

# 基于内容和协作的推荐系统的研究

黄文荣, 周长征

(西南财经大学经济信息工程学院, 四川 成都 610074)

**摘 要:** 对推荐系统理论研究与技术开发现状进行了综述, 分析与评述了各种推荐技术的原理方法、特点、优势和不足, 阐述了推荐系统研究的主要内容, 指出了现有推荐系统存在的问题和研究的发展方向。

**关键词:** 推荐技术; 协作推荐系统; 向量空间模型; 相关系数

## 0 引言

协作推荐技术的思想可以追溯到 20 年前, Goldberg 和他的同事们将浏览过的网页文本内容和点击偏好定义为用户爱好, 然后用于推荐系统。之后, Group Lens 和 Ringo 等分别将其继续发展成协作推荐技术。世纪之交, 国内已有专家学者在讨论中文文本推荐的相关技术。2001 年 5 月, 香港召开的第 10 届 WWW 大会上, Sarwar 教授等人 (Group Lens 研究小组), 发表了题为《Item-based Collaborative Filtering Recommendation Algorithms》的论文。此后, 在研究方面, 大多学者将协作推荐技术应用于推荐系统, 研究其应用理论框架及与实践相结合的问题。

协作推荐系统应用前景广阔, 如许多知名的大型网上商店: Amazon、CDNow、和 Yahoo! 等等。目前, 有从网上移植到网下的趋势, 应用于人们的日常生活的方方面面: 如在大型商场中, 可对客户或会员进行终端拦截式的自动化销售, 从而提高用户的满意度和商场的业绩; 在图书馆文献查询服务系统中, 则可为其客户提供精确或模糊的相关文献搜索、推荐。

推荐技术通常分为三类: 基于内容的推荐技术 (CBF)、协作推荐技术 (CF) 以及混合推荐技术 (HF)。CBF 着眼于产品内容之间的相关性; CF 则侧重于多个用户之间爱好的相似性; HF 则综合了 CBF 和 CF 的优点而具有灵活性和精确性<sup>[1]</sup>。

本文在第一部分介绍了基于内容推荐模型——通过用户自身的爱好信息进行推荐; 在第二部分, 阐述了基于内存和基于模型的协作推荐机制; 最后, 指出了现有推荐系统存在的问题和研究的发展方向。

## 1 基于内容的推荐技术

基于内容的推荐技术 (Content-Based Filtering) 主要采用概率统计 (词频、相关系数)、人工智能、自然语言处理和机器学习等技术进行推荐。基于内容的推荐技术典型过程就是: 首先, 收集用户爱好信息、构建和维护用户概要信息库; 其次, 建立清晰、完整的用户爱好模型 (即表示成用户向量空间中的一个用户向量); 接着, 对文本集内的文本进行分词、词频统计、加权等过程来生成每一文本的文本向量; 然后, 计算用户向量和文本向量之间的相关系数, 将相关系数高的文本发送给该用户模型的用户; 最后, 根据用户的反馈信息进行修正, 以提高推荐信息的效率和质量。

### 1.1 构建用户概要信息库

基于内容的推荐技术通常用来处理文字类的产品 (如: 网

页文章)。如果是非文字类的产品 (如: 多媒体文件) 则可利用它的相关说明。例如: 音乐推荐系统可以根据该音乐的相关描述: 标题、风格、专辑名、艺术家等等来建立用户概要信息库, 通过它可以清楚地了解到该用户的音乐品位和爱好。

由于用户手工输入和保护隐私的限制, 基于内容的推荐系统经常不得不面对收集用户概要信息的困境。此时, 系统会选取一些用户浏览过或所感兴趣的文章, 再从文章中抽取重要的词汇以建立用户概要信息库。例如: Fab 网页推荐系统, 它先从用户所喜欢的网页的文章中通过一定的算法摘录 100 多个词汇, 从而构成用户概要信息库<sup>[2]</sup>。例如: TF-IDF 算法, 基本思路是: 如有  $N$  篇向用户推荐的文档, 将其中的每一篇文档  $j (j \in N)$  表示成向量空间中的一个向量  $\vec{w}_j$ , 即将用户感兴趣的文档表示成关键词向量, 向量的每一维由文档  $j$  中的一个关键词条  $i (i \in R, R$  为文档  $j$  中所有不重复的关键词条数目) 和其权重  $W_{ij}$  组成。其典型公式为:

$$W_{ij} = TF_{ij} \times IDF_i \quad (1.1)$$

$$TF_{ij} = \frac{f_{ij}}{\max_z f_{zj}} \quad (1.2)$$

$$IDF = \log \frac{N}{n_i} \quad (1.3)$$

$$\vec{w}_j = (W_{1j}, W_{2j}, \dots, W_{k_jj}) \quad (1.4)$$

前一项乘数  $TF_{ij}$  是标准词频 (Term Frequency/Normalized Frequency),  $f_{ij}$  表示关键词条  $i$  在文档  $j$  中出现的次数,  $\max_z f_{zj}$  表示在文档  $j$  中的所有词条中出现频率最高的词条出现的次数; 后一项  $IDF$  是词条  $i$  的逆向文档频率 (Inverse Document Frequency), 表示包含关键词条  $i$  的文档数量越多, 其在区分文档中的作用越小——主要用来排除虚词等一些毫无意义的语法词, 其中  $n_i$  表示在所有  $N$  篇文档中所包含词条  $i$  的文档总数量。

但是, TF-IDF 算法一开始面临关键词条选取问题, 需要压缩掉无关数据维, 降低数据噪音。后有学者将 (1.1) 改进为 (1.1'), 效果较好:

$$W_{ij} = \left( a + \frac{(1-a)f_{ij}}{\max_i f_{ij}} \right) \times \log \frac{N}{n_i} \quad (1.1')$$

其中:  $a$  为权重, 一般可取 0.5。

通过将用户感兴趣的文档表示成关键词向量, 可以建立起用户概要信息库。接下来就是选取合适的推荐产品。

## 1.2 选取和推荐产品

建立用户概要信息库后,就可与要推荐产品进行匹配与推荐。此时要用到相关系数。相关系数是指产品与用户概要信息之间的相关联的程度,计算的方法有:布尔代数法、向量空间模型VSM(余弦法)、概率模型、神经网络模型、模糊数集模型等等。目前,在文本类信息处理方面上,应用广泛的是向量空间模型(VSM),它基于统计分类方法,将给定的文档转换成一个维数很高的关键词向量,计算出两个向量间的相关系数,即所推荐的文档与用户概要信息特征向量的相关性。

由(1.1)或(1.1')可以方便地根据用户概要信息计算用户特征向量 $\bar{W}_u=(W_{1,u}, W_{2,u}, \dots, W_{M,u})$ 和根据产品(文档)信息计算每一件产品的特征向量(文档关键词条向量) $\bar{W}_p=(W_{1,p}, W_{2,p}, \dots, W_{M,p})$ 。之后,定义它们之间的余弦相关系数 $S_i(u, p)(j \in N)$ :

$$\begin{cases} \text{Max } S_i(u, p) \\ \text{S.T. } S_i(u, p) = \frac{\sum_{i=1}^M (W_{i,u} \times W_{i,p})}{\sqrt{\sum_{i=1}^M W_{i,u}^2} \times \sqrt{\sum_{i=1}^M W_{i,p}^2}} \end{cases} \quad (1.5)$$

其中,  $M$  为文档和用户概要信息库中的关键词条的总数目。求出  $\text{Max } S_i(u, p)$  所对的产品(文档)即为推荐的产品(文档)。

## 1.3 基于内容的推荐技术的局限性

虽然基于内容的推荐技术对于文本内容的推荐简单有效,推荐结果直观、容易解释、不需要领域知识,同时能推荐新的或不是很流行的产品(文档),且能为具有特殊兴趣爱好的用户进行推荐;但是,它仍有不足之处,要求所有内容能容易抽取成有意义的特征,要求特征内容有良好的结构性,并且用户的爱好必须能够用内容特征形式来表达。其余还有<sup>[6]</sup>:

### 1.3.1 过分依赖文本中的关键词条、词频等信息

有时一些高质量的产品会因其估值过低而被忽略;有时一些低质量的产品却会因其估值很高而被推荐。这是因为,它包含了大量的高频的关键词条。长此以往,必然丧失用户的信任。解决办法之一是,引入加权数,综合考虑作者、日期、出版来源、引用或已阅次数等参数。

### 1.3.2 对于非文本类的产品景象不够好

例如:一些多媒体产品(图像、影音)本身很少包含关键词条。此时,可引入其属性信息、说明简介和相关评论推荐,以克服信息不足的弊端。

## 2 协作推荐技术

协作推荐技术(Collaborative Filtering)是信息推荐和推荐系统中应用较早和较为成功的技术之一。它的基本思想是物以类聚:如果一些用户对一些产品(文档)的评分比较相似,则他们对其他产品(文档)的评分也比较相似,并据此推荐;同时,用户也比较偏爱与其已购买的东西相类似的产品。

由此可以看出,协作推荐技术(CF)与基于内容的推荐技术(CBF)不同点在于:后者主要从产品内容出发进行产品匹配推荐;前者则是从用户角度出发,由用户向用户自动推荐产品,且在此过程中,不用再填写调查表格等等,所需要的信息都是系统从用户购买模式或浏览行为等隐式获得的。协作推荐技术

最大特点是,对推荐的产品没有特殊的要求,能处理一些非结构化的复杂对象(如:多媒体)和一些非数字化的产品。

协作推荐技术(CF)与基于内容的推荐技术(CBF)的应用过程类似:收集用户信息、选择和推荐匹配产品。所不同的是:协作推荐技术CF更多地隐式收集信息以及由于匹配对象的不同而采用不同的算法——基于内存的协作推荐技术和基于模型的协作推荐技术。

### 2.1 基于内存的协作推荐技术

基于内存的协作推荐技术(Memory-based Collaborative Filtering)因其使用简单、结果直观、且有良好的质量,而广泛应用于许多商业网站。美国的Minnesota大学的Group Lens研究实验较早地开发了基于内存的协作推荐技术;它采用了修正的Pearson相关系数法计算用户与其“最近邻居”的用户群之间的相关系数。

与基于内容的推荐技术不同的是:基于内存的协作推荐技术的核心是用户与用户之间的推荐,而基于内容的推荐技术是产品之间的推荐。故此,它们所适用的方法也不尽相同,但总的来说,基于内存的协作推荐技术的精度要高于前者。

基于内存的协作推荐技术可以分为三个步骤:建立“最近邻居”的用户群;预估用户的评分值;推荐相关产品。

#### 2.1.1 建立“最近邻居”的用户群

“最近邻居”的用户群是指与推荐对象兴趣最相同(或共同产品的评分最相似)的一组用户。它的核心思想是计算两个用户的相关系数。过程如下:首先,搜寻用户对以往的共同产品的历史评分数据;其次,基于这些评分数据项,选择一个合适的相关系数计算方法,计算和保存用户之间的相关系数值。然后,统计分析这些值,选取最佳的相关系数值下限,将其相关系数值在下限以上或较高的用户聚类为“最近邻居”用户群。

“最近邻居”用户群的相关系数算法较多,例如前文所提及的余弦法、Pearson相关系数法(及其修正法),还有泊松相关系数法等等。通常情况下,会根据数据集的不同选择不同的算法。这里主要介绍一下应用广泛的修正的Pearson相关系数法:设有 $(n+1)$ 个用户: $X, Y_j(j \in n)$ ;交叉对 $m$ 件产品给出评分;用户 $X, Y_j(j \in n)$ 对产品 $i$ 的评分为 $R_{y_j, i}$ ;  $\bar{R}_X, \bar{R}_{Y_j}$ 为用户 $X, Y_j(j \in n)$ 评分的算术平均值;用户 $X, Y_j(j \in n)$ 之间的相关系数为 $S_p(x, y_j)$ 。则

$$\begin{aligned} S_p(x, y_j) &= \frac{\text{cov}(x, y_j)}{\sigma_x \sigma_{y_j}} \\ &= \frac{\sum_{i=1}^m [(R_{x, i} - \bar{R}_X) \times (R_{y_j, i} - \bar{R}_{Y_j})]}{\sqrt{\sum_{i=1}^m (R_{x, i} - \bar{R}_X)^2} \times \sqrt{\sum_{i=1}^m (R_{y_j, i} - \bar{R}_{Y_j})^2}} \quad (2.1) \end{aligned}$$

计算出 $S_p(x, y_j)$ 后,舍去 $S_p(x, y_j) < \varphi$ 的用户,其余即为“最近邻居”的用户群中的用户。下限 $\varphi$ 的取值视实际情况综合平衡,过高则用户数量不够而影响精度,过低则用户数据庞大而影响运算速度,一般可取为0.6~0.8之间。用户样本数 $n$ 也可与 $\varphi$ 一起考量,为保证算法的一般性,可随机抽样。这样的益处在于,不再将用户局限于固定的某一归类,毕竟随着时间变幻,人们的爱好可能会随之而变——特别是一些时尚、流行的元素。

### 2.1.2 预估用户的评分值

当“最近邻居”的用户群中的用户( $k$ 个,  $k < n$ )及其相关系数确定之后,就可基于用户对以往产品的评分来预估用户最可能感兴趣的产品的评分值。设:  $P_{x,i}$  为用户  $X$  对未评分(或未用过)的产品  $i$  的预估评分值。

$$P_{x,i} = \bar{R}_x + \frac{\sum_{j=1}^k [(R_{y,j} - \bar{R}_y) \times S_p(x, y_j)]}{\sum_{j=1}^k |S_p(x, y_j)|} \quad (2.2)$$

这里,应用到这样一个假设:每个用户对未曾接触过的产品谈不上是爱好还是厌恶,故此,取其所有的评分值的均值。关键的是后面一项修正值,它用“最近邻居”的用户群来修正用户  $X$  对产品  $i$  的预估评分值——与用户  $X$  兴趣相投(正相关)的人所爱好的产品,用户  $X$  也有好感;与用户  $X$  格格不入(负相关)的人所爱好的产品,用户  $X$  评分值却较低。

### 2.1.3 推荐相关产品

至此,得出用户  $X$  关于产品的预估评分值  $P_x$  之后,选取一个或一组评分值  $P_x$  较高的产品推荐给用户即可。

### 2.1.4 例证:音乐推荐系统

设:用户  $x, y (j \in 3)$  交叉对 6 件产品(Music 1~ Music 6)给出评分如表 1。试预估用户  $x$  对 Music6 的评分值。

表 1

	$x$	$y_1$	$y_2$	$y_3$	$y_4$
Music 1	5	2	4		
Music 2	1	4	2		
Music 3			3	3	
Music 4	4	1			
Music 5	2	5			
Music 6	?	2	5		
Music 7			5		
Music 8					

解题过程如下:先求出其均值( $\bar{R}_x = 3; \bar{R}_y = 3$ ),再求出其相关系数( $S_p(x, y_1) = -0.8$ ),最后求预估评分值( $P_{x,6} = 4.56$ ):

$$\bar{R}_x = \frac{5+1+4+2}{4} = 3$$

$$\bar{R}_y = \frac{2+4+1+5}{4} = 3$$

$$S_p(x, y_1) = \frac{(5-3) \times (2-3) + (1-3) \times (4-3) + (4-3) \times (1-3) + (2-3) \times (5-3)}{\sqrt{(1-3)^2 + (5-3)^2 + (2-3)^2 + (4-3)^2} \times \sqrt{(4-3)^2 + (2-3)^2 + (5-3)^2 + (1-3)^2}} = -0.8$$

$$P_{x,6} = \bar{R}_x + \frac{(R_{y,6} - \bar{R}_{y_1}) \times S_p(x, y_1) + (R_{y,6} - \bar{R}_{y_2}) \times S_p(x, y_2)}{|S_p(x, y_1)| + |S_p(x, y_2)|} = \frac{(2-3) \times (-0.8) + (5-3) \times 1}{|1| + |-0.8|} = 4.56$$

### 2.2 基于模型的协作推荐技术

基于模型的协作推荐技术(Model-based Collaborative Filtering),先用历史数据训练用户得到一个模型,再用此模型进行预测和推荐产品。在早期的假设和研究的许多方法中,共同采用了聚类思想:将评价相近的用户或产品进行分组和推荐。

例如:贝叶斯聚类分析模型(BC: Bayesian Clustering)、贝叶斯网络模型(BN: Bayesian Network)、和 Aspect 模型(即:面向方面模型)等等。BC 模型的基本思想是假设多项式混合参数模型的评分值是已知或可观察的,并采用了 EM 算法(期望值最大化)。它能有效地提高推荐算法的实时响应速度,扩展性和实现性能上比传统算法略显优势,一定程度上也可以降低数据的稀疏性,但同时也会降低推荐质量,聚类的边缘推荐精度较低。BN 模型建立在在可被观察的产品的数据中寻找产品之间的关联规则的基础上。它增加用户在产品空间上的重叠数目,降低了数据稀疏性——但牺牲了推荐精度,且面临可扩充性问题。Aspect 模型则在潜在模型中引入用户群和产品组等概念。

随着 Aspect 模型被成功地广泛应用,研究者又开发了许多概率模型来加以完善。Flexible mixture 模型扩展了聚类算法模型<sup>[4]</sup>:它放弃了用户和产品的聚类必须只能归结为某个单一的类中的假设。Preference-based graphic 模型着重于用户行为之间的区别:在兴趣相同的用户群中,各个用户对产品的评分值是不同的,某些人的评分值高于(或低于)其他用户的评分值。Popescul 在 Aspect 模型中引入了产品的属性信息,从而有效地提高了推荐质量,克服了评分值数据稀疏问题<sup>[5]</sup>。Hofmann 更进一步,将 Aspect 模型发展成 PLSA 模型(即:潜在语义分析概率模型):它支持多种属性评分值来取代某单一类别的评分值(例如:二进制评分数据等等)。

Pennock 将基于内存的协作推荐技术和基于模型的协作推荐技术结合起来使用。Yu 等利用概率的方式将协作推荐技术的两种算法结合使用:首先,使用机器自动化学用户爱好以建立用户们的概率模型;然后将用户概要信息的混合模型用于推荐产品。

除此之外,基于模型的协作推荐技术还有降维算法、潜在类别模型(Latent Class Models)、概率相关模型、线性回归模型、最大熵模型等算法。这些模型算法的重点都在于改善、解决协作推荐算法中的稀疏性问题和可扩展性问题。其中:降维算法引入了奇异值分解技术 SVD(Singular Value Decomposition),从产品评分矩阵中抽取一些本质的特征,来降低初始数据的维数;潜在类别模型算法可以捕获存在于用户与产品之间的潜在关系,并据此计算用户喜欢某一产品的可能性。

基于模型的协作推荐技术的优点是建立的模型比原始数据库小得多,能有效缓解推荐算法的实时性问题。缺点是训练后的模型具有滞后效应(跟不上用户爱好变化);因此为了保证模型的有效性,必须定期对模型进行必要的更新训练。

### 2.3 协作推荐技术的局限性

协作推荐技术使用了其他用户对产品的评分值,能处理复杂的非结构化对象,推荐个性化、自动化程度高,因此,它可广泛地用于各种类型的产品的推荐。但是,它也有自身的缺点:

稀疏性问题(Sparsity problem):在许多推荐系统当中,只有较少的有效的数据可用来预估用户对产品的评分值,用户的评分数据较少的话很难找到“最近邻居”用户群,从而降低了推荐效果。解决办法之一是:使用用户概要信息和产品属性信息来计算用户之间和产品之间的相关系数。另外,也可以通过用户

```

//池中含nThreads个线程:
exec=Executors.newFixedThreadPool(nThreads);
//向执行框架提交任务列表中的各个任务:
for(FutureTask<BigDecimal> task:listOfTask)
    exec.execute(task);
//等待各并发任务完成,提取各自的计算结果:
for(int k=0; k<listOfTask.size(); k++)
    try { result[k]=listOfTask.get(k).get();
        //get()方法依赖于任务的完成状态:
        //若已完成,立即获取结果;
        //若未完成,自我阻塞直到任务完成。
    } catch (Exception e) {
        System.out.println(e);
    }
//根据 Gauss 公式,合并形成最终结果:
BigDecimal pi;
BigDecimal bd48=new BigDecimal(48);
BigDecimal bd32=new BigDecimal(32);
BigDecimal bd20=new BigDecimal(20);
pi=bd48.multiply(result[0]);
pi=pi.add(bd32.multiply(result[1]));
pi=pi.subtract(bd20.multiply(result[2]));
return pi;
}
}

```

如表1所示,其数据是在CPU主频1.66GHz、内存1.24GB的双核计算机上,以不同的线程池尺寸(nThreads)构造的执行框架中运行上述程序精确计算 $\pi$ 到小数点后1000位所用的平均时间(对每一个nThreads参数的取值,程序运行10次再取运

行时间的平均值)。

表1 线程池尺寸对执行时间的影响

nThreads	平均运行时间(ms)
1	1015
2	610
3	635
6	644

实验数据表明:当线程池中的线程数nThreads为1时,没有充分利用双核处理器的能力,而nThreads大于2时,由于池中的线程数大于实际可用的处理单元数,所以与nThreads等于2时相比,额外增加了少量线程调度开销(对于nThreads>2的各种取值,时间递增不十分明显,所以表1中仅列出了取值3和6时的参考数据)。

#### 4 结束语

与“一任务一线程”的多线程编程方式相比,基于Java线程池执行框架的并发机制具有显著的优点:实现任务描述、提交与任务执行之间的解耦;便于对任务和线程的管理,提高系统效率;程序代码结构清晰易于维护。

#### 参考文献:

- [1] 郑和根.操作系统概论[M].高等教育出版社,2004.
- [2] Sun Microsystems Inc. Java Platform, Standard Edition 6 API Specification[EB/OL].<http://java.sun.com/javase/6/docs/api/index.html>,2008.
- [3] [美]Brian Goetz, Tim Peierls等著,韩婧,方妙伟. Java并发编程实践[M].电子工业出版社,2007.



(上接第7页)

过去的交易和反馈信息,使用ARF技术(Associative Retrieval Framework)或RSA算法(Related Spread Activation)传递用户之间的相关性。

冷启动问题(Cold start problem):它是稀疏问题的极端情况,包括新用户问题和新产品问题(如表1中的用户y4和Music7、8)。新用户问题:当一个新用户加入系统之后,由于没有相关的用户评分信息,也就无从建立其“最近邻居”的用户群,从而也无法向其推荐任何产品。新产品问题:当一个新的产品首次出现的时候(或早期),用户评分没有(或较少),因此单一的协作推荐技术无法(很难)对其进行预测评分和推荐。

信息遗失问题(Ignorance problem):由于采用了聚类技术,协作推荐系统会忽略一些较为关键的信息,造成信息遗失。如表1,根据用户y2与y3的相关系数和预估评分值,向用户y3推荐Music6或Music7。此时,系统可能会随机选择其中之一。但是,所忽略的信息却表明:Music3属于摇滚风格,Music7也属于摇滚风格,而Music6为轻音乐风格,用户y3可能更喜欢Music7。

#### 3 结束语

在过去的数十年中,推荐技术得到显著进展,出现了基于

内容的推荐技术、协作推荐技术和混合推荐技术等,并开发出了许多商业化的推荐系统。然而,在处理数据的扩展性、稀疏性和推荐结果的有效性等方面仍有许多需要改进的地方。

#### 参考文献:

- [1] Q. Li, B. Kim. Clustering approach for hybrid recommender system[C]. Proceedings of IEEE/WI International Conference, 2003:33~38
- [2] M. Balabanovi c, Y. Shoham. Fab: content-based, collaborative recommendation[J]. Communications of the ACM, 1997, 40: 66~72
- [3] G. Adomavicius and A. Tuzhilin. Toward 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: 734~749
- [4] L. Si, R. Jin. Flexible mixture model for collaborative filtering[C]. Proceedings of International Conference on Machine Learning (ICML), 2003.
- [5] A. Popescul, L. Ungar, D. Pennock, S. Lawrence. Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments[J]. Proceedings of 17th Conference on Uncertainty in Artificial Intelligence, 2001: 458.



作者: [黄文荣](#), [周长征](#)  
作者单位: [西南财经大学经济信息工程学院, 四川, 成都, 610074](#)  
刊名: [计算机时代](#)  
英文刊名: [COMPUTER ERA](#)  
年, 卷(期): 2009 (2)

## 参考文献(5条)

1. [Q.Li;B.Kim](#) [Clustering approach for hybrid recommender system](#) 2003
2. [M.Balabanovi;C.Y.Shoham](#) [Fab:content-based, collaborative recommendation](#)[外文期刊] 1997
3. [G.Adomavicius;A.Tuzhilin](#) [Toward the next generation of recommender systems:a survey of the state-of-the-art and possible extensions](#)[外文期刊] 2005 (6)
4. [L.Si;R.Jin](#) [Flexible mixture model for collaborative filtering](#) 2003
5. [A.Popescul;L.Ungar;D.Pennock;S.Lawrence](#) [Probabilistic models for unified collaborative and content-based recommendation in sparse-data environments](#) 2001

## 本文读者也读过(7条)

1. [李晓昀. 余颖. LI Xiao-yun. YU Ying](#) [推荐技术的研究与探讨](#)[期刊论文]-[计算机时代](#)2009 (11)
2. [胡迎松. 韩苹. 赵勇](#) [基于内容和协同过滤的混合推荐技术](#)[会议论文]-2005
3. [方力. FANG Li](#) [基于信息推荐技术的元数据扩展设计研究](#)[期刊论文]-[情报科学](#)2008, 26 (5)
4. [王艳. 景韶光. 李雪耀. 于军](#) [基于分类方法的内容过滤推荐技术](#)[期刊论文]-[情报杂志](#)2005, 24 (8)
5. [徐永良. Xu Yongliang](#) [推荐技术在电子商务中的应用研究](#)[期刊论文]-[电脑与电信](#)2007 (9)
6. [曹毅](#) [基于内容和协同过滤的混合模式推荐技术研究](#)[学位论文]2007
7. [杨艳. 李建中](#) [数字图书馆中基于文献拓扑的个性化推荐技术](#)[会议论文]-2004

引用本文格式: [黄文荣. 周长征](#) [基于内容和协作的推荐系统的研究](#)[期刊论文]-[计算机时代](#) 2009 (2)