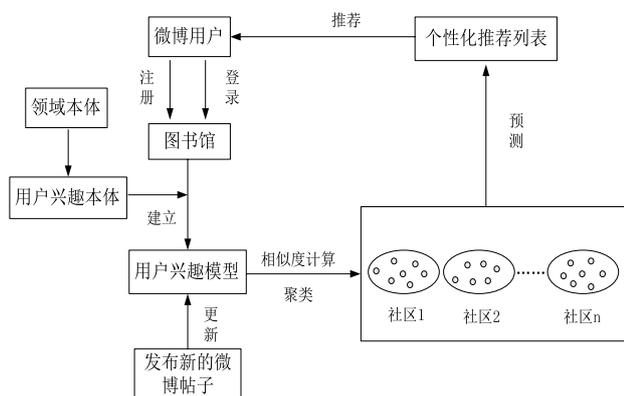


图1 个性化推荐框架

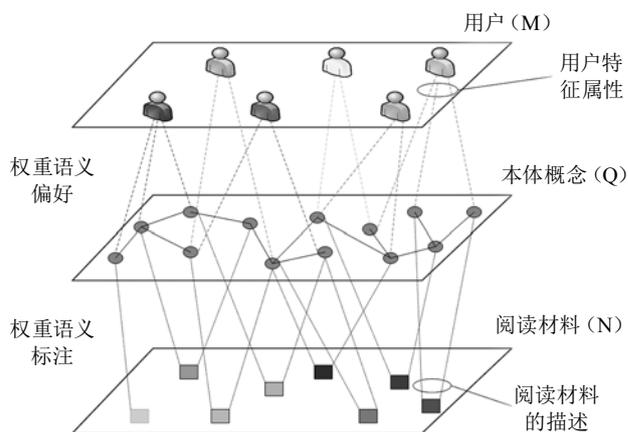


图2 基于本体的用户特征和阅读材料的描述

趣特征属性进行相似度计算,使具有相同语义特征的用户聚集成一个社区,从而形成多层语义社区网络;

(2) 随着用户微博信息的不断更新,用户的兴趣模型发生变化,该模型能自动识别和适应该变化,并更新推荐列表。

3 用户兴趣模型表示机制

基于本体的知识表示相较于基于关键词或者项目的模型显得更加丰富,不容易引起歧义。它为用户兴趣由粗到细粒度的表示提供了足够的依据,是处理用户偏好之间细微差别的关键。此外,本体是标准化的(如RDF和OWL),提供的推理机制可以加强个性化特征,这些特征在我们的个性化推荐模型中将被利用。

基于本体的个性化框架^[9-10],用户偏好向量表示为 $\vec{u}_i=(u_{i,1}, u_{i,2}, \dots, u_{i,M})$ 其中 $u_{i,j} \in [0,1]$,衡量的是在领域本体 O 中,用户 $u_i \in U$ 对概念 $c_j \in O$ (类或实例)的兴趣程度, Q 是本体里总的概念数。同样,我们定义检索空间中阅读材料 $d_k \in D$ 的概念权重向量 $\vec{d}_i=(d_{k,1}, d_{k,2}, \dots, d_{k,N})$,它表示同一个向量空间中的用户偏好。基于上述逻辑表示,用户对阅读材料内容的偏好程度可以通过比较用户的特征和阅读材料的标注向量来衡量,这些方法能以个性化的方式优先、过滤和排序阅读材料。

下图2形象地表明了基于本体的双重空间的知识表示, M 和 N 分别表示系统中登记用户和阅读材料的数量。

目前,用户在使用很多推荐系统时,必须采用手工方式为自己添加多个个性标签。由于不明白添加个性标签对推荐结果的影响,用户通常不愿意花费时间添加,更不用说为标签添加权重。如果能够自动识别用户

的偏好特征,用户配置稀疏的问题就能迎刃而解了。

由于每个微博用户的特征属性都有所不同,文章运用了基于用户不同兴趣特点的个性化本体概要方法^[11],发现隐藏的知识 and 特征属性间的关系。通过挖掘用户的微博文本信息,捕捉用户的特征属性,然后建立个性化的本体概念领域。知识发现提供一个个性化本体表示,其能够自动地推理和适应用户行为的隐式变化,无需用户的干预。语义偏好扩展机制,通过发现本体里与其它概念的语义关系,明确本体概念领域里的用户特征属性。该扩展基于约束传播激活(CSA)策略^[12-13],每次遍历关系时,通过为用户偏好强度施加一个衰减因子控制其扩展,如图3所示。

这样,系统输出的阅读材料排名列表不仅考虑了当前用户的偏好,而且通过用户特征属性和领域本体考虑到了用户所隐藏的偏好,使用户的推荐更加多样化。用户的特征属性是很简化的,不同用户之间特征属性的匹配度也很低。基于本体的特征属性表示比基于关键字的特征属性表示具有更好性能。该扩展不仅对用户个性化的表现很关键,而且对下一部分描述的聚类策略至关重要。

4 多层语义社区网络

4.1 用户的聚类

社区网络中,普遍认为有共同兴趣的人们之间存在其他相关联的兴趣^[14]。例如,对旅游感兴趣的人也有可能对摄影、美食或语言有兴趣。事实上,这种假设是多数推荐系统的技术基础^[15-18]。为了根据多个用户的共同

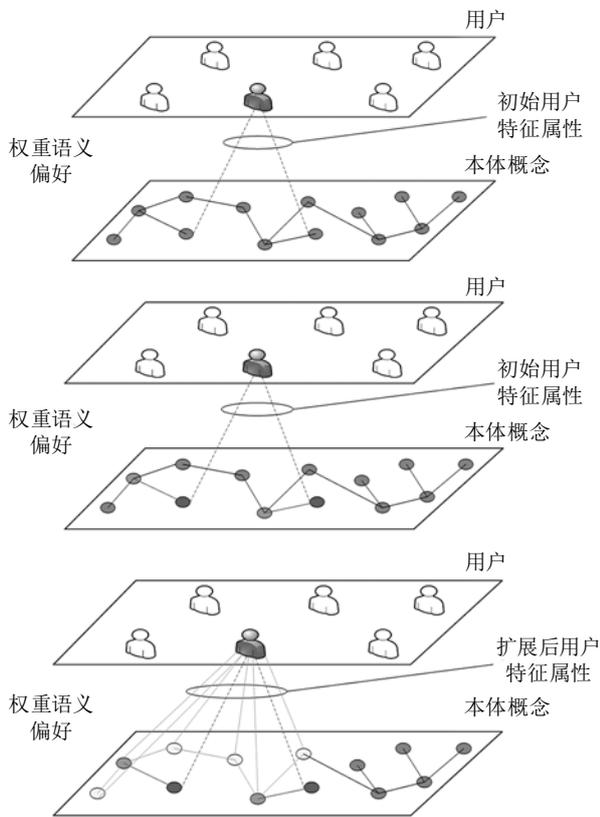


图3 用户语义特征扩展

兴趣组聚类出本体领域概念群，我们也假定这种设想成立。

利用概念之间的联系和用户对概念的偏好，基于出现在用户偏好里的概念相关性，聚类成语义空间。之后，根据用户特征属性在概念群中的投影将其划分成用户的兴趣子集。然后，根据产生的兴趣子集比较用户，可以发现两个用户之间的权值关系。

向量 $c_j=(c_{j,1}, c_{j,2}, \dots, c_{j,Q})$ 表示至少一个用户对概念向量 c_j 的偏好比，其中 $c_{j,i}=u_{i,j}$ ， $c_{j,i}$ 表示在用户 u_i 的语义特征中概念 c_j 的权重。基于这些向量，通过应用分层聚类策略^[19-20]形成多层语义社区网络。获得的聚类表示大量用户共享的概念用户向量空间中的偏好（兴趣主题）群体。

一旦创建了本体领域概念集群，每个用户将被分配到一个特定的集群中。用户 $u_i=(u_{i,1}, u_{i,2}, \dots, u_{i,N})$ 和集群 Cr 的相似度计算如下：

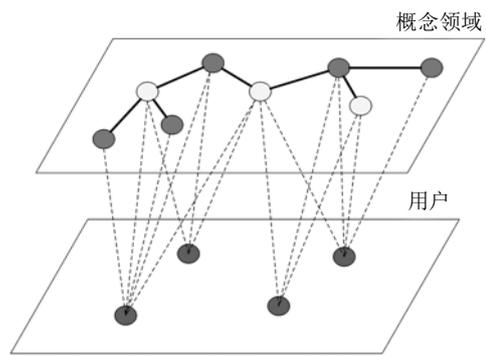
$$sim(u_i, Cr) = \frac{(\sum_{c_k \in Cr} u_{i,j})}{|Cr|} \quad (1)$$

这里， c_j 表示的概念相当于用户偏好向量 $u_{i,j}$ 的组成， $|Cr|$ 是所在概念集群中的概念总数。用户被分配到与其相似度最高的概念集群中，这样就创建了具有共同兴趣偏好的用户群。

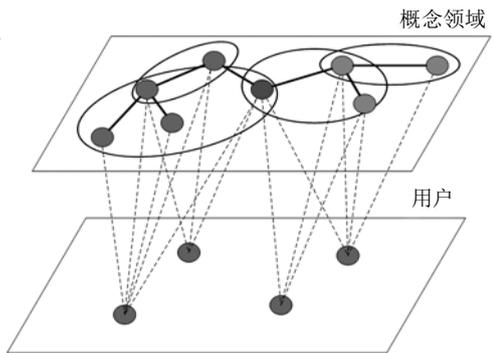
根据上节用户语义特征扩展，图4阐述了用户的聚类过程。

第一步，挖掘用户的微博文本信息，提取用户的兴趣偏好，扩展用户语义偏好，发现隐藏在用户兴趣爱好之间的语义关系，建立个性化的本体领域概念。

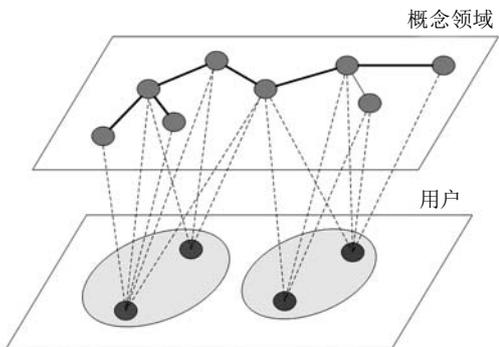
第二步，依据用户的偏好空间向量，将语义本体领域



(a) 语义偏好扩展



(b) 语义概念聚类



(c) 用户聚类

图4 建立用户兴趣模型

概念划分成不同的语义群,每个语义群代表一个语义特征属性,每个语义特征属性又包含多个不同的本体概念。

第三步,根据用户的偏好权重,聚类成不同的社区。

4.2 语义社区网络的形成

概念领域和用户集群被用来发现语义社区网络。一方面,利用用户的偏好权重,用户成员对每个集群的聚类程度和集群间的相似性发现两种不同的社会网络关系:个人的和团体的。另一方面,利用概念集群将用户的特征属性划分成若干个语义片段。每个片段对应一个概念集群,代表一个用户偏好的子集,这些用户共享的兴趣偏好引发了聚类过程。通过在用户特征属性中引入更多这样的结构,可以定义在不同层次中用户间的关系,获得一个用户的多层网络(见图5)。

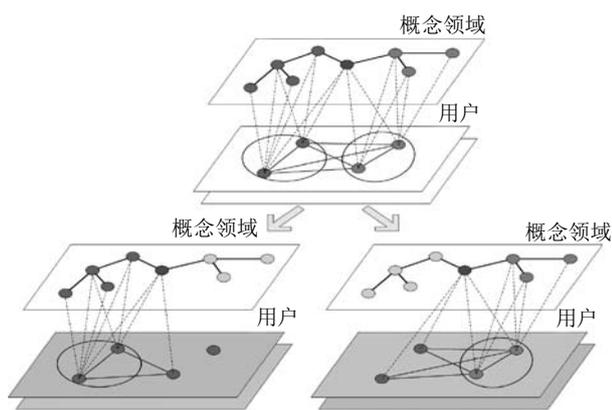


图5 语义社区网络的形成

图5描述了聚类出的两个用户群的情况,每个用户群中的用户特征属性被划分到两个语义层。在每个语义层,得到用户间的权重关系,建立不同的兴趣社区。这些社区有很多潜在的应用,可以在协同过滤推荐中被利用,因为社区不仅建立了用户之间的相似性,而且为不同的信息需求提供了不同语义上下文。下节将介绍两种不同的推荐模型。

5 个性化协同过滤推荐

个性化服务是推荐系统根据用户的微博信息行为,动态调节用户兴趣模型,自动适应用户的动作,并为用户推荐可能感兴趣的内容的过程。个性化推荐系统可以使用户快速、准确地得到所需信息。在个性化推荐中,协同过滤是当前应用最成功的技术。协同过滤技

术主要强调人与人之间的交互,结合基于用户兴趣的潜在社区网络关系和语义项的偏好信息,将对协同过滤推荐系统带来重要的影响。

5.1 用户兴趣度遗忘

由于用户的微博信息不断更新,导致用户的兴趣模型也不断更新,从而用户的兴趣偏好发生变化。为了适应用户不断变化的兴趣,我们引进用户概念兴趣度遗忘因子^[21]。将个性化本体中的概念节点进行遗忘,以适应用户兴趣的变化,提高推荐的准确度。遗忘因子公式为:

$$f(t)=e^{-\frac{\log_2}{T_{now}-T_{visited}}(T_{now}-T_{created})} \quad (2)$$

其中 T_{now} 表示当前时间, $T_{created}$ 表示本体概念节点在个性化本体中被创建的时间, $T_{visited}$ 表示概念节点被更新的时间,取 $T_{now_i}-T_{visited}$ 作为半衰期(用户兴趣遗忘一半的天数)。

5.2 个性化推荐模型

协同过滤分析用户兴趣,在用户群中找到与指定用户的相似(兴趣)用户,综合这些相似用户对此信息的喜好程度预测。

语义社区网络的形成可以使我们进一步研究社区网络在推荐系统中作用。考虑到用户兴趣模型不断更新的特点,我们提出了两种混合推荐模型,加入兴趣遗忘因子,使推荐模型更加适用于微博的特性。

(1) 基于用户的混合推荐模型

该混合推荐模型利用了基于本体描述的阅读材料的相似性和多层用户的相似性。利用由用户兴趣衍生出的潜在社区网络的关系,并结合语义项偏好信息,对协同过滤推荐具有很好的作用。基于上述的多层语义社区网络,我们提出两种推荐模型,它们考虑了生成的社区网络中用户之间的联系和其所处的不同场景,生成了阅读材料的有序排名列表。第一种模型称为HUP,是基于所有用户的语义特征生成唯一一排名列表;第二种模型称为HUP-r,是对每个语义集群 C_r 输出一个排名列表。

HUP

混合推荐系统利用用户 u_i 的特征属性返回唯一的排序列表。这种模型比较了当前用户和其他用户的兴趣,并考虑了他们之间的相似性,衡量了所有用户对不

同阅读材料的兴趣度, 以及随着用户模型的更新, 用户对阅读材料兴趣的锐减度。同时, 还比较了每个概念集群来衡量阅读材料和集群的相似性。我们还结合了两种方法在不同的语义层中推荐阅读材料, 一种是根据阅读材料的特征, 一种是根据用户偏好间的联系。

阅读材料 d_k 的偏好值是每个概念集群中用户与其他用户的相似性的间接偏好加权和, 计算公式如下:

$$pref(d_k, u_r, t) = \sum_r nsim(d_k, C_r) \cdot \frac{\sum_r nsim(u_r, u_i) \cdot sim_r(d_k, u_i) \cdot f(t)}{f(t)} \quad (3)$$

这里: $sim(d_k, C_r) = \frac{\sum_{qj \in C_r} d_{kj}}{|d_k| \sqrt{|C_r|}}$, $nsim(d_k, C_r) = \frac{sim(d_k, C_r)}{\sum_r sim(d_k, C_r)}$ 分别是阅读材料 d_k 和概念集群 C_r 的单一化、归一化的相似性。 $sim_r(u_r, u_i) = \cos(u_i', u_r')$, $nsim_r(u_r, u_i) = \frac{sim_r(u_r, u_i)}{\sum_r sim_r(u_r, u_i)}$ 表示用户 u_i 、 u_r 在社区层 r 中的单一化、归一化相似性。 $sim_r(d_k, u_r) = \cos(d_k', u_r') = \frac{d_k' \cdot u_r'}{|d_k'| \cdot |u_r'|}$ 表示阅读材料 d_k 和用户 u_i 在社区层 r 中的相似性。

HUP-r

混合推荐系统通过比较每个社区网络层用户和阅读材料的相似性, 预测用户的偏好值, 对每个社区返回一个排名列表。排名靠前的是集群中具有最大相似性的用户。计算公式如下所示:

$$pref_r(d_k, u_r, t) = \frac{\sum_r nsim_r(u_r, u_i) \cdot sim_r(d_k, u_i) \cdot f(t)}{f(t)} \quad (4)$$

该计算公式不考虑阅读材料与社区 C_r 的关系, r 是 $sim(u_i, C_r)$ 的最大值。

与HUP模型类似, 该模型利用了用户兴趣之间的关系和用户感兴趣的阅读材料。不同的是该模型在不同的社区层分别进行推荐。如果当前语义社区能够较好的识别某个阅读材料, 与整体推荐模型相比, 该模型会得到更好的精度召回率。

(2) 基于项目的混合推荐模型

该模型忽略用户的兴趣特征, 适用于没有兴趣偏好的新用户或用户的兴趣偏好过于笼统, 不利于用户与社区网络比较的情况。

HIP

混合推荐系统忽略用户的特征属性并返回唯一的有序列表。阅读材料 d_k 的排名由它与社区的相似性、它与每个社区中用户特征的相似性确定。由于当前用户与其他用户之间没有联系, 每个用户特征的影响因子由用户数量 M 平均。

$$pref(d_k, t) = \frac{1}{M} \sum_r nsim(d_k, C_r) \cdot \frac{\sum_r sim_r(d_k, u_i) \cdot f(t)}{f(t)} \quad (5)$$

该模型适用于当前用户特征属性不明确的情况下, 在不同语义社区层中收集所有用户对阅读材料的满意度。虽然在精度召回率方面不如前面两种模型, 但这个相较于之前需要手动配置用户特征属性, 已经得到了明显的改善。

HIP-r

混合推荐系统忽略用户的特征属性, 每个社区返回一个有序推荐列表, 与社区相似度最接近的用户排在前面。用户和阅读材料语义属性相似性的计算表示如下:

$$pref_r(d_k, t) = \frac{1}{M} \cdot \frac{\sum_r sim_r(d_k, u_i) \cdot f(t)}{f(t)} \quad (6)$$

这个模型是最简单的, 它只衡量了每个社区中最适合用户的阅读材料对象, 代表了基于项目对象的协同过滤系统。

6 实验

文章选择来自新浪微博 (<http://weibo.com>) 和豆瓣的数据集, 验证所提出框架的有效性。在新浪微博数以亿计的注册用户中, 选择上海理工大学图书馆作为数据收集的种子, 通过新浪微博开放的API获取关注用户, 剔除垃圾数据, 只保留上海理工大学图书馆资源对其开放的粉丝数据。由于用户发布的帖子数据形式多样, 在剩余的用户中, 再挑选2000位用户, 且这2000位用户发布的微博帖子原创性文字较多, 更新频率高。抽取其近3个月的微博数据, 进行分词和关键词过滤这两个预处理过程。本文采用中国科学院的汉语词法分析系统ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) 对微博文本进行分词, 其支持中文分词、词性标注等功能, 分词准确度可达98.45%, 满足实验要求。我们根据豆瓣的分类标签, 将主题属性定义为文学、流行、文化、生活、经管、科技, 每个主题属性又包含不同的项目和关键词。

本文利用自然语言处理工具, 抽取微博关键词并度量其权重作为用户的偏好评分, 使用向量空间模型^[22]构建用户偏好评分向量模型, 并计算用户之间的相似关系。表1表示了20组用户(每组100个用户)对不同主题属性的偏好程度。

表2显示了主题属性相关的初始概念, 需要注意的是用户的主题属性不一定包含它下面的所有概念。

表1 不同用户对不同主题属性的兴趣度及其预期聚类

	文学	文化	流行	生活	经管	科技	期刊类别
G ₁	0.774	0.721	0.212	0.167	0.045	0.165	1
G ₂	0.812	0.655	0.134	0.461	0.105	0.227	1
G ₃	0.146	0.124	0.763	0.840	0.242	0.487	2
G ₄	0.087	0.156	0.225	0.195	0.779	0.714	3
G ₅	0.245	0.543	0.892	0.843	0.112	0.168	2
G ₆	0.104	0.194	0.835	0.495	0.527	0.136	2
G ₇	0.677	0.482	0.124	0.118	0.469	0.078	1
G ₈	0.488	0.185	0.217	0.258	0.625	0.834	3
G ₉	0.065	0.105	0.536	0.207	0.765	0.488	3
G ₁₀	0.137	0.055	0.793	0.456	0.148	0.508	2
G ₁₁	0.758	0.563	0.115	0.498	0.206	0.125	1
G ₁₂	0.095	0.235	0.555	0.718	0.164	0.620	2
G ₁₃	0.147	0.208	0.124	0.462	0.429	0.715	3
G ₁₄	0.514	0.558	0.612	0.236	0.068	0.173	1
G ₁₅	0.405	0.522	0.126	0.352	0.095	0.115	1
G ₁₆	0.108	0.125	0.602	0.665	0.193	0.485	2
G ₁₇	0.216	0.179	0.483	0.524	0.324	0.136	2
G ₁₈	0.105	0.456	0.215	0.226	0.535	0.508	3
G ₁₉	0.485	0.578	0.139	0.202	0.524	0.133	1
G ₂₀	0.126	0.483	0.815	0.187	0.125	0.223	2

表2 参考主题属性的初始概念

主题属性	概念
文学	诗人、作家、小说
文化	哲学、寺庙、历史、人文
流行	唱片、音乐节、好莱坞、美剧、摇滚、时尚
生活	美食、旅行、购物、健身、宠物
经管	经济、投资、理财、股票、基金、房地产
科技	互联网、IT、乔布斯、Android、手机、笔记本、相机

接下来，利用这20组用户的特征属性测试所提出方法的有效性。首先，通过CSA策略扩展新的概念，增加概念领域并促进用户聚类。文中采用基于欧式距离计算概念相似度和基于平均连锁法测量聚类相似度的分层策略。表3总结了用户实际分配到的聚类及相应的相似性。实验表明，获取到的结果与表1中列出的期望值是完全重合的，所有的用户均被分配到了与其相关

表3 用户与社区聚类的相似度

社区聚类	用户							
1	G ₁	G ₂	G ₇	G ₁₁	G ₁₄	G ₁₅	G ₁₉	
	0.733	0.568	0.507	0.468	0.408	0.384	0.216	
2	G ₃	G ₅	G ₆	G ₁₀	G ₁₂	G ₁₆	G ₁₇	G ₂₀
	0.746	0.722	0.604	0.530	0.425	0.427	0.364	0.234
3	G ₄	G ₈	G ₉	G ₁₃	G ₁₈			
	0.728	0.663	0.484	0.376	0.263			

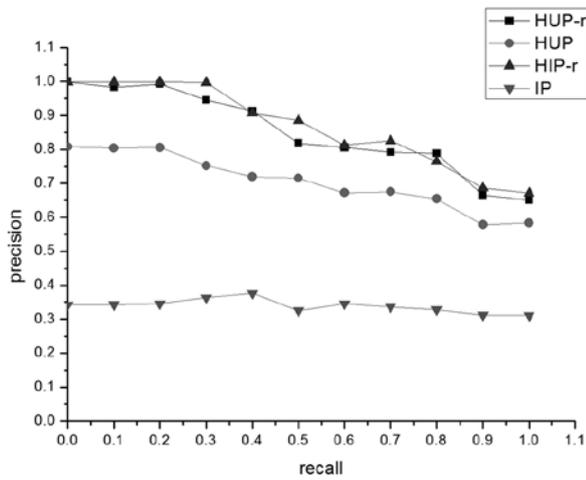
的聚类中，用户的相似性值反映了它们与每个聚类的相似度。

概念形成之后，每组用户被划分到特定的社区，利用信息检索模型获得了与社区聚类相关的概念内容，并将其分组，如表4所示。从表中可以看出，获得的大多数概念内容不包含在最初的用户概念领域中，但对聚类的创建有很大的帮助。

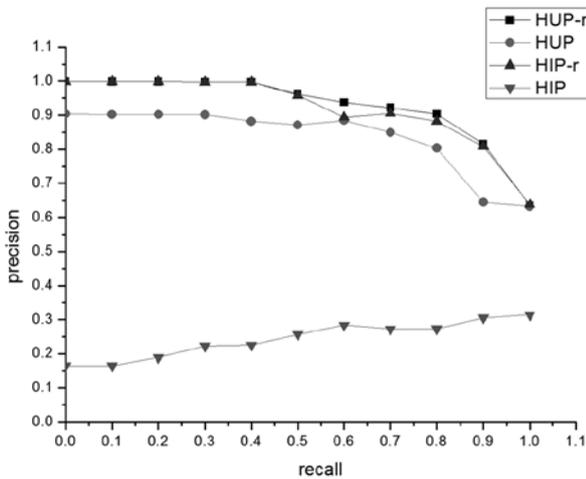
表4 聚类及所属概念领域

主题属性	概念
1	文学：作家、韩寒、三毛、莫言、萧红、海子；小说、三体、围城、百年孤独、活着、平凡的世界
	文化：哲学、不能承受的生命之轻、人类简史；寺庙、佛教、拉萨、一步一如来；历史、明朝那些事、三国演义、中国历代政治得失、大英博物馆世界简史、木已成舟
2	流行：唱片、摇滚、音乐、滚石30、CD、音乐节；钢琴、莫扎特、贝多芬；好莱坞、电影、在黑暗中醒来、奥黛丽赫本、梦露；美剧、权利的游戏、生活大爆炸、福尔摩斯、柯南
	生活：美食、吃货、火锅、甜品；旅行、香港、台湾、北京、上海、日本、韩国、清迈、欧洲；购物、代购、闲置、二手、淘宝、买；健身、跑步、减肥、瑜伽、游泳；宠物、猫、狗、花
3	经管：信用卡、钱、富爸爸穷爸爸、货币战争、创业、伟大的博弈、MBA、巴菲特
	科技：互联网、硅谷、IT；手机、苹果、小米、魅族、华为、三星、Ios、Android；相机、单反、佳能、尼康； mac、联想；乔布斯、雷军、老罗、浪潮之巅

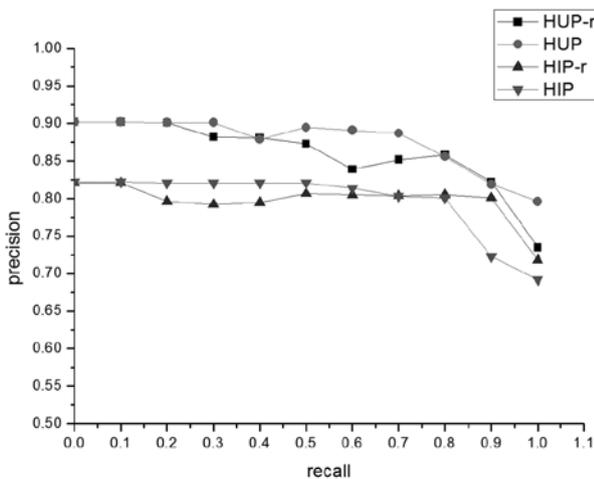
通过计算三个社区聚类中的用户的平均精确度/召回度曲线，我们总结出了四种不同的模型，如图6所示。从实验结果中可以看出：根据不同社区返回特定推荐列表模型（HUP-r和HIP-r）的精确率和召回率都优于只返回一个推荐列表的模型；利用社区网络中用户关系



(a) 社区1中用户的平均精确率和召回率



(b) 社区2中用户的精确率和召回率



(c) 社区3中用户的精确率和召回率

图6 社区聚类中用户的精确度和召回度

(HUP 和HUP-r) 的模型与没有使用用户特征属性相似度的模型相比, 结果得到了明显的提升。

7 结语

本文提出了基于微博的个性化信息推荐系统, 该系统可以自动识别和适应用户微博信息的变化, 建立基于本体的个性化兴趣模型, 形成多层语义社区网络, 并提出了基于用户的混合过滤推荐模型和基于项目的混合过滤推荐模型, 每种推荐模型分别有两种预测方法, 对用户进行阅读材料的智能推荐。最后, 通过收集新浪微博数据和豆瓣数据, 对所提出的系统进行了有效性的验证。实验表明, 该推荐系统可以自动识别用户的兴趣特征, 用户不需要手动标注其特征。此外, 系统能够自动适应用户兴趣的变化, 及时调整个性化推荐模型。

研究结果对数字化图书馆的发展有启示意义: (1) 图书馆应重视新兴媒体对读者的影响, 与新媒体融合, 让两者相互激发, 利用微博数据提高对用户兴趣偏好的自动化识别; (2) 图书馆作为海量信息的提供者, 应当依据用户兴趣偏好的变化, 构建针对用户的偏好空间, 提供个性化的服务。未来的研究工作将从动态视角处理用户社区网络关系, 而不是作为静态网络, 进一步提高个性化推荐的多样性和效率。

参考文献

- [1] Ioannidis Y, Kourtrika G. Digital Library Information- Technology Infrastructure [J]. INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES, 2005, 5(4):266.
- [2] 张富国,徐升华.基于信任的电子商务推荐多样性研究 [J]. 情报学报,2010, 29(2):350-355.
- [3] Zhang Fuzhi,Bai Long,Gao Feng.A User Trust-Based Collaborative Filtering Recommendation Algorithm[J].LECTURE NOTES IN COMPUTER SCIENCE, 2009(5927):411-424.
- [4] Bellogin A,Cantador I,Pablo C. A Study of Heterogeneity in Recommendations for a Social Music Service [C]. PROCEEDINGS OF THE 1st INTERNATIONAL WORKSHOP ON INFORMATION HETEROGENEITY AND FUSION IN RECOMMENDER SYSTEMS. New York: ACM,2010:1-8.
- [5] Phelan O, McCarthy K, Bennett M, et al. Terms of a Feather: Content-Based News Recommendation and Discovery Using Twitter [J].

- LECTRUE NOTES IN COMPUTER SCIENCE, 2011(6611):448-459.
- [6] Su Mon Kywe, Ee-Peng Lim, FeidaZhu. A Survey of Recommender Systems in Twitter [J]. LECTRUE NOTES IN COMPUTER SCIENCE, 2012(7710):420-433.
- [7] 曾琦. 基于微博的图书馆资源推荐系统设计[J]. 图书馆学研究, 2012(14):25-28.
- [8] 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.
- [9] Vallet D, Castells P, Fernández M, et al. Personalized Content Retrieval in Context Using Ontological Knowledge[C]. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on "The Convergence of Knowledge Engineering, Semantics and Signal Processing in Audiovisual Information Retrieval", 2007, 17(3):336-346.
- [10] Vallet D, Mylonas P, Corella M A, et al. A Semantically-Enhanced Personalization Framework for Knowledge-Driven Media Services[C]. In Proceedings of IADIS International Conference on WWW / Internet (ICWI 2005), Lisbon, Portugal, 2005.
- [11] Hawalah A, Fasli M. Using User Personalized Ontological Profile to Infer Semantic Knowledge for Personalized Recommendation [J]. LECTRUE NOTES IN BUSINESS INFORMATION PROCESSING, 2011(85):282-295
- [12] Cohen P R, Kjeldsen R. Information Retrieval by Constrained Spreading Activation in Semantic Networks[J]. INFORMATION PROCESSING AND MANAGEMENT, 1987, 23(2):255-268.
- [13] Crestani F, Lee P L. Searching the Web by Constrained Spreading Activation [J]. INFORMATION PROCESSING & MANAGEMENT, 2000, 36(4):585-605.
- [14] Liu H, Maes P, Davenport G. Unraveling the Taste Fabric of Social Networks [J]. INTERNATIONAL JOURNAL ON SEMANTIC WEB AND INFORMATION SYSTEMS, 2006, 2 (1):42-71.
- [15] Balabanovic M, Shoham Y. Content-Based Collaborative Recommendation [J]. COMMUNICATIONS OF THE ACM, 1997, 40(3):66-72.
- [16] Linden G, Smith B, York J. Amazon.com Recommendations: Item-to-Item Collaborative Filtering[J]. IEEE INTERNET COMPUTING, 2003, 7(1):76-80.
- [17] Montaner M, López B, Lluís de la Rosa, J. Taxonomy of Recommender Agents on the Internet[J]. ARTIFICIAL INTELLIGENCE REVIEW, 2003, 19(4): 285-330.
- [18] Sarwar B M, et al. Item-Based Collaborative Filtering Recommendation Algorithms[C]. WWW '01 Proceedings of the 10th international conference on World Wide Web, 2001:285-295.
- [19] Duda R O, Hart P, Stork D G. Pattern Classification[M]. Wiley-Interscience, 2001.
- [20] Ungar L, Foster D. Clustering Methods for Collaborative Filtering[C]. Proceedings of the Workshop on Recommendation Systems at the 15th National Conference on Artificial Intelligence, 1998.
- [21] 蒋萍, 崔志明. 智能搜索引擎中用户兴趣模型分析与研究[J]. 微电子学与计算机, 2004, 21(11):24-26.
- [22] Salton G, Wong A, Yang C S. A Vector Space Model for Automatic Indexing[J]. COMMUNICATIONS OF THE ACM, 1975, 18(11):613-620.

作者简介

黄婷, 女, 1987年生, 研究方向: 信息检索、数据挖掘, E-mail: huangtingsst@163.com。

Microblog-based Personalized Recommendation for Intelligent Digital Library

HUANG Ting

(University of Shanghai for Science and Technology Library, Shanghai 200093, China)

Abstract: As the traditional library can't provide accurate and diverse personalized service, we propose a personalized information recommendation model based on micro-blog. By mining microblog users' text messages and decrypting their common preferences of concepts space based on ontology, we build a multilayered semantic social network model. The applicability of the proposed model to a hybrid collaborative filtering system is empirically suitable to be used in personalized recommendation of intelligent digital library.

Keywords: Microblog; Ontology Concept; Multilayered Semantic Social Network; Intelligent Digital Library; Personalized Recommendation

(收稿日期: 2015-09-09)