

文献推荐系统综述*

□ 李琳娜 张志平 / 中国科学技术信息研究所 北京 100038
刘春霞 / 郑州铁路职业技术学院 郑州 450052

摘要: 文献推荐系统帮助用户在海量文献环境下发现个性化的信息, 已经成为文献检索系统的重要组成部分。文献推荐技术研究在信息检索、文献计量学与电子商务推荐系统研究成果综合演变下发展起来。首先讨论了一般个性化推荐技术; 进一步对文献推荐技术已经取得的研究成果进行了系统的分析与总结; 同时, 介于评价测度与方法是推荐系统的重要组成部分, 给出了常用的文献推荐系统的评价测度; 最后, 对文献推荐系统研究现状作出总体评价并指出将来的发展方向。

关键词: 文献推荐, 推荐系统, 协同过滤, 基于内容的推荐, 混合推荐

DOI: 10.3772/j.issn.1673—2286.2012.05.006

1 引言

随着互联网技术飞速发展, 电子文献数量越来越庞大, 如何帮助用户尤其是科研工作者从海量的电子文献中快速有效地找到其所需的相关文献已成为急需解决的问题。

个性化推荐技术能够有效解决信息过载问题, 是一个信息检索、人机交互、数据挖掘和用户建模等多学科交叉发展的领域, 多年来已经在研究领域取得了丰富的研究成果, 特别是在电子商务领域已取得了很好的应用效果, 目前已形成了较为丰富的方法与技术体系^[1]。

近年来, 针对文献领域的个性

化推荐技术在借鉴一般个性化推荐技术的基础上, 取得了一定的研究成果。一般电子商务个性化推荐最有代表性的技术是协同过滤技术^[2], 即根据用户对推荐项目空间的打分情况计算得出用户兴趣之间的相似性, 进一步地, 对一个用户的推荐将根据与该用户最相似的其他用户的行为给出推荐。文献推荐与电子商务个性化推荐相比而言, 文献推荐有文献的标题、摘要、作者、引文等多方面信息可以利用。在电子商务领域, 用户对产品的打分直接展示了用户对产品的喜爱程度。而在文献推荐领域, 研究者对一篇文献的打分不仅仅依赖于该篇文献是否有趣, 还依赖于该篇文献是否符合研究者的认知

需求。文献推荐技术相比于电子商务领域的个性化推荐技术面临着两个核心问题: 第一, 用户数目相对于文献数目的比率比电子商务领域用户数目相对于产品数目的比率要小1到2个数量级。在文献推荐领域, 通常是数万用户相对于数千万篇文献; 第二, 用户平均评价的文献数量非常少, 稀疏性问题比在电子商务领域更加严重。所以, 文献领域的个性化推荐技术与一般电子商务领域的个性化推荐技术既有共同点又有区别。

本文内容组织如下: 本文第2部分讨论了一般个性化推荐技术方法; 进一步地, 在第3部分对文献推荐技术已经取得的研究成果进行了系统的讨论分析与总结; 同

* 基金项目: 科技部项目“面向外科技文献信息的知识组织体系建设与应用示范”课题之一“信息资源自动处理、智能检索与STKOS应用服务集成”资金支持。

时, 鉴于评价测度与方法是推荐系统的重要组成部分, 并且是推荐系统发展路向与研究价值判断, 具有量化性旨归, 本文在第4部分系统地给出了文献推荐系统的评价测度; 最后, 对文献推荐领域研究现状作出总体评价并指出将来的发展方向。

2 个性化推荐技术一般方法

根据推荐算法的不同, 推荐系统可以分为如下几类^[1,2]:

(1) 基于内容的推荐系统。其推荐策略是: 推荐系统首先根据用户过去选择的项目, 利用项目的内容信息建立该用户的偏好模型, 然后根据偏好模型向用户推荐符合其偏好的项目。建构用户模型的主要方法是分析用户过去选择的项目的内容和打分, 采用信息检索中的基于关键词的分析技术或者概率模型建构用户偏好模型。尽管基于内容的推荐系统在许多领域取得了很好的应用, 但在长期的探索中, 研究者也观察到了基于内容的推荐方法存在的限制。首先, 由于该方法需要采用推荐项目的内容信息来建构用户偏好模型, 而对于图形、图像、视频等这些多媒体类型项目的内容信息很难较好地获取, 因此, 基于内容的推荐方法存在着推荐项目类型的限制。其次, 由于基于内容的推荐方法根据用户过去选择的项目内容建构用户偏好模型, 使得被推荐的项目受限于那些与该用户过去选择的项目相似的项目集合。对用户具有潜在兴趣的项目, 如果用户没有选择过与其相似的项目, 则不会被推荐; 再次, 对系统的新用户而言, 由于其

没有任何已选择的项目信息可以利用来建构其偏好模型, 故基于内容的推荐系统无法对新用户提供推荐服务。

(2) 协同过滤推荐系统。其推荐策略是: 向用户推荐与其偏好相似的其他用户选择的项目。更具体地, 对某一用户而言, 根据与其偏好相似的用户群对某项目的打分估计其对该未见项目的打分。目前, 协同过滤技术已成为最广泛采用的推荐技术, 电子商务领域已实现的比较成功的协同过滤推荐系统有Amazon、CDNow和LastFM等。由于协同过滤推荐方法根据用户之间的相似性进行推荐, 故其具有向用户推荐新颖性项目的的能力, 并且也能够像用户推荐图形、图像、视频等这些多媒体类型的项目。但是其仍然存在无法向没有任何已选择项目信息的新用户提供推荐服务的能力; 另外, 对于没有被任何用户选择过的新产品, 也无法被推荐; 再者, 由于用户打过分的项目通常是整个项目空间中很少的一部分, 因此在发现相似用户群时, 会出现相似测度计算偏差问题。

(3) 混合推荐系统。由于基于内容的推荐技术和协同过滤技术有各自的优缺点, 从而有许多学者提出了混合推荐技术。根据组合方法的不同, 可将现已存在的混合推荐系统分为四类: 简单地合并基于内容的推荐技术和协同过滤推荐技术各自产生的推荐结果来整合两种方法; 将基于内容的特征集成到协同过滤技术中, 比如利用项目的内容特征计算用户之间的相似度可以解决协同过滤技术面临的稀疏性问题; 将协同特征集成到基于内容的推荐技术中, 该方法主要

对基于内容的用户偏好进行维度约简; 构建将基于内容的推荐技术和协同推荐技术融合起来的统一框架。

3 文献推荐系统研究现状

文献推荐系统研究在上述一般推荐技术框架下展开, 同时根据文献系统用户空间与文献空间的特征, 引入文献计量学信息, 使用多种机器学习与数据挖掘技术, 创建出新颖的文献推荐技术, 形成了独特的推荐系统子领域。本文以方法范式为宏观框架, 以所使用的技术和信息为纲目, 将当前文献推荐领域的研究成果归结为如下几类。

3.1 基于内容的文献推荐技术

基于内容的文献推荐技术在推荐过程中主要采用文献的标题、摘要、类别、作者等特征信息或者采用用户的性别、年龄、职业等人口统计学信息。相关的研究成果有: Sullivan等^[3]首次提出将激活-扩散模型应用到文献推荐领域。用户的输入不再是一些关键词, 而是其感兴趣的文档。在设计关联矩阵的时候不仅考虑到了文本信息, 还同时考虑到了引文、共引强度、共被引强度等信息。实验结果表明基于文本信息的算法要优于基于引文信息的算法, 并且同时基于文本信息和引文信息的算法要优于仅仅基于单一信息的算法。Watanabe等^[4]开发了文献支持系统Papits, 该系统具有论文共享、论文推荐、论文抽取、论文分类等

功能。它用scale-free网络基于用户的浏览记录构建用户模型,然后通过计算用户模型与文献模型之间的相似度,将具有较高相似度的文献推荐给用户。Gori和Pucci^[5]根据文献之间的引用关系,将整个文献数据库表示为图,将相应的图矩阵进行归一化处理后,使用改进的PageRank算法进行推荐。Goodrum^[6]介绍了Citeseer系统利用文献间的引用关系发现相关文献的方法。Strohman等^[7]认为单纯的基于文本信息的推荐方法或者基于引文信息的推荐方法都有各自的缺陷,提出了将二者融合的推荐技术。Yu等^[8]以向量空间模型为基础表示用户偏好及文献,但基于中图分类法和子网对向量的每个元素在歧义层、同义词层和上下位层进行扩展。Sun等^[9]根据审稿人的发表记录,基于发表文献的标题、摘要、关键词及作者信息将其偏好用向量空间模型表示,将需要审稿的文献也用向量空间模型表示,然后将与审稿人偏好相似度较高的文献推荐给她审阅。Chandrasekapan等^[10]面对Citeseer系统中有发表记录的用户设计了相关文献推荐系统,该系统本质上是一个基于内容的推荐,但在计算文献与用户偏好之间的相似度的时候不再采用传统的基于向量空间模型的方式,而是将文献及用户偏好表示为概念树,通过计算概念树之间的编辑距离来计算文献及用户偏好之间的相似度。实验证实该方法比基于向量空间模型推荐技术的准确率要高。Pudhiyaveetil等^[11]进一步将用户范围扩展到所有有浏览记录的用户。Basu等^[12]针对如何将需要评审的很多会议论文提交给相应评

审人的问题,提出了将评审人的多种信息及论文的多种信息融合的推荐技术。Pohl等^[13]基于文献的共被查看信息向用户推荐研究文献,实验验证了基于共被查看信息的推荐要优于基于共被引信息的推荐。

3.2 基于协同过滤的文献推荐技术

高凯等^[14]首先用FPT(Frequent-Pattern-Time)树发现用户的共同兴趣,然后基于神经网络的向后传播算法进行推荐。Vellino^[15]基于先前的研究认为用PageRank算法生成的偏好矩阵的TOP-N推荐结果要优于布尔型偏好矩阵的推荐结果,而本文的实验却证实了在文献推荐领域,PageRank算法生成的偏好矩阵的TOP-N推荐结果要劣于布尔型偏好矩阵的推荐结果。Agarwal等^[16]针对文献推荐领域用户空间相对于庞大的文献空间非常小的状况,提出了子空间聚类算法。Chen等^[17]首先使用蚁群算法对用户进行聚类,然后在每一个簇内使用Apriori算法发现频繁模式,进一步发现关联规则从而完成相关推荐。文献^[18]首先根据用户的背景知识使用自适应共振理论将用户聚类,然后对每个簇使用Apriori算法发现频繁模式,进一步发现关联规则,从而完成推荐。Conry等^[19]针对会议论文评阅问题,提出了预测模型RMSE。

3.3 混合型文献推荐技术

Huang等^[20]提出了基于图模型的文献推荐系统,图的第一层是书及书之间的关联,第二层是用户及

用户之间的关联,两层节点之间的链接是用户对书的打分或者预测打分,具体的推荐策略是图搜索技术。Weng等^[21]提出了将本体和扩散激活模型融合的推荐技术。该项研究首先用本体建构用户偏好模型,然后用扩散激活模型确立用户之间的相似程度,进一步完成基于用户的协同推荐。Middleton等^[22]提出了基于本体方法构建用户偏好的推荐方法。他们设计了两个能够进行实时文献推荐的实验系统Quickstep和foxtrot,根据检测到的用户浏览行为和用户的反馈信息构建用户偏好。Liao等^[23-25]设计了文献推荐系统PORE。该系统将中图分类法作为图书馆本体,首先根据用户的借书记录计算用户对每个类别的喜欢程度,将大于预先设定的阈值的那些类别保留重新组织成用户个性化本体,然后计算用户对每个类别的关键词的兴趣度值,最后根据计算的关键词兴趣度值计算用户对每本书的偏好值,从而进一步完成个性化推荐。McNee等^[26]通过离线实验和在线实验分析了四个基于引文网络和协同过滤技术向一篇目标文献推荐相关参考文献的算法,离线实验结果表明基于不同的评价标准,推荐算法的比较结果会有很大的差异并且推荐列表的顺序也影响到了用户对系统的评价;在线实验结果表明大部分用户认为推荐算法对其有一定帮助作用。Hess等^[27]提出了将引文信息和可信度信息相融合的推荐技术。文献^[28]介绍了CYCLADES系统及它的个性化推荐及提醒功能。该系统的显著特点是数字图书馆环境下的协同及各种层次的个性化推荐功能。用户及用户群可以搜索、共享、组织

自己的信息空间。它可以根据用户及用户群的偏好向其推荐各种信息。系统不仅能向用户推荐相关的文献,还可以向用户推荐相似的兴趣。Naak等^[29, 30]认为依赖于用户打分信息的推荐存在着没有考虑用户打分原因——两个用户对同一篇文章打分相同时,初衷可能并不相同,进一步设计了文献管理系统Papyres,其推荐策略为级联式的混合推荐技术,首先是基于内容的文献推荐,其次是协同过滤技术对基于内容推荐结果的进一步净化。在协同过滤推荐方面,不仅仅从用户对文献的总体打分考虑,还从用户对文献的贡献、原创性、可读性、组织、技术质量等角度综合考虑。Gipp等^[31]实现了文献推荐系统Scienstein,该系统是一个集成引文分析、作者分析、源分析、隐式打分、显式打分等多方面信息的混合推荐系统。张等^[32]提出了根据协同标签进行语义相似度计算的推荐系统,该系统本质上是一个混合推荐系统,采用语义概念表示用户偏好及文献,首先用协同过滤技术发现相似邻居,然后采用基于内容的推荐技术进行推荐。Wang等^[33]基于概率主题模型提出了融合协同过滤技术和内容分析技术的混合文献推荐方法。该方法用潜在因子模型实现协同过滤技术,概率主题模型实现内容分析技术,从而将二者融合在一个统一的概率模型框架下,具体地将对一个用户的推荐归约为计算潜在变量的条件概率。

3.4 其他研究成果

Nuria等^[34]论述了具有个性化浏览和搜索功能的数字图书馆系统的实现,并给出了数字图书馆个

性化本体的构建原则。McNee等^[35]分析了文献推荐系统所面临的问题,通过对ACM数字图书馆130个用户的使用调查得出了只向用户推荐一定对其有用的文献才能提高用户对系统的信任程度的结论。Shimbo等^[36]评估了基于几个核的链接分析技术在文献推荐系统中的应用性能。文献^[37]介绍了卡尔斯鲁厄大学数字图书馆推荐系统BibTip,该系统是一个分布式结构,所采用的算法基于重复购买理论及泊松框架中独立交易过程基本理论分布以及对数级数分布算法。Matsatsinis等^[38]为了克服当前的推荐系统无法较好地平衡用户-系统交互和用户隐私的问题,很好地将决策理论中的多标准分析方法应用到文献推荐领域。

4 文献推荐系统的评价

面对不同的应用背景,如何选择合适的文献推荐算法是一项非常具有挑战性的工作。目前研究者对自己提出的文献推荐系统的评价主要还是借鉴一般推荐系统的评价方法,最常用的就是预测准确率,此外还有覆盖率、可扩展性等^[39-41]。下面介绍这些评价指标。

4.1 预测准确率

由于预测准确率可以离线测量,故其是介绍文献推荐系统的文献中讨论最多的度量方法。可以将文献中讨论的预测准确率分为如下两类:打分预测准确率和使用预测准确率。

(1) 打分预测准确率

RMSE (Root Mean Squared Error) 是最常用的一项度量指标,

假设系统的测试集为系统预测用户 u 对文献 i 的打分为 $\hat{r}_{ui}$, u 对 i 的真实打分为 r_{ui} ,那么RMSE定义为:

$$RMSE = \sqrt{\frac{1}{|\mathcal{P}|} \sum_{(u, i) \in \mathcal{P}} (\hat{r}_{ui} - r_{ui})^2} \quad (1)$$

MAE是另一个普遍采用的度量指标,定义为:

$$MAE = \sqrt{\frac{1}{|\mathcal{P}|} \sum_{(u, i) \in \mathcal{P}} |\hat{r}_{ui} - r_{ui}|} \quad (2)$$

RMSE对大的错误惩罚很严重,比如,若需要对四篇文献打分,系统A对三篇文献的打分错误值都为2,对另一篇文献的打分正确;系统B对一篇文献的打分错误值为3,对其余三篇文献的打分正确,则RMSE倾向于系统A,而MAE倾向于系统B。

归一化RMSE (Normalized RMSE, NRMSE) 与归一化MAE (Normalized MAE, NMSE) 是用打分尺度将RMSE与MAE归一化,二者对推荐系统的度量结果分别与对应的RMSE和MAE的度量结果一致。

(2) 使用预测准确率

在一些应用领域,需要评价的不是系统对文献的打分,而是用户是否采纳系统推荐的文献。具体评价方法为:选择一个测试用户,将他对一些文献的选择结果隐藏起来,然后对比系统向用户推荐的文献与用户实际感兴趣的文献是否一致。在离线实验环境下,推荐项与隐藏项之间有表1所示四种关系。

表1 准确率-召回率数据表

	Recommended	Not recommended
Used	True-Positive(tp)	False-Negative(fn)
Not used	False-Positive(fp)	True-Negative(tn)

可以定义如下指标:

$$Precision = \frac{\#tp}{\#tp + \#fp} \quad (3)$$

$$recall = \frac{\#tp}{\#tp + \#fn} \quad (4)$$

$$F1 = \frac{\#fp}{\#fp + \#tn} \quad (5)$$

由于用户没有浏览过的文献并不代表他对该文献不感兴趣,或许是他没有搜索到该文献,因此表1中会存在 fp 值被过高估计的问题。

推荐的文献数量的增加能提高召回率,但是会降低准确率。在有些情况下,推荐给用户的文献数量是事先确定的,这种情况下常用的度量为Precision at N。而在某些情况下,推荐给用户的文献数量不是预先确定的,故需要在一定的推荐文献数量范围内评价算法。因此,一些研究者提出了“或”的评价方式,前者即是准确率-召回率曲线,后者为受试者工作特征(Receiver Operating Characteristic, ROC)曲线。虽然二者都度量用户实际感兴趣的文献占推荐的所有文献的比率,但准确率-召回率更侧重于该比率值,而受试者工作特征曲线更侧重于用户不感兴趣的文献占推荐的所有文献的比率。

给定两个推荐算法,需要对每个算法计算这两个曲线,如果一个曲线完全优于另一个曲线,可以轻易确定较优的算法。但当两个曲线有交叉的时候,需要根据实际情况确定较优的算法。

当有多个测试用户的时候,可以根据给定的推荐长度N,计算每个用户的准确率与召回率,然后取其平均值,该方式称为CROC

(Customer Receiver Operating Characteristic) 曲线。

4.2 覆盖率

一些推荐算法的准确率会随着数据量的增加而提高,但是却只能发现用户所有感兴趣的文献中的很少一部分或者只能向一部分用户提供较好的推荐结果。覆盖率用来度量推荐算法是否能够尽可能多地发现用户感兴趣的文献,或者为尽可能多的用户提供较好的推荐服务。它又分为项空间覆盖率、用户空间覆盖率。

(1) 项空间覆盖率

通常,项空间覆盖率指系统能推荐的项占整个项空间的比率,又称为目录覆盖率。目录覆盖率最简单的度量方法是所有曾被推荐的项的比率。更常用的度量是在实验过程中被推荐的项占整个项空间的比率。

目录覆盖率的另一个度量方式是推荐多样性,它度量用户对被推荐的文献感兴趣的概率,若文献 i 被选择的概率为 $p(i)$,那么基尼指数(Gini Index)为:

$$G = \frac{1}{n-1} \sum_{j=1}^n (2j-n-1)p(i_j) \quad (6)$$

这里, $i_1, i_2, \dots, i_n$ 是随 $p(i)$ 值增加的文献的排列。

(2) 用户空间覆盖率

用户覆盖率指系统能为其提供推荐服务的用户占有所有用户的比率,在某些情况下,由于某些原因,系统无法为一些用户提供推荐服务。比如,系统为某些用户提供推荐的准确率较低,因此,为了获得较高的用户空间覆盖率,需要在覆盖率与准确率之间折中。

4.3 可扩展性

由于文献推荐系统帮助用户在大规模的数据集中选择其感兴趣的文献,所以系统的一个目标就是要在真实数据集上运行。因此,为了提高系统的运行速度,可能要降低算法的其他属性,如准确率或覆盖率。

随着数据集的增长,很多推荐算法为了保持运行速度,都需要更多的计算能力或存储空间,在计算机科学领域,对这方面的研究称为计算复杂度。

由于很多应用领域需要推荐系统提供在线推荐功能,测量系统能够提供推荐的速度是一项非常重要的任务。一个方法是测量系统能在一秒内推荐的文献数量,也可以度量响应时间——能够实时提供一项推荐的时间需求。

5 结束语

随着Internet的迅猛发展,网络电子文献的规模和复杂度在不断提高。对很多科研工作者而言,文献检索是一项相当繁重的工作。文献推荐系统通过分析用户的浏览记录及用户的个人信息,采用合适的推荐算法,可以在一定的准确度上向用户推荐其感兴趣的文献,不仅可以在很大程度上减轻用户的文献检索工作,还可能帮助用户发现其不可能发现的一些对其有用的文献。本文介绍了文献推荐系统的研究现状。虽然这些推荐系统在实验分析上都具有很好的推荐效果,但真正投入实际应用的并不多。文献推荐系统的研究仍面临着如下问题:

(1) 研究之宏观一致性: 近

十年来的文献推荐系统研究社区是个性化推荐系统研究共同体的一部分。其所用的核心推荐策略是在一般个性化推荐以及其他领域推荐系统推荐策略的谱系范围之内。

(2) 研究之宏观完备性: 一般推荐系统使用的基于内容、协同过滤以及混合型推荐策略, 在文献推荐系统中都有呈现。但是从研究现状可以看到, 在一般推荐系统领域采用最广泛的协同过滤推荐技术在文献推荐领域并没

有体现出来, 这主要是因为文献推荐领域的数据相比于一般电子商务领域的数据稀疏性问题更加突出。

(3) 一般个性化推荐系统研究当前已经开始广泛研究基于社区网络分析的方法, 极大地丰富了推荐系统的方法谱系。这一研究探索并未在文献推荐系统研究社区内广泛和深入展开, 个别研究案例只是初步的探索。

(4) 用户偏好模型时变因素未能充分探讨: 随着用户生活、研

究和学习经历、探索上下文的变迁, 用户偏好模型必然处于动态变化之中, 当前研究未能充分对此演化机制展开探讨。

(5) 评价体系的简单化: 当前文献推荐系统的评价体系仍然借鉴于一般个性化推荐系统的评价方法, 对于其特定的应用背景来说是不充分的, 需要引入更加全面的评价体系, 才能系统的研究评价一个推荐方法及其系统的内在性能, 并为优化和拓展研究提供指导。

参考文献

- [1] LOUSAME F P, SANCHEZ E. A Taxonomy of Collaborative-Based Recommender Systems [J]. Studies in Computational Intelligence, 2009, 229: 81-117.
- [2] ADOMAVICIUS G, TUZILIN A. Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(6): 734-749.
- [3] WOODRUFF A, GOSSWEILER R, PITKOW J, et al. Enhancing a Digital Book with a Reading Recommender [C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, NY, USA. ACM, 2000: 153-160.
- [4] WATANABE S, ITO T, OZONO T, et al. A Paper Recommendation Mechanism for the Research Support System Papits [C]// 2005 International Workshop on Data Engineering Issues in E-Commerce, Tokyo, Japan. IEEE, 2005: 71-80.
- [5] GORI M, PUCCI A. Research Paper Recommender Systems: A Random-Walk based Approach [C]// Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Washington, USA. IEEE Computer Society, 2006: 778-781.
- [6] GOODRUM A. Scholarly Publishing in the Internet Age: a Citation Analysis of Computer Science Literature [J]. Information Processing & Management, 2001, 37(5): 661-675.
- [7] STROHMAN T, CROFT W B, JENSEN D. Recommending Citations for Academic Papers [C]// Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, NY, USA. ACM, 2007, 705-706.
- [8] YU Z T, ZHENG Z Y, GAO S X, et al. Personalized Information Recommendation in Digital Library Domain Based on Ontology [C]// IEEE International Symposium on Communications and Information Technology, 2005, 1249-1252.
- [9] SUN Y H, NI W J, MEN R. A Personalized Paper Recommendation Approach based on Web Paper Mining and Reviewer's Interest Modeling [C]// Proceedings of the 2009 International Conference on Research Challenges in Computer Science, Washington, DC, USA. IEEE Computer Society, 2009, 49-52.
- [10] CHANDRASEKARAN K, GAUCH S S, LAKKARAJU P, et al. Concept-based Document Recommendations for Citeseer Authors [C]// Proceedings of the 5th International Conference on Adaptive Hypermedia and Adaptive Web-based Systems. Berlin, Heidelberg. Springer-Verlag, 2008, 83-92.
- [11] PUDHIYAVEETIL A K, GAUCH S, LUONG H, et al. Conceptual Recommender System for CiteseerX [C]// Proceedings of the Third ACM Conference on Recommender Systems, NY, USA. ACM, 2009, 241-244.
- [12] BASU C, HIRSH H, COHEN W W, et al. Technical Paper Recommendation: a Study in Combining Multiple Information Sources [J]. Artificial Intelligence Research. 2001, 14(1): 231-252.
- [13] POHL S, RADLINSKI F, JOACHIMS T. Recommending Related Papers Based on Digital Library Access Records [C]// Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, ACM New York, NY, USA, 2007, 417-418.
- [14] GAO K, WANG Y C, WANG Z Q. Similar Interest Clustering and Partial Back-Propagation-based Recommendation in Digital Library [J]. Library Hih Tech. 2005, 23(4): 587-597.
- [15] VELLINO A. Recommending Journal Articles with PageRank Ratings [EB/OL]. Research report. 2009. <http://cuvier.cisti.nrc.ca/~vellino/documents/PageRankRecommenderRecSys2009-Vellino.pdf>
- [16] AGRARWAL N, HAQUE E, LIU H, et al. Research Paper Recommender Systems: A Subspace Clustering Approach [C]// International Conference on Web-Age Information Management, Springer, 2005, 3739: 475-491.
- [17] CHEN C C, CHEN A P. Using Data Mining Technology to Provide a Recommendation Service in the Digital Library [J]. Electronic Library: Library and Information Studies. 2007, 25(6): 711-724.
- [18] TSAI C S, CHEN M Y. Using Adaptive Resonance Theory and Data-Mining Techniques for Materials Recommendation based on the E-Library Environment[J/OL]. Electronic Library: Library and Information Studies, 2008, 26(3): 287-302.
- [19] CONRY D, KOREN Y, RAMAKRISHNAN N. Recommender Systems for the Conference Paper Assignment Problem [C]// Proceedings of the third ACM Conference on Recommender Systems, ACM New York, NY, USA, 2009, 357-360.
- [20] HUANG Z, CHUNG W Y, ONG T H, et al. A Graph-based Recommender System for Digital Library [C]// Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, NY, USA. ACM, 2002: 65-73.

- [21] WENG S S, CHANG H L. Using Ontology Network Analysis for Research Document Recommendation [J]. Expert Systems with Applications, 2008, 34(3): 1857-1869.
- [22] MIDDLETON S E, SHADBOLT N R, DE ROURE D C. Ontological User Profiling in Recommender Systems [J]. ACM Transactions on Information Systems, 2004, 22(1): 54-88.
- [23] LIAO I E, LIAO S C, KAO K F, et al. A Personal Ontology Model for Library Recommendation System [J]. Digital Libraries: Achievements, Challenges and Opportunities, 2006, 4312: 173-182.
- [24] LIAO S C, KAO K F, LIAO I E, et al. Pore: a Personal Ontology Recommender System for Digital Libraries [J]. Electronic Library: Library and Information Studies, 2009, 27(3): 496-508.
- [25] LIAO I E, HSU W C, CHENG M S, et al. A Library Recommender System based on a Personal Ontology Model and Collaborative Filtering Technique for English Collections [J]. Electronic Library: Library and Information Studies, 2010, 28(3): 386-400.
- [26] MCNEE S M, ALBERT I, COSLEY D, et al. On the Recommending of Citations for Research Papers [C]// Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work, NY, USA. ACM, 2002, 116-125.
- [27] HESS C, STEIN K, SCHLIEDER C. Trust-Enhanced Visibility for Personalized Document Recommendations [C]// Proceedings of the 2006 ACM Symposium on Applied Computing, ACM New York, NY, USA. 2006, 1865-1869.
- [28] AVANCINI H, CANDELA L, STRACCIA U. Recommenders in a Personalized, Collaborative Digital Library Environment [J]. Journal of Intelligent Information Systems, 2007, 28(3): 253-283.
- [29] NAAK A, HAGE H, AIMEUR E. A Multi-criteria Collaborative Filtering Approach for Research Paper Recommendation in Papyrus [C]// Proceedings of MCETECH, 2009, 26: 25-39.
- [30] NAAK A, HAGE H, AIMEUR E. Papyrus: A Research Paper Management System [C]// Proceedings of 10th IEEE Conference on E-Commerce Technology and the Fifth IEEE Conference on Enterprise Computing, 2008, 201-208.
- [31] GIPP B, BEEL J, HENTSCHEL C. Scienstein: a Research Paper Recommender System [C]// Proceedings of the International Conference on Emerging Trends in Computing. Virudhunagar, India. IEEE, 2009, 309-315.
- [32] ZHANG M, WANG W C, LI X M. A Paper Recommender for Scientific Literatures Based on Semantic Concept Similarity [C]// Proceedings of the 11th International Conference on Asian Digital Libraries: Universal and Ubiquitous Access to Information, Springer-Verlag Berlin, Heidelberg, 2008, 359-362.
- [33] WANG C, BLEI D M. Collaborative Topic Modeling for Recommender Scientific Articles [C]// Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM New York, NY, USA. 2011, 448-456.
- [34] NURIA F F, ENRIC M P, JULIA M A. Towards Personalization in Digital Libraries through Ontologies [J]. Library Management, 2005, 26(4/5): 206-217.
- [35] MCNEE S M, KAPOOR N, KONSTAN J A. Don't Look Stupid: Avoiding Pitfalls when Recommending Research Papers [C]// Proceedings of the 2006 20th anniversary conference on Computer Supported Cooperative Work, 2006, 171-180.
- [36] SHIMBO M, ITO T, MATSUMOTO Y. Evaluation of Kernel-based Link Analysis Measures on Research Paper Recommendation [C]// Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, ACM New York, NY, USA. 2007, 354-355.
- [37] GEYER-SCHULZ A, NEUMANN A, THEDE A. An Architecture for Behavior-Based Library Recommender Systems [J]. Information Technology and Libraries, 2003, 22(4): 165-174.
- [38] MATSATSINIS N F, LAKIOTAKI K, DELIAS P. A System based on Multiple Criteria Analysis for Scientific Paper Recommendation [C]// Proceedings of the 11th Panhellenic Conference on Informatics, Patras, Greece, 2007, 135-149.
- [39] SHANI G, GUNAWARDANA A. Evaluating Recommendation Systems [OL]. www.research.microsoft.com/pubs/115396/EvaluationMetrics.TR.pdf.
- [40] MCNEE S M, RIEDL J, KONSTAN J. Accurate is not enough: how Accuracy Metrics have Hurt Recommender Systems [C]// Proceedings of the CHI' 06 Extended Abstracts on Human Factors in Computing Systems, ACM New York, NY, USA, 2006, 1097-1101.
- [41] SULLIVAN D O, SMUTH B, WILSON D. Preserving Recommender Accuracy and Diversity in Sparse Datasets [J]. International Journal on Artificial Intelligence Tools, 2004, 13(1): 219-236.

作者简介

李琳娜 (1981-), 博士后。研究方向: 数字图书馆, 推荐系统。E-mail: liln@istic.ac.cn
 张志平 (1963-), 研究员。研究方向: 智能信息检索, 数字图书馆技术。
 刘春霞 (1979-), 硕士。研究方向: 推荐系统, 数据挖掘。

A Survey of Document Recommender Systems

Li Linna, Zhang Zhiping / Institute of Scientific and Technical Information of China, Beijing, 100038
 Liu Chunxia / Zhengzhou Railway Vocational College, Zhengzhou, 450052

Abstract: Document recommender systems can help people acquire documents that correspond to their information need from huge amount of documents and have become the key part of document retrieval systems. It is developed under the comprehensive evolution of information retrieval, bibliometrics and E-commerce. General recommender systems were introduced firstly in this paper. The research to document recommender system were systematically analyzed and discussed subsequently. Because measurement plays important role for recommender systems, the often used methods to evaluate document recommender system is investigated at the same time. Finally, an overall assessment for document recommender system and its future direction is discussed.

Keywords: Document recommender, Recommender systems, Collaborative filtering, Content-based recommender, Hybrid recommender

(收稿日期: 2012-01-18)