

# 基于网络读书社区阅读行为的 书目推荐模型研究\*

邹婧琳 范炜

(四川大学公共管理学院, 成都 610064)

**摘要:** 书目推荐旨在建立图书与读者之间的双向连接。新兴网络读书社区中用户的参与互动过程积累了大量用户行为数据, 为书目推荐提供了丰富的数据依据。文章针对网络读书社区中的主要用户阅读行为, 综合考虑用户评分、阅读状态、标签、时间信息对用户阅读偏好的体现, 提出用户阅读偏好程度和偏好方向相结合的用户偏好模型, 形成了个性化书目推荐。采集豆瓣读书数据进行模型检验, 结果表明, 该模型有效地提高书目推荐的准确性和多样性, 对网络读书社区实施个性化书目推荐具有参考价值。

**关键词:** 网络读书社区; 阅读行为; 书目推荐

**中图分类号:** G252

**DOI:** 10.3772/j.issn.1673—2286.2014.04.007

阅读是中国自古以来的优良传统, 爱读书、会读书已成为当代公民基本文化素养的重要组成部分。随着图书出版行业的繁荣发展与互联网信息传播的快速便捷, 读者在面对海量的图书资源时陷入了选择的困境。如何帮助读者选择适合他们的图书, 并关注他们的个性化阅读需求, 有效的引导和推荐手段必不可少。

图书情报领域的书目推荐又称书目导读, 是引导和影响读者阅读行为的重要服务形式。传统的图书馆书目推荐服务主要借助于读者的借阅记录、书目榜单等信息, 不能很好地反映读者真实的阅读意愿。新兴的网络读书社区是由用户主导的读书交流和知识发现的虚拟社区之一, 其中所有的内容、分类、筛选, 都由用户产生和决定。这些阅读行为数据是用户贡献内容(UGC)的一个重要类型, 也为书目推荐提供了客观的数据依据。

本文关注网络读书社区中的主要阅读行为, 从用户阅读偏好程度和偏好方向两个角度构建用户偏好模型, 继而形成书目推荐。

## 1 书目推荐现状分析

书目推荐可视为现代推荐系统的一种类型。现代

推荐系统通过用户的历史行为数据来预测用户的需求, 为特定用户推荐针对性的信息资源<sup>[1]</sup>。推荐系统在互联网服务业得到广泛应用, 关注个性化小众需求带来利基经济效益, 长尾理论给出了最好的诠释。以电子商务网站Amazon为例, 它为我们展现了丰富的推荐服务形式, 其中基于用户的购买和浏览记录, Amazon会为其推荐购买了此商品的用户也经常或同时购买的其他商品, 及浏览了此商品的其他用户还查看或购买过的商品。以在线影片租赁网站Netflix为例, 它根据用户的评分和反馈, 为用户推荐相似的电影, 其宣称有60%的用户能通过其推荐找到自己感兴趣的电影视频。

随着数字图书馆的建设发展, 现代推荐系统融入书目导读服务之中。基于内容特征的书目推荐, 斯坦福大学的Fab<sup>[2]</sup>、麻省理工学院的Letizia<sup>[3]</sup>等系统利用资源内容进行信息过滤, 自动为用户提供与检索内容相似的资源; 基于读者基本信息的书目推荐, 中国人民大学的KingBase DL系统<sup>[4]</sup>通过用户的注册信息, 如专业、研究方向等, 为用户提供推荐服务; 基于读者的借阅历史信息的书目推荐, 赵麟<sup>[5]</sup>利用图书馆用户的借阅数据分析用户的行为模式, 基于最大频繁模式挖掘算法进

\*本研究得到四川大学大学生创新创业训练计划项目“大学生书目导读与互动推荐系统”(编号: 20131241)资助。

行书目的关联推荐。景民昌等<sup>[6]</sup>利用图书借阅时间构建用户兴趣模型,并基于协同过滤算法的推荐实现说明了模型的有效性。从现有研究看,图书馆的书目推荐系统,对用户的借阅历史记录数据的挖掘进行了一定的探索,但图书馆资源的低利用率和用户行为的稀疏性对其推荐造成了一定的局限性。

以Web2.0为特征的网络读书社区,如豆瓣读书<sup>[7]</sup>、LibraryThing<sup>[8]</sup>、Goodreaders<sup>[9]</sup>等,充分利用大众参与,带来了丰富的用户阅读行为数据,如评分、打标签、分享、评论等。针对社交网络中丰富的用户行为数据,Nan Zheng等<sup>[10]</sup>在社会化标签系统中融合标签和时间信息来预测用户偏好,并实验证实了标签和时间信息的结合能提高推荐的准确性。Zi-Ke Zhang等<sup>[11]</sup>针对“用户-项目”和“项目-标签”之间的信息关联提出基于扩散的资源分配推荐策略,并实验证明了该方法可以明显提升推荐结果的准确性、多样性和惊喜性。Heung-Nam Kim等<sup>[12]</sup>综合评分和标签信息进行用户主题发现,并实验证明其方法在冷启动和准确性方面的推荐效果。

从现有研究看,学者们主要针对网页、电影、论文等领域的用户行为数据进行混合推荐的探索,并在一定程度上提升了推荐的准确性,这为书目推荐提供了一种重要的研究思路。网络读书社区的兴起及其API的数据开放性,使用户阅读行为数据更加容易获取。因此,本文利用网络读书社区中的用户行为数据,采用混合推荐的思路,对多个推荐要素进行集成,构建用户偏好模型,来提升书目推荐效果。

## 2 网络读书社区书目推荐模型

网络读书社区中,人们通过集体智慧(Collective Intelligence)来获得满足他们兴趣的图书推荐。集体智慧主要是指为创造新的想法,将一群人的行为、偏好或思想组合在一起,从中对数据进行搜索、组合和分析<sup>[13]</sup>。根据用户的交互行为,一般通过用户的显性反馈行为和隐性反馈行为来预测用户的阅读兴趣偏好。显性反馈行为是指用户明确表示对物品喜好的行为,在网络读书社区中的主要方式就是对图书进行评分、收藏、打标签、评论等。隐性反馈行为与其相对应,是指那些不能明确反应用户喜好的行为,在网络读书社区中较为典型的是浏览、点击等行为。相比于隐性反馈行为,显性反馈行为产生的数据量虽然较少,但能明确反映用户兴趣,并具有实时读取性和正负反馈性。

用户评分、阅读状态、标签信息、时间信息是四种常见的显性反馈行为数据,能够较为明显地体现用户的阅读兴趣。

用户评分是获得用户阅读兴趣最直接的方法,一般采用五分制反馈量表,从“非常不喜欢”到“非常喜欢”,分别对应分值1-5。通过调节评分尺度,从不同的粒度刻画用户对图书的偏好程度,但这种显式评分需要用户额外投入,获得的可用图书评分信息可能比较少。

阅读状态主要分为“想读、正在读、已读”三种,反映用户当前所处的阅读阶段和整体的阅读状况,可以将此数据作为用户对图书的隐式评分。根据阅读状态的不同,对各状态赋予不同的权值,将此数据与评分数据综合,可在一定程度上缓解数据稀疏性的问题。

标签是一种无层次化结构的、用来描述信息的自然语词。用户对图书进行打标签的行为时,一方面表达了用户的阅读兴趣,另一方面也表达了对图书的语义理解。标签作为一种重要的多维特征表现形式<sup>[14]</sup>,联系着读者与图书。通过标签能较全面了解用户多方面的阅读兴趣,并且标签使用的频繁程度也能反映用户对其相关领域图书的感兴趣程度,而用户也更倾向于关注标注了其个人常用标签的图书。

时间信息对书目推荐而言也是一种非常重要的上下文信息。用户的阅读兴趣不是一成不变的,随着时间的迁移,用户所处年龄段、身份等情景的变化会影响其当前的阅读兴趣与图书选择;而且图书本身也具有其生命周期。一般而言,用户近期的阅读行为能反映出其当前的主要阅读兴趣,而当前的阅读兴趣比很久之前的阅读兴趣更值得关注,通过时间信息来了解用户阅读兴趣的阶段变化,能保证实时性的推荐图书与用户阅读兴趣的一致性。

用户评分和阅读状态信息侧重于表达用户对书籍的感兴趣程度,标签信息侧重于表达用户具体的阅读兴趣方向,而时间信息则在上述两个方面体现了用户阅读兴趣的一种迁移,这四种阅读行为数据为预测用户阅读偏好提供了数据基础。用户阅读偏好的预测主要从深度和广度两个方面进行考虑。阅读偏好的深度体现用户对图书感兴趣的程度,即用户偏好程度;阅读偏好的广度体现用户阅读图书的多样性,即用户偏好方向。

因此,本文从用户阅读偏好程度和偏好方向两个角度,将评分、阅读状态、标签和时间四种阅读行为数据作为书目推荐要素,形成集成推荐模式,如图1所示。其中,将用户评分和阅读状态信息进行处理,统一其度

量基准, 获得用户的阅读偏好程度; 对标签信息进行处理, 获得用户的阅读偏好方向; 将时间因子分别引入, 从而获得最终的用户偏好程度指标和用户偏好方向指标。对两个指标进行加权处理, 获得联系用户与图书之间关系的总偏好值, 构成用户偏好模型。

(1) 用户偏好程度指标集成

设用户对图书的原始评分 $r_{u,i}$ , 用户的阅读状态为 $S_{u,i}$ 。用户对图书的原始评分 $r_{u,i}$ , 采用五分制反馈量表, 取值范围为1-5的整数。一般将低于3分的评分值视为负向评分, 表示用户不喜欢此图书。因此, 将原始评分数据分为正负反馈数据, 以原始区间「1,2,3,4,5」中的分值3为中心映射到以分值1为中心、以 $t$ 为间距的区间「 $1-2t, 1-t, 1, 1+t, 1+2t$ 」上, 变换后的评分值用 $r'_{u,i}$ 表示。

用户的阅读状态 $S_{u,i}$ 分为“想读(wish), 在读(reading), 已读(read)”三种, 阅读状态的选定相当于用户对图书的隐性评分, 且属于正反馈数据。考虑用户评分与阅读状态的关联, 对不同的阅读状态赋予相应的权值来表现用户的阅读偏好, 通过将两个度量指标统一在同一个基准上讨论。不同阅读状态的取值函数 $y_{u,i}$ 表示为

$$y = \begin{cases} 1ne, & \text{when } S_{u,i} = wish; \\ \ln(t+1), & \text{when } S_{u,i} = read; \\ \ln(2t+1), & \text{when } S_{u,i} = reading. \end{cases} \quad (1)$$

用户处于“想读”状态, 对图书本身并不是十分了解, 此时不能给图书进行评分。该状态只是表达其模糊的阅读意愿, 并没有明确体现偏好, 对用户的阅读偏好程度影响较小, 但它的正向反馈性可作为评分数据的补充。用户“正在读”或“已读完”的状态, 表明用户已经将书籍与自身的兴趣进行了匹配, 结合评分数据在一定程度上能反映用户的阅读偏好程度。而“正在读”状态最能体现用户当前的阅读偏好, “已读完”状态次之。根据当前的评分量表机制,  $t$ 取值最好为5以内的正整数, 本文取 $t=2$ 。

时间信息体现用户阅读兴趣的动态变化, 也是用户历史行为数据可靠性的指标。根据用户阅读兴趣的衰减性确定其生命周期, 时间衰减值 $f(t_{u,i})$ 服从指数衰减, 可表示为

$$f(t_{u,i}) = \exp\{-\alpha(T-t_{u,i})\} \quad (2)$$

其中:  $T$ 为系统给定的近期时间,  $t_{u,i}$ 为用户对图书产生阅读行为的时间,  $\alpha$ 为时间衰减参数。根据不同读书社区中的具体情况对 $\alpha$ 进行取值。 $\alpha$ 取较大的值时, 默认该群体中用户的阅读兴趣变化较快, 反之 $\alpha$ 取较小的值。

将用户评分和阅读状态两个指标进行集成, 并引入时间衰减值, 最终用户偏好程度指标 $r''_{u,i}$ 表示为

$$r''_{u,i} = (r'_{u,i} + y_{u,i}) \cdot f(t_{u,i}) \quad (3)$$

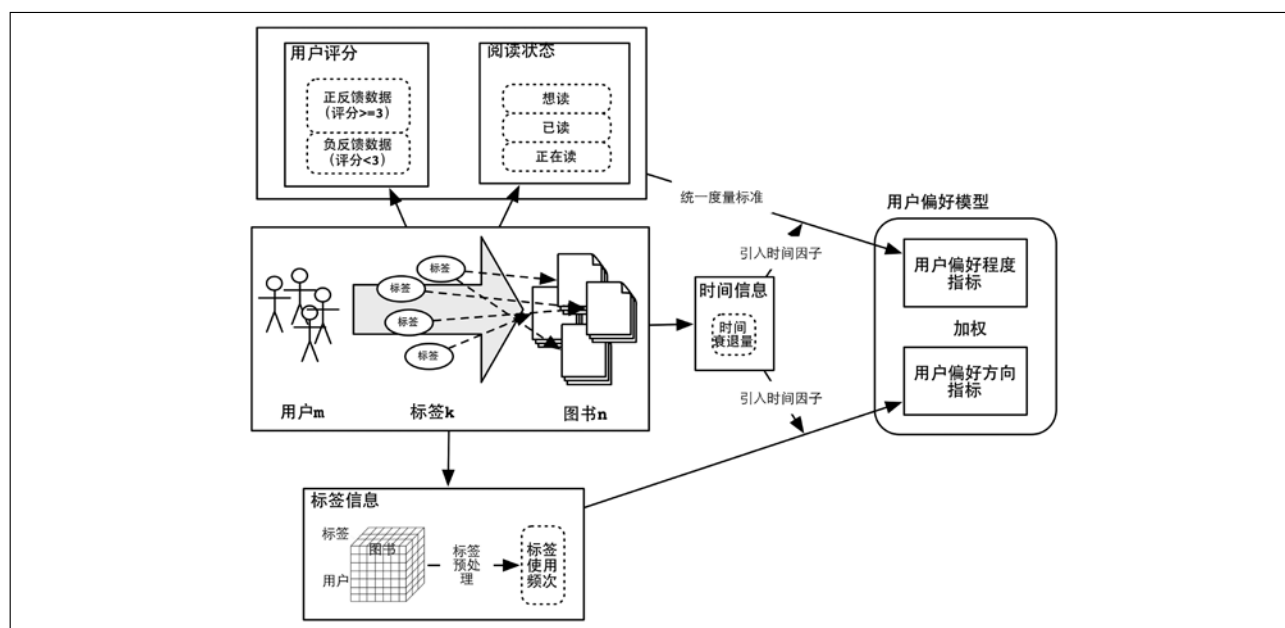


图1 四种阅读行为数据集成模式

(2) 用户偏好方向指标集成

标签信息是对图书内容的多角度揭示,也是对用户阅读兴趣偏好细分的体现。利用标签在用户和图书之间的使用次数,并借鉴TF-IDF (term frequency-inverse document frequency) 的思想,对热门标签和热门图书进行了一定程度的修正,防止其对推荐图书新颖性的影响,可获得用户阅读偏好方向上的预测值 $d_{u,i}$ ,表示为

$$d_{u,i} = \sum_{g \in G(u) \cap G(i)} \frac{n_{u,g}}{\log(1+n_g^{(u)})} \cdot \frac{n_{i,g}}{\log(1+n_i^{(u)})} \quad (4)$$

其中:  $G(u)$ 为用户 $u$ 使用过的标签集合,  $G(i)$ 为图书 $i$ 被标注过的标签集合,  $n_{u,g}$ 为用户使用标签 $g$ 的次数,  $n_{i,g}$ 为图书 $i$ 被标签 $g$ 所标注的次数,  $n_g^{(u)}$ 为使用过标签 $g$ 的用户人数,  $n_i^{(u)}$ 为对图书 $i$ 进行过标注行为的用户人数。

一般而言,用户当前的阅读偏好与近期的兴趣方向关系更大,因此要考虑到时间信息对用户阅读偏好方向的影响,在 $d_{u,i}$ 中引入时间的衰减函数 $f(t_{u,i})$ ,最终的用户偏好方向指标 $p_{u,i}$ 表示为

$$p_{u,i} = d_{u,i} \cdot f(t_{u,i}) \quad (5)$$

(3) 总偏好值计算

综合考虑将用户偏好程度指标 $r''_{u,i}$ 和用户偏好方向指标 $p_{u,i}$ 进行整合,获得用户 $u$ 对图书 $i$ 的总偏好值 $A_{u,i}$ ,可表示为

$$A_{u,i} = \lambda \cdot r''_{u,i} + (1-\lambda) \cdot p_{u,i} \quad (6)$$

其中:  $\lambda \in [0,1]$   $\lambda$ 为调节因子,用于调整 $r''_{u,i}$ 和 $p_{u,i}$ 之间的权重,其具体取值根据网络读书社区的具体运行情况而定。通过调整 $\lambda$ 值,可使总偏好值发挥出最佳的推荐效果。

(4) 书目推荐形成

协同过滤算法是推荐系统中运用最广泛的技术之一,它的基本思想是利用用户-物品评分矩阵计算相似度,基于相似用户或相似物品形成推荐<sup>[15]</sup>。本文采用基于用户的协同过滤算法,利用用户 $u$ 对图书 $i$ 的总偏好值,形成针对不同用户的个性化书目推荐。

a. 确定邻居集

设所有用户的集合为 $U = \{u_1, u_2, \dots, u_m\}$ ,所有物品的集合为 $I = \{i_1, i_2, \dots, i_n\}$ ,用户-图书偏好矩阵为

$$R_{m \times n} = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix},$$

若用户对图书无阅读行为操作,则 $A_{i,j}=0$ ,否则为用户偏好模型所得的总偏好值。

根据给定的矩阵 $R_{m \times n}$ ,计算用户之间的相似度,来发现用户之间阅读兴趣的相似性。推荐系统中常用的相似度计算方法有Jaccard系数、余弦相似度、Pearson相关系数等。本文采用余弦相似度度量指标,用户 $u$ 与用户 $v$ 之间的相似度 $Sim(u,v)$ 表示为

$$Sim(u,v) = \frac{\sum_{i \in I} A_{u,i} \cdot A_{v,i}}{\sqrt{\sum_{i \in I} A_{u,i}^2} \sqrt{\sum_{i \in I} A_{v,i}^2}} \quad (7)$$

$Sim(u,v)$ 值越大,表示两个用户的阅读兴趣偏好越相近。根据 $Sim(u,v)$ 值确定与目标用户阅读兴趣相似的前 $K$ 个邻居,构成目标用户的邻居集 $S(u,k)$ 。

b. 形成书目推荐列表

得到与目标用户阅读兴趣相似的邻居集后,计算目标用户对该集中的用户感兴趣的图书的偏好程度,通过过滤、排名形成推荐。用户对其邻居喜欢图书的偏好程度 $pred(u,i)$ 表示为

$$pred(u,i) = \sum_{v \in S(u,k) \cap U(i)} Sim(u,v) \cdot A_{v,i} \quad (8)$$

其中:  $S(u,k)$ 表示与用户 $u$ 阅读兴趣最相似的前 $K$ 个用户,  $U(i)$ 表示对图书 $i$ 有过阅读行为的用户集合。对 $pred(u,i)$ 值进行排序,向目标用户推荐其没有阅读行为且 $pred(u,i)$ 值高的前 $n$ 本图书。

### 3 书目推荐模型检验

(1) 实验数据集获取

实验数据采集自豆瓣网。豆瓣网是一个以生活、文化为内容的创新网络服务平台,它的核心用户群是具有良好教育背景的都市青年,2012年月度其覆盖用户数已超过1亿。在“豆瓣读书”中,用户可以给图书打分、写评价、为图书添加标签分类等,所有的内容都通过用户参与产生。同时,其提供的豆瓣图书API能以用户或图书为中心获取相应的用户行为数据。较大的用户基础与活跃的用户参与使其数据具有一定的可行性和实验

的客观性。

通过调用豆瓣图书API,以用户为参数,采集用户对图书产生的行为数据。每条记录包括用户ID、图书ID、用户评分、阅读状态、标签和时间六个数据项。原始数据集包括了500个用户对5873本图书的阅读行为数据。其中,存在行为数据过少的用户及大量同义、近义的标签,为减少这些噪音,对其进行预处理:

- 剔除阅读行为数据过少的用户。保证每个用户至少对5本图书进行过评分,且每个用户对其所有标注过的图书至少使用过10个标签。

- 统一标签中相同语词的繁简形式。比如将标签“東野圭吾”、“东野圭吾”统一为简写形式。

- 统一标签中因词义或分隔符造成的同义、近义标签。比如将“数据挖掘”、“DM”等标签统一表示为“数据挖掘”;将“夏洛蒂·勃朗特”、“夏洛蒂.勃朗特”、“夏洛蒂勃朗特”等标签统一为“夏洛蒂·勃朗特”。

经处理后,最终的实验数据集包括120个用户对867本图书的2488条阅读行为数据记录。实验数据片断如表1所示。

(2) 实验分析及结果

以用户为分割标准,采用随机化抽样的方法将实验数据集均匀分为10份。轮流将其中9份数据作为训练集,其中1份作为测试集。在训练集上训练用户偏好模型,在测试集上对用户行为进行预测,统计出相应的离线实验评测指标。将10次实验的平均值作为最终的评测结果,通过评测指标值来判断推荐效果。

评测推荐结果的优劣有很多指标。本文选用预测

准确率、多样性作为评价依据。预测准确率采用文献[16]提出的F1指标综合准确率和召回率进行度量。准确率是指推荐列表中用户喜欢的图书与测试集中推荐的所有图书的比率;召回率是指推荐列表中用户喜欢的图书与测试集中用户喜欢的所有图书的比率。多样性主要是评测用户内部的多样性<sup>[17]</sup>,即推荐书目列表对用户阅读兴趣的覆盖面的大小,通过推荐列表中图书之间的相似度来度量。

实验中,对每组数据集中的每个用户推荐前10本图书。考虑到用户偏好模型中的要素集成调节因子 $\lambda$ 和用户的邻居数K可能会对推荐效果产生影响,实验中,在5—100的区间中以5为间距调整不同的K值,在0—1的区间中以0.1为间距调整不同的 $\lambda$ 值,分别进行实验。为验证本文书目推荐模型的有效性,将本文的要素集成推荐模型与传统基于评分的推荐模型进行实验比较。先设定 $\lambda$ 为0.5,调整邻居数量,实验结果如图2所示。

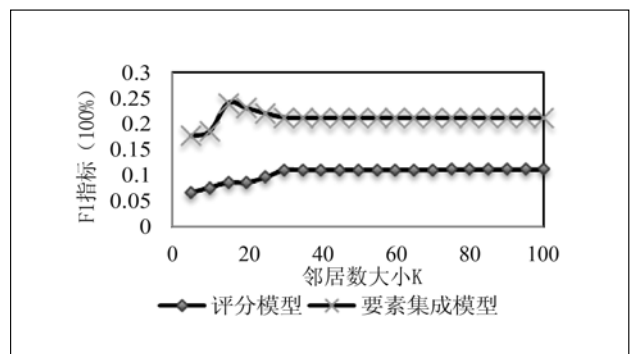


图2 预测准确率结果对比

表1 用户阅读行为数据片断

	2250587	2980812	10598170	1262986	6398075
	一个人的好天气	踮脚张望的时光	给女儿的17封告别信	我不是教你诈	一转念
53323716	猫,青山七惠, 日本文学, 4, read, 2012.4.13	小说,寂地,3, read,2012.4.2	0,read, 2013.4.23,		思维,杂文, 经济学,4,read, 2013.7.24
2658645		寂地,绘本, 4, read,2008.7.4		刘墉,处世,智慧, 励志,4,read, 2008.9.25	
3499806	日本文学,一个人的好天气,4,read 2011.11.11		阿富汗,传记, 0,wish,2012.5.25 0,read,2011.8.25	刘墉,台湾, 智慧,5,read, 2010.3.18	
1982297		绘本,3,read, 2011.12.25			经济学,经济 随笔,0,wish, 2011.7.14

从图2的显示结果可以看出,对于不同数量的邻居,评分推荐模型与本文要素集成推荐模型的预测准确率的变化趋势大体相同。但本文要素集成推荐模型所对应的预测准确度相比于利用传统的用户评分的推荐模型有了明显的提高。F1值越大表明推荐的效果越好。F1值的提高,说明在传统的评分数据中,通过集成用户的阅读状态、标签和时间信息,能够更准确地把握用户的阅读兴趣偏好。

此外,实验结果也显示,当邻居数目小于30时,F1值波动较大,推荐效果受到了邻居数量的影响。而当邻居数目大于30时,F1值趋于稳定。由此可知,在邻居数足够多的情况下,与目标用户有着相似阅读兴趣的用户群体对推荐结果有一定的影响。本文取邻居数为 $K=30$ ,来调整用户偏好模型中的要素集成调节因子 $\lambda$ ,观测预测准确度和多样性的变化情况。实验结果如图3、图4所示。

图3、图4的实验结果表明,对本文用户偏好模型中要素集成的权重进行调节,推荐结果的预测准确性和多样性都受到影响。在 $\lambda$ 取值为0.5-0.8之间时,推荐效果较佳;取值在0.2-0.4之间时,推荐效果较差;单独考虑用户阅读的偏好程度或方向(取值为0或1),推荐效果适中。但从整体来看,预测准确率都高于图2中所示的评分模型所得的预测准确度。 $\lambda$ 所对应的F1值的变化分布与实验数据集的特征及实际运行需求关联较大,就本文实验数据集而言,在 $\lambda$ 取值为0.7左右能获得较为准确的推荐效果。在实际的运用情况下,可根据不同的读书社区的状况进行调整。 $\lambda$ 值使推荐结果的多样性有小幅度的波动,但基本上能在0.6左右。该值越大,表明推荐列表的多样性越好,用户发现自己感兴趣图书的概率越大。本实验的多样性评测值说明书目推荐列表在一定程度上满足用户广泛的阅读兴趣,但是否考虑到用户的主要兴趣,还有待于通过实际的用户问卷调查来观测用户的满意度进行证明。

以用户1255061为例,为其推荐前10本图书,结果如表2所示。

综合以上实验结果可知,将四种阅读行为数据作

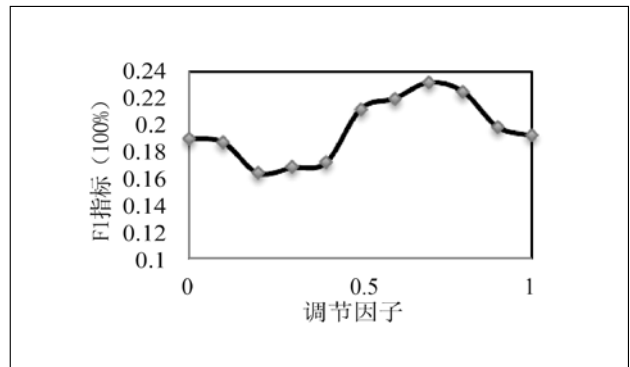


图3 预测准确度变化情况

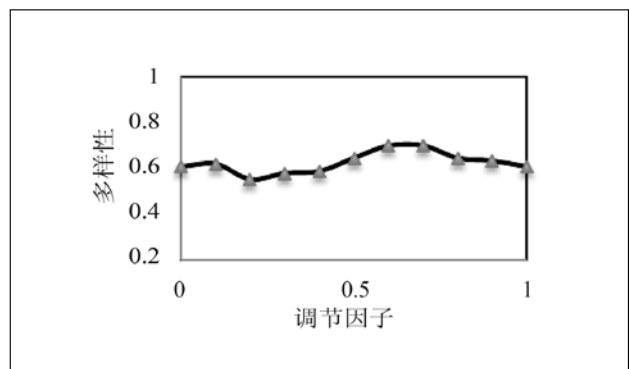


图4 多样性变化情况

为推荐要素进行集成,构建用户偏好模型,能在一定程度上较为准确地把握用户多样化的阅读偏好,并形成针对不同用户的个性化书目推荐列表。

## 4 结语

相较于传统的图书馆书目推荐形式,虚拟读书社区的多交互性,丰富了用户的阅读行为数据,带来了更好的书目推荐的可能。本文综合考虑用户评分、阅读状态、标签和时间信息对用户阅读偏好的影响,从用户偏好程度和用户偏好方向两个角度出发,考虑阅读兴趣的时间偏移状况,对四个阅读行为数据进行处理,构建书目推荐集成模型。通过实验证明该用户偏好模型在一

表2 用户1255061 TOP10推荐图书列表

图书ID	6710437	1205054	1829226	6388661	20427187	1433640	1102852	4238362	6710437	1786670
图书名称	撒哈拉的故事	悲惨世界	肖申克的救赎	这些人,那些事	看见	猜猜我有多爱你	新编日语	活着	送你一颗子弹	百年孤独

定程度上能提高推荐结果的准确性,并能较为广泛地覆盖用户的阅读兴趣。该推荐模型对于当前网络读书社区有一定的分析应用性,有助于把握用户阅读兴趣的迁移,发现用户当前主要的阅读兴趣,形成较准确的书目推荐。

由于豆瓣API数据获取存在一定技术限制,实验数据集相对较小。随着用户参与度的提升,不同维度与层次的更多阅读行为数据集成,未来进一步探讨体现用户阅读兴趣的多种深层阅读行为,有助于扩充和修正书目推荐模型,更准确地把握用户的阅读兴趣。

### 参考文献

- [1] SHAPIRA B. Recommender systems handbook [M]. Springer, 2011.
- [2] BALABANOVIC M, SHOHAM Y. Fab: content-based, collaborative recommendation [J]. Communications of the ACM, 1997, 40(3): 66-72.
- [3] LIEBERMAN H. Letizia: An agent that assists web browsing [J]. IJCAI, 1995(1): 924-929.
- [4] 中国人民大学数字图书馆个性化服务系统[DB/OL]. [2010-03-25]. <http://202.12.18.49/>.
- [5] 赵麟.基于最大频繁模式挖掘算法进行书目推荐系统的设计与实现[J].现代图书情报技术,2010(5):23-28.
- [6] 景民昌,于迎辉.基于借阅时间评分的协同图书推荐模型与应用[J].图书情报工作,2012,56(3):117-120.
- [7] 豆瓣读书[EB/OL]. [2013-11-24]. <http://book.douban.com>.
- [8] LibraryThing [EB/OL]. [2013-11-24]. <http://www.librarything.com/>.
- [9] Goodreads [EB/OL]. [2013-11-24]. <https://www.goodreads.com/>.
- [10] ZHENG N, LI Q. A recommender system based on tag and time information for social tagging systems [J]. Expert Systems with Applications, 2011, 38(4): 4575-4587.
- [11] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs [J]. Physica A: Statistical Mechanics and its Applications, 2010, 389(1): 179-186.
- [12] KIM H N, ALKHALDI A, EL SADDIK A, et al. Collaborative user modeling with user-generated tags for social recommender systems [J]. Expert Systems with Applications, 2011, 38(7): 8488-8496.
- [13] SEGARAN T. Programming collective intelligence: building smart web 2.0 applications [M]. O'Reilly Media, 2007.
- [14] VIG J, SEN S, RIEDL J. Tagsplanations: explaining recommendations using tags [C]// Proceedings of the 14th international conference on Intelligent user interfaces. ACM, 2009: 47-56.
- [15] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms [C]// Proceedings of the 10th international conference on World Wide Web. ACM, 2001: 285-295.
- [16] PAZZANI M, BILLSUS D. Learning and revising user profiles: The identification of interesting web sites [J]. Machine learning, 1997, 27(3): 313-331.
- [17] ZHOU T, SU R Q, LIU R R, et al. Accurate and diverse recommendations via eliminating redundant correlations [J]. New Journal of Physics, 2009, 11(12): 123008.

### 作者简介

邹婧琳,女,四川大学公共管理学院信息管理技术系本科生。

范炜,男,1981年生,管理学博士,四川大学公共管理学院信息管理技术系讲师,研究方向:信息组织与信息检索。通讯作者,E-mail:fanwscu@163.com。

## Bibliography Recommendation Model Derived from Reading Behaviors of Online Social Book Community

ZOU JingLin FAN Wei

(Department of Information Management Technology, School of Public Administration, Sichuan University, Chengdu 610064, China)

Abstract: Bibliography recommendation aims to build a two-direction connection between books and readers. Due to users' participation in the online social reading community, huge data is newly created every day, which provides many references for bibliography recommendation. This paper analyzes users' primary reading behaviors and studies how users' rating, reading state, tag and action time present their reading preferences. Then it proposes a bibliography recommendation model, combining with users' reading preference intentions and levels. Finally, it uses the dataset from douban.com to evaluate the recommendation model. The experimental results show that the model achieves better recommendation in terms of accuracy and diversity, and the model could provide an action reference for online social reading community.

Keywords: Online social book community; Reading behaviors; Bibliography recommendation

(收稿日期: 2014-04-01)