

## 互联网推荐系统比较研究<sup>\*</sup>

许海玲<sup>1+</sup>, 吴 潇<sup>2</sup>, 李晓东<sup>1</sup>, 阎保平<sup>1,3</sup>

<sup>1</sup>(中国科学院 计算机网络信息中心 CNNIC 实验室, 北京 100190)

<sup>2</sup>(中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

<sup>3</sup>(中国科学院 计算机网络信息中心, 北京 100190)

### Comparison Study of Internet Recommendation System

XU Hai-Ling<sup>1+</sup>, WU Xiao<sup>2</sup>, LI Xiao-Dong<sup>1</sup>, YAN Bao-Ping<sup>1,3</sup>

<sup>1</sup>(Laboratory of CNNIC, Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>3</sup>(Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100190, China)

+ Corresponding author: E-mail: hailingxu@gmail.com

Xu HL, Wu X, Li XD, Yan BP. Comparison study of Internet recommendation system. *Journal of Software*, 2009,20(2):350-362. <http://www.jos.org.cn/1000-9825/3388.htm>

**Abstract:** This paper makes a comprehensive survey of the recommender system research aiming to facilitate readers to understand this field. First the research background is introduced, including commercial application demands, academic institutes, conferences and journals. After formally and informally describing the recommendation problem, a comparison study is conducted based on categorized algorithms. In addition, the commonly adopted benchmarked datasets and evaluation methods are exhibited and most difficulties and future directions are concluded.

**Key words:** recommender system; social network; information overload; collaborative filtering; personalization

**摘 要:** 全面地总结推荐系统的研究现状,旨在介绍网络推荐的算法思想、帮助读者了解这个研究领域。首先阐述了推荐系统研究的工业需求、主要研究机构和成果发表的期刊会议;在讨论了推荐问题的形式化和非形式化定义之后,对主流算法进行了分类和对比;最后总结了常用数据集和评测指标,领域的重难点问题和未来可能的研究热点。

**关键词:** 推荐系统;社会网络;信息过载;协同过滤;个性化

**中图法分类号:** TP393 **文献标识码:** A

---

• Supported by the National Natural Science Foundation of China under Grant Nos.60773056, 60802028, 60873165 (国家自然科学基金); the National Basic Research Program of China under Grant No.2007CB311100 (国家重点基础研究发展计划(973)); the National High-Tech Research and Development Plan of China under Grant No.2007AA01Z416 (国家高技术研究发展计划(863)); the Knowledge Innovation Program of the Chinese Academy of Sciences under Grant No.CNIC\_QN\_07001 (中国科学院知识创新工程青年人才领域前沿项目); the Beijing New Star Project on Science & Technology of China under Grant No.2007B071 (北京市科技新星计划)

Received 2008-01-22; Accepted 2008-05-05

互联网规模和覆盖面的迅速增长带来了信息超载(information overload)的问题:过量信息同时呈现使得用户无法从中获取对自己有用的部分,信息使用效率反而降低.现有的很多网络应用,比如门户网站、搜索引擎和专业数据索引本质上都是帮助用户过滤信息的手段.然而这些工具只满足主流需求,没有个性化的考虑,仍然无法很好地解决信息超载的问题.推荐系统(recommender system)作为一种信息过滤的重要手段,是当前解决信息超载问题的非常有潜力的方法.推荐系统与以搜索引擎为代表的信息检索(information retrieval)系统最大的区别在于:1) 搜索注重结果(如网页)之间的关系和排序,推荐还研究用户模型(user profile)和用户的喜好,基于社会网络(social network)进行个性化的计算(personalization);2) 搜索的进行由用户主导,包括输入查询词和选择结果,结果不好用户会修改查询再次搜索.而推荐是由系统主导用户的浏览顺序,引导用户发现需要的结果.高质量的推荐系统会使用户对该系统产生依赖.因此,推荐系统不仅能够为用户提供个性化的服务,而且能够与用户建立长期稳定的关系,提高用户忠诚度,防止用户流失.

推荐系统最典型的应用是在 B2C 电子商务领域,具有良好的发展和应用前景,商家根据用户的兴趣、爱好推荐顾客可能感兴趣或满意的商品(如书籍、音像等).顾客的需求通常是不明确的、模糊的,如果商家能够把满足用户模糊需求的商品推荐给用户,就可以把用户的潜在需求转化为现实需求,从而达到提高产品销售量的目的.目前,几乎所有的大型电子商务系统,如 Amazon,eBay 等,都不同程度地使用了各种形式的推荐系统.其中 Amazon 研究电子商务的推荐系统长达 10 年时间.各种提供个性化服务的 Web 站点,如电影、音乐网站,也需要推荐系统的大力支持.表 1 中按照应用领域分类列举了一些典型的商用推荐系统.

Table 1 Mainstream commercialized recommender systems

表 1 主流的商用推荐系统一览表

Field	Systems
	Commercialized recommendation systems
E-Commerce	Amazon.com, eBay, Levis, Ski-europe.com
Web page	Fab, Foxtrot, ifWeb, MEMOIR, METIOREW, ProfBuilder, QuIC, Quickstep, R2P, Siteseer, SurfLen
Music	CDNOW, CoCoA, Ringo, Music.Yahoo.com
Movie	Netfilx.com, Moviefinder.com, MovieLens, Reel.com
News	GroupLens, PHOAKS, P-Tango

在学术界,自 20 世纪 90 年代中期出现第一批关于协同过滤的文章<sup>[1-3]</sup>以来,推荐系统在电子商务、网络经济学和人类社会学等领域一直保持很高的研究热度并逐渐成为一门独立的学科.各种推荐算法涵盖包括认知科学、近似性理论、信息检索<sup>[4]</sup>、管理科学<sup>[5]</sup>、市场营销建模<sup>[6]</sup>等在内的众多研究领域<sup>[7]</sup>.近几年来,国际学术界针对计算机网络信息整合的推荐相关的研究大量出现:1) ACM 设立推荐系统年会(ACM recommender systems);2) 计算机领域的人机交互、数据挖掘和机器学习顶级会议(SIGCHI,KDD,SIGIR,WWW 等)中,推荐算法的文章逐年增加;3) 国际数据分析领域的高阶期刊(如 IEEE Trans. on Knowledge and Data Engineering,ACM Trans. on Information System 等)刊载数篇推荐系统方面的文章.信息领域做推荐系统领先的研究单位(学者)包括:纽约大学(Alexander Tuzhilin)、明尼苏达州立大学的 GroupLens 研究小组(Joseph A. Konstan,John Riedl 等)、美国密歇根大学(Paul Resnick)、卡内基梅隆大学(Jaime Callan)、微软研究院(Ryen W. White)等.其中,美国密歇根大学在 2006 年开授了由 Paul Resnick 主讲的推荐系统的课程.推荐系统,结合社会网络和语义网络的研究,面向互联网发展中出现的新问题和新技术需求,具有广泛的研究和应用前景.

本研究调研了推荐系统在计算机网络和信息领域的主流研究与应用进展.本文第 1 节中给出推荐系统的形式化定义.第 2 节根据推荐算法的类别分类陈述最新的学术进展.第 3 节讨论使用的数据集以及实验评测方法,对当前推荐系统的研究难点进行归纳并对比各种推荐方法的优、缺点.第 4 节对推荐系统有待深入的研究点和发展趋势进行初步预测.

1 推荐系统概念和形式化定义

目前被广泛引用的推荐系统的非形式化概念是 Resnick 和 Varian 在 1997 年<sup>[8]</sup>给出的:“它是利用电子商务网站向客户提供商品信息和**建议**,帮助用户决定应该购买什么产品,模拟销售人员帮助客户完成购买过程”.推

荐有 3 个组成要素:推荐候选对象、用户、推荐方法.通用的推荐系统模型流程如图 1 所示.用户可以向推荐系统主动提供个人偏好信息或推荐请求,或者用户不提供,而是推荐系统主动采集.推荐系统可以使用不同的推荐策略进行推荐,如将采集到的个性化信息和对象数据进行计算得到推荐结果,或者直接基于已建模的知识数据库进行推荐.推荐系统将推荐结果返回给用户使用.

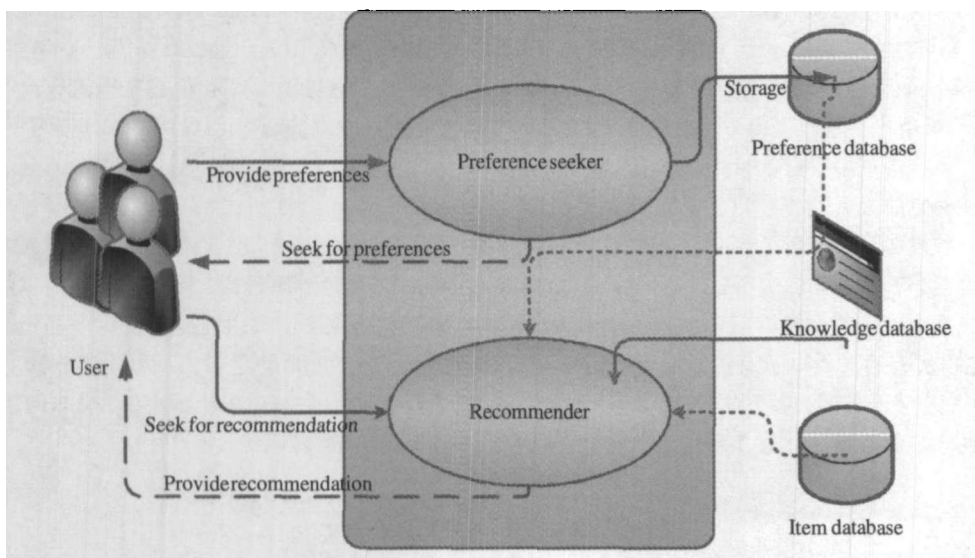


Fig.1 Recommendation system general model

图 1 推荐系统通用模型

此外,文献[7]给出了推荐系统的形式化定义:设  $C$  是所有用户(user)的集合, $S$  是所有可以推荐给用户的对象(object)的集合.实际中, $C$  和  $S$  集合的规模通常很大,如上百万的顾客以及上亿种歌曲等.设效用函数  $u()$  可以计算对象  $s$  对用户  $c$  的推荐度(如提供商的可靠性(vendor reliability)和产品的可得性(product availability)等),即  $u: C \times S \rightarrow R, R$  是一定范围内的全序的非负实数,推荐要研究的问题就是找到推荐度  $R$  最大的那些对象  $s^*$ ,如式(1).

$$\forall c \in C, s^* = \arg \max_{s \in S} u(c, s) \quad (1)$$

用户和对象的度量与采样可以使用不同的属性和特征,这根据实际面对的问题不同而不同.推荐算法研究的中心问题是效用度  $u$  的计算,并非遍历整个  $C \times S$  的整个空间,而是分布到一个流形子空间(manifold)上.对于某个数据集而言,必须先对  $u$  进行外推(extrapolation),也就是说,对象必须具备用户以前作的评分(rating),未评定(unrated)的对象的评分必须先根据已标注的对象进行标注外推后才可以使使用.各类推荐算法在外推和评分预测(rating propagation)上采用了不同的策略,设计了不同的效用函数,这些将在下一节中分类介绍.

## 2 现有的推荐算法

推荐算法是整个推荐系统中最核心和关键的部分,在很大程度上决定了推荐系统类型和性能的优劣.目前,对推荐系统的分类并没有统一的标准,很多学者从不同角度对推荐方法进行了不同的划分<sup>[9-11]</sup>.但主流的推荐方法基本包括以下几种:基于内容推荐、协同过滤推荐、基于知识推荐和组合推荐.本节我们将分类讨论推荐算法的研究成果,下一节我们将讨论这几类推荐算法各自的优、缺点和推荐系统研究的重点、难点问题.

### 2.1 基于内容的推荐

基于内容的推荐(content-based recommendation)是指根据用户选择的对象,推荐其他类似属性的对象作为推荐,属于 Schafer 划分中<sup>[10]</sup>的 Item-to-Item Correlation 方法.这类算法源于一般的信息检索方法<sup>[4]</sup>.不需要依据用户对对象的评价意见.对象使用通过特征提取方法得到的对象内容特征来表示,系统基于用户所评价对象的

特征,学习用户的兴趣,从而考察用户资料与待预测项目相匹配的程度。

对象内容特征(Content(s))的选取在目前的研究中以对象的文字描述为主,比如信息检索中最经典的文本特征是词频-倒排文档频率(term frequency-inverse document frequency,简称 TF-IDF)<sup>[8]</sup>。另一方面,用户的资料模型 ContentBasedProfile(c)取决于所用机器学习方法,常用的有决策树、贝叶斯分类算法<sup>[12,13]</sup>、神经网络、基于向量的表示方法等,数据挖掘领域的众多算法都可以应用。结合对象内容特征和用户资料模型,最终的效用函数可以定义为<sup>[7]</sup>

$$u(c,s) = \text{score}(\text{ContentBasedProfile}(c), \text{Content}(s)) \quad (2)$$

Score 的计算有不同的方法,比如使用最简单的向量夹角余弦的距离计算方法:

$$u(c,s) = \cos(\vec{w}_c, \vec{w}_s) = \sum_{i=1}^K w_{i,c} w_{i,s} / \left( \sqrt{\sum_{i=1}^K w_{i,c}^2} \sqrt{\sum_{i=1}^K w_{i,s}^2} \right) \quad (3)$$

最后得到的  $u$  数值用于排序对象,将最靠前的若干个对象作为推荐。

基于内容推荐的其他研究还包括自适应过滤<sup>[14,15]</sup>和阈值设定<sup>[16,17]</sup>等,前者关注如何通过不断到来的对象增量地计算 ContentBasedProfile(c),使其更加准确;后者研究用户查询文字和对象特征的匹配方法,从而更精确地计算 Content(c)。

## 2.2 协同过滤推荐

协同过滤推荐(collaborative filtering recommendation)技术是推荐系统中最为成功的技术之一,它于 20 世纪 90 年代开始研究并促进了整个推荐系统研究的繁荣。大量论文和研究都属于这个类别。

协同过滤的基本思想是:找到与当前用户  $c_{cur}$  相似(比如兴趣和口味相似)的其他用户  $c_j$ ,计算对象  $s$  对于用户的效用值  $u(c_j,s)$ ,利用效用值对所有  $s$  进行排序或者加权等操作,找到最适合  $c_{cur}$  的对象  $s^*$ 。其基本思想非常易于理解,在日常生活中,我们往往会利用好朋友的推荐来进行一些选择。协同过滤正是把这一思想运用到推荐系统中来,即基于其他用户对某一内容的评价向目标用户进行推荐。

基于协同过滤的推荐系统可以说是从用户的角度进行推荐的,并且是自动的,也就是说,用户所获得的推荐是系统从用户购买或浏览等行为中隐式获得的,不需要用户主动去查找适合自己兴趣的推荐信息,如填写一些调查表格等。其另外一个优点是对推荐对象没有特殊的要求(而基于内容的推荐需要对推荐对象进行特征分析),能够处理非结构化的复杂对象,如音乐、电影等。同时,研究用户之间的关系需要大量的用户访问行为的历史数据,与社会网络研究有交叉点,有丰富的研究基础和广阔的前景。对协同过滤最早的研究有 Grundy system<sup>[18]</sup>,后来的研究成果包括 Tapestry system<sup>[19]</sup>,GroupLens<sup>[20]</sup>,Ringo<sup>[11]</sup>,PHOAKS system<sup>[21]</sup>,Jester system<sup>[22]</sup>等。总体而言,此类推荐算法可以分为两类<sup>[7]</sup>:启发式(heuristic-based or memory-based)方法和基于模型(model-based)的方法。

### 1) 启发式方法

启发式方法<sup>[23,24]</sup>的基本思想是使用与新用户  $c$  相似的用户  $c'$  对一个对象  $s$  的评价来预测  $s$  对新用户  $c$  的效用,进而判断是否推荐  $s$  给  $c$ 。显然,启发式方法的研究主要包括两点:(1) 计算用户之间的相似度;(2) 对所有与用户  $c$  相似的用户  $c'$  对对象  $s$  的评分进行聚合计算,以得到  $s$  对新用户  $c$  的效用的统计预测方法。

在用户相似度  $\text{sim}(c,c')$  这个研究点上,主流的思路是根据用户对同一对象的评分的差异来判断用户兴趣的相似性。评分属于用户的浏览历史行为,可以是打分、观看次数、停留时间等。最基本的两种计算  $\text{sim}(c,c')$  的方法是基于关联的(correlation-based)和基于余弦距离的(cosine-based)方法。基于关联的方法研究用户  $c$  和  $c'$  共同评分过的所有对象的评分相似度来计算关联<sup>[1,3]</sup>。而基于余弦距离的方法直接把评分作为向量来计算余弦距离,进而得到用户相似度<sup>[23,25]</sup>。

统计预测方法的计算公式可以形式化地表示如下<sup>[7]</sup>:

$$r_{c,s} = \text{aggr}_{c',s}(c' \in C) \quad (4)$$

之前的研究设计了很多计算 aggr 的启发式函数,几个比较典型的例子是:

$$r_{c,s} = \frac{1}{n} \sum_{c' \in C} r_{c',s} \quad (5)$$

$$r_{c,s} = k \sum_{c' \in C} \text{sim}(c, c') \times r_{c',s} \quad (6)$$

$$r_{c,s} = \bar{r}_c + k \sum_{c' \in C} \text{sim}(c, c') \times (r_{c',s} - \bar{r}_{c'}) \quad (7)$$

这3类aggr函数都是利用以前用户的评价和用户之间的相似度来启发式地计算效用值.其中,式(5)是最简单的形式;式(6)简单地引入用户相似度加权,是应用最广的方法;考虑到不同的用户在不同情况下作的评分可能有不同的尺度,式(7)提出进行平均归一化的操作以消除这种尺度影响.

除了这两个研究点之外,近年来一些学者同时也发展了其他启发式方法,以提高启发式推荐的性能,如缺省投票(default voting)、用户倒排评分(inverse user frequency)、实例扩展(case amplification)<sup>[23]</sup>和主流加权预测(weighted-majority prediction)<sup>[24]</sup>等.

## 2) 基于模型的方法

这类方法利用用户  $c$  对众多对象的评分来学习一个  $c$  的模型(model)<sup>[26-30]</sup>,然后使用概率方法对新的对象  $s$  的推荐效用进行预测.文献[7]对这种方法的形式化描述如式(8)所示:

$$r_{c,s} = E(r_{c,s}) = \sum_{i=0}^n i \times \Pr(r_{c,s} = i | r_{c,s'}, s' \in S_c) \quad (8)$$

这样,基于模型的方法把一个用户归类到一种模型下或者一个类型中.其他的算法还包括利用机器学习方法<sup>[26]</sup>和统计模型<sup>[30]</sup>、贝叶斯模型<sup>[31]</sup>、概率相关模型<sup>[27]</sup>、线性回归模型<sup>[25]</sup>和最大熵模型<sup>[32]</sup>. Shani 在文献[33]中还把推荐选择看作序列决策问题(sequential decision problem),使用马尔可夫决策过程方法(Markov decision processes)加以解决.图模型方法,包括概率隐形式语义分析(probabilistic latent semantic analysis)<sup>[28]</sup>和 LDA(latent dirichlet allocation)<sup>[29]</sup>,也应用于协同过滤推荐算法的研究.

## 2.3 基于知识的推荐

基于知识的推荐(knowledge-based recommendation)<sup>[34]</sup>在某种程度上可以看成是一种推理(inference)技术.它不是建立在用户需要和偏好基础上推荐的,而是利用针对特定领域制定规则(rule)来进行基于规则和实例的推理(case-based reasoning).例如,文献[34]中利用饭店的菜式方面的效用知识,推荐饭店给顾客.效用知识(functional knowledge)是一种关于一个对象如何满足某一特定用户的知识,因而能够解释需求和推荐的关系,用于推荐系统.效用知识在推荐系统中必须以机器可读的方式存在(ontology 本体知识库),例如 quickstep and foxtrot systems<sup>[35]</sup>使用关于学术论文主题的 ontology 本体知识库向读者作推荐.

## 2.4 组合推荐

组合推荐(hybrid recommendation)的一个最重要原则就是通过组合后应能避免或弥补各自推荐技术的弱点(见第3.4节).研究和应用最多的是内容推荐和协同过滤推荐的组合<sup>[9,36-38]</sup>.尽管从理论上有很多种推荐组合方法,但不同的组合思路适用于不同的应用场景.我们将研究人员提出的组合思路大致分为如下3类:

- 1) 后融合:融合两种或两种以上的推荐方法各自产生的推荐结果.如使用基于内容的方法和协同过滤方法分别得到推荐列表,融合列表的结果决定最后推荐的对象.
- 2) 中融合:以一种推荐方法为框架,融合另一种推荐方法.如以基于内容的方法为框架,融合协同过滤的方法,或者以协同过滤的方法为框架,融合基于内容的方法.
- 3) 前融合:直接融合各种推荐方法.如将基于内容和协同过滤的方法整合到一个统一的框架模型下.

### 2.4.1 后融合组合推荐

在后融合组合推荐中,最简单的做法就是分别用基于内容的方法和协同过滤推荐方法去产生一个推荐预测结果,然后用某种方法组合其结果.文献[37]使用了评分结果的线性组合,而文献[38]使用了投票机制来组合这些推荐结果.除此之外,也可以分别考察两个推荐列表,判断使用其中的哪个推荐结果.比如, Daily Learner system<sup>[39]</sup>计算推荐结果的可信度,然后选择一个列表的结果.这种结果层次上的融合我们称为后融合组合推荐.

## 2.4.2 中融合组合推荐

目前,中融合的组合推荐主要有两种,以基于内容的方法为框架,融合协同过滤的方法和以协同过滤的方法为框架,融合基于内容的方法.前者利用降维技术把基于内容的对象特征进行精简化.例如,文献[40]使用了LSI(latent semantic indexing)算法,在基于内容的框架中使用精化的用户特征向量.后者为了克服协同过滤的稀疏问题(详见第3.3节),把用户当作对象,使用基于内容的特征提取方法把用户本身的特征(如年龄、工作情况等人口统计学特征(demographic features))使用到相似度计算中,而不是仅仅依赖用户的点击行为.Good等人在文献[41]中引入多种不同的用户描述符来归类用户,挖掘用户的内在联系,从而得到更好的推荐效果.文献[42]使用独立的基于内容的特征来补偿用户提供的简单的 rating,也属于此类方法.

## 2.4.3 前融合组合推荐

近年来,这类推荐方法最受学者的关注.在文献[36]中,研究者把用户的年龄和电影的类型放到一个统一的分类器中训练学习.另外一种方法<sup>[43]</sup>使用了贝叶斯混合效果回归模型,并通过马尔可夫蒙特卡洛方法得到这个模型的参数.文献[43]将用户和对象的特征都放到一个统计模型下来计算效用函数,研究者使用用户属性  $z$ 、对象属性  $w$  及其交互关系(如选择关系) $x$  来计算效用  $r$ .对象  $j$  对于用户  $i$  的效用值  $r_{ij}$  计算式可以表示为

$$r_{ij} = x_{ij}\mu + z_i\gamma_j + w_j\lambda_i + e_{ij}, \text{ Where } e_{ij} \sim N(0, \sigma^2), \lambda_i \sim N(0, \lambda), \gamma_j \sim N(0, \tau) \quad (9)$$

这其中的3种正态分布的变量分别用于描述数据的噪声、用户属性的异质性和对象属性的异质性.式(9)表述效用值是由这几个因素共同决定的.这3种分布的3个参数由马尔可夫蒙特卡洛方法估算得到.

近年来,一些方法比较的工作<sup>[9,42,38]</sup>讨论并实验了各种方法与组合策略,得出结论:组合策略能够取得比纯基于内容或协同过滤方法更好的效果.这种在方法层次上融合的方法我们称为前融合组合推荐.

# 3 推荐系统的重点、难点问题和主流算法对比

## 3.1 推荐系统的评测标准数据集

推荐系统学术研究常用的数据集包括:

### 1) MovieLens<sup>[44]</sup>

MovieLens 数据集中,用户对自己看过的电影进行评分,分值为1~5.MovieLens 包括两个不同大小的库,适用于不同规模的算法.小规模库是943个独立用户对1682部电影作的10000次评分的数据;大规模库是6040个独立用户对3900部电影作的大约100万次评分.

### 2) EachMovie<sup>[45]</sup>

HP/Compaq 的 DEC 研究中心曾经在网上架设 EachMovie 电影推荐系统对公众开放.之后,这个推荐系统关闭了一段时间,其数据作为研究用途对外公布,MovieLens 的部分数据就是来自于这个数据集的.这个数据集有72916个用户对1628部电影进行的2811983次评分.早期大量的协同过滤的研究工作都是基于这个数据集的.2004年HP重新开放 EachMovie,这个数据集就不提供公开下载了.

### 3) BookCrossing<sup>[46]</sup>

这个数据集是网上的 Book-Crossing 图书社区的278858个用户对271379本书进行的评分,包括显式和隐式的评分.这些用户的年龄等人口统计学属性(demographic feature)都以匿名的形式保存并供分析.这个数据集是由Cai-Nicolas Ziegler 使用爬虫程序在2004年从Book-Crossing 图书社区上采集的.

### 4) Jester Joke<sup>[22]</sup>

Jester Joke 是一个网上推荐和分享笑话的网站.这个数据集有73496个用户对100个笑话作的410万次评分.评分范围是-10~10的连续实数.这些数据是由加州大学伯克利分校的Ken Goldberg 公布的.

### 5) Netflix<sup>[47]</sup>

这个数据集来自于电影租赁网址Netflix 的数据库.Netflix 于2005年底公布此数据集并设立百万美元的奖金(netflix prize<sup>[47]</sup>),征集能够使其推荐系统性能上升10%的推荐算法和架构.这个数据集包含了

480 189 个匿名用户对大约 17 770 部电影作的大约 10 亿次评分.

6) Usenet Newsgroups<sup>[48]</sup>

这个数据集包括 20 个新闻组的用户浏览数据.最新的应用是在 KDD 2007 上的论文<sup>[49]</sup>.新闻组的内容和讨论的话题包括计算机技术、摩托车、篮球、政治等.用户们对这些话题进行评价和反馈.

7) UCI 知识库<sup>[50]</sup>

UCI 知识库是 Blake 等人在 1998 年开放的一个用于机器学习和评测的数据库,其中存储大量用于模型训练的标注样本,在文献[49]中被用于推荐系统的性能测试数据.

### 3.2 推荐系统的性能评测方法

推荐系统的性能指标一般有推荐的效果/精确度(effectiveness)和推荐的效率(efficiency),使用的指标有 mean absolute error(MAE),root mean squared error(RMSE)和 correlation.由于不同的研究工作针对不同的问题,使用不同的数据集,所以具体评测方法变化很大.比较普遍的评测方法来自于机器学习等领域的一般方法,比如数据集被分割为训练集(probe set)和测试集(quiz set).推荐算法的模型在训练集上进行学习和参数调整,然后在测试集上计算精确度和运行效率,从而达到评测目的.文献[23]使用两种评测方法来比较几种协同过滤的算法性能,第 1 种评测得到每次推荐绝对误差的平均值,第 2 种评测计算整个推荐列表的推荐精度.

### 3.3 推荐系统的重点、难点问题

随着近年来对推荐系统研究的开展,很多研究中的重点、难点问题得到研究者的关注和共识<sup>[7]</sup>,主要包括:

1) 特征提取问题

虽然在信息检索中,文本等对象特征提取技术已经很成熟,但是推荐系统的对象不一定具有文本特征或者文本不足以作为描述<sup>[1]</sup>,此时特征的选择出现了问题.尤其是网络上广泛存在的多媒体数据如音乐、视频、图像等,自动化的特征提取方法需要结合多媒体内容分析领域的相关技术.另一个问题是特征的区分性问题,大规模数据情况下不同对象的特征错配会影响系统性能.

2) 模型过拟合问题(可扩展性问题)

推荐系统中推荐算法无法完全掌握用户每个方面的兴趣和需求,因为用户之前没有对足够多类别的对象进行评价.过拟合现象是指系统推荐给用户的对象与用户刚刚看过的不是太相似,就是太不相关.模型过拟合(过学习)的问题本质上来源于数据的不完备性,这在实际应用是无法完全避免的.在信息检索领域这类问题广泛存在,解决的主要方法是引入随机性,使算法收敛到全局最优或者逼近全局最优.随机方法包括遗传算法<sup>[51]</sup>等.Daily Learner 相关的文献[15,39]针对这个问题考察了被推荐的对象的相关性(relevant)和冗余性(redundancy),认为被推荐的对象首先不能与用户看过的对象重复(冗余),其次必须有相关性以相互联系.推荐的多样性是必不可缺的.

3) 新用户问题

系统没有存储或者存储很少新用户的信息,包括查看对象的历史记录和新用户对对象的评分,基于模型的方法无法获得训练数据而基于规则的方法难以进行推理.近期一些研究特别针对这个问题提出了解决方法.文献[52,53]利用对象熵(entropy)、受欢迎程度(popularity)、用户个性属性等来改进效果.

4) 新对象问题

新用户和新对象问题都属于冷启动问题.在推荐系统尤其是协同过滤系统中,新对象加入数据库后必须等待一段时间才有用户查看并进行评价(点击、打分、评论等都是评价的手段).在评价达到一定数量之前无法对此对象进行分析和推荐.不同于新用户问题,这类问题一般考虑使用组合推荐的方法来应对.

5) 稀疏问题

在任何大型的推荐系统中,对于一个用户,总有大量的对象没有经过用户的评价或者查看,而且这类数据常常比已经有此用户评价的数据量更大<sup>[7]</sup>.用户之间由于选择的差异性非常大造成稀疏情况,即任意两个用户的评分差别都非常大.文献[38]提出初步的解决方法,将用户的年龄、国籍、性别等个人信息

增加作为用户相似度计算的根据,称为基于人口统计学的过滤方法(demographic filtering).文献[26,54]使用主分量分析(SVD)降维方法尝试把稀疏的关系矩阵降维到低维,以得到用户之间潜在的关系.

3.4 各类推荐方法的对比

各类推荐方法都有其各自的优、缺点,针对不同的数据集,效果也有所不同.每种方法因为算法本身的特征可能不适合在所有数据集上作推荐.如在基于内容的推荐方法中,自动化的特征提取方法很难应用于多媒体数据,即使在容易提取特征的文本数据的情况下,也无法仅仅通过词频统计的方式区分文档质量<sup>[4]</sup>.除此之外,为用户推荐的内容仅限于与该用户曾经选择的对象相似的对象,结果多样性差.而对于没有选择过任何对象的新用户,推荐尤其困难.协同过滤的方法从某种程度上克服了基于内容方法自动化程度低、推荐结果不丰富等弊端.但是,协同过滤是基于大量历史数据集的,因而存在稀疏问题和冷启动问题.在冷启动方面,由于协同过滤是依靠人与人之间选择内容的相似度进行推荐的,因此,与基于内容的方法相比,不但存在新用户问题<sup>[53]</sup>,而且还存在新对象问题,即刚刚加入的对象如果没有被任何人选择过,就很难被推荐<sup>[7]</sup>.基于知识的推荐是一种静态的推荐方法,不存在冷启动和稀疏问题,但知识很难建模.组合推荐策略由于组合方式不同,其性能特点差异很大,故不在此讨论范围内.几种推荐方法的优、缺点具体比较见表 2.

Table 2 Comparison of typical recommendation algorithms  
表 2 典型推荐算法对比

Recommendation algorithms	Advantages	Disadvantages
Content based	Recommendation result is intuitive and easy to interpret; No need for users' access history data; No new item problem and no sparsity problem; Supported by the mature technology of classification learning.	Limited by the features extraction methods; New user problem; The training of classifier needs massive data; Poor scalability.
Collaboration filtering	No need for professional knowledge; Performance improving as the increasing of the user number; Automatic; Easy to find user's new interesting point; Complex unstructured item can be processed. eg. Music, Video, etc.	Sparsity problem; Poor scalability; New user and new item problem; The recommendation quality limited by the history data set.
Knowledge based	Can map users' needs to product; Non-Product attribute is considered.	Knowledge is difficult to obtain and model; Static without flexibility.

4 推荐系统研究发展的热点方向

推荐系统的研究发展多年,曾经一度进入低潮期.近年来,机器学习、大规模网络应用需求和高性能计算的发展推动了这个研究领域的新进展,可以深入并可能取得成果的方向很多,主要包括:

1) 引入更精确适用的用户和对象特征(new profiles of user and item)<sup>[7]</sup>

针对特定问题适用的用户和对象特征通常可以作为模型训练的样本.典型的协同过滤方法<sup>[1-3]</sup>并没有使用用户和对象特征,而是利用用户的评分.文献[13,55]只是使用简单的特征,如对象描述的关键词和用户的统计学特征等.而结合数据挖掘的高层特征一般是基于网络上下文的分析的<sup>[56]</sup>,比如发现用户浏览网页和对象的时序模式.这类方法需要精准的用户浏览历史数据和先进的数据挖掘算法,尚未在基于内容和协同过滤的研究中广泛采用.

2) 推荐的多维度研究<sup>[7]</sup>

当前的大部分研究都是基于对象-用户的二维度量空间的,未考虑相关信息(contextual information).然而,用户对对象的评价和选择常常由很多环境因素来决定,比如某个对象在特定时段很流行,用户在某个地方浏览对象的时候偏向于选择某类对象等.环境因素是无法从用户和对象的自有特征得到的,正如文献[57,58]所指出的,推荐使用的特征维度有必要根据具体问题来增加.文献[57]提出在多维度量空间上来定义效用函数  $u()$ ,以电影推荐为例:除了看电影的人的特性和电影本身的固有描述特征以外,还有环境因素  $d$  来决定如何计算效用函数  $u()$ , $d$  包括:1) 是一个人看电影还是和其他人一起看;2) 是去电影



院看还是在网上看等;3) 看电影的时间段.基于这种多维特征的思想,文献[43]提出使用贝叶斯模型的方法进一步扩展了这个思路.问题在于,维度的扩展与应用场景相关,其可扩展性还有待进一步改进.

### 3) 推荐系统安全问题

协同过滤类别的推荐算法必须使用用户的行为历史记录,但用户出于保护隐私的考虑常常无法提交完整而正确的信息给推荐系统.文献[59,60]认为良好的隐私保护机制是推荐系统获取优质数据和忠实用户的关键之一,并提出了相应的算法.同时,用户的信用度(reputation)也是推荐算法的重要参考数据.文献[61]提出两种信用计算模型并用于提高推荐精度.另外,熟悉推荐算法的攻击者恶意利用捏造的评分和对象属性等数据欺骗推荐系统,达到被频繁推荐的目的.文献[62,63]中定义这类行为为用户欺诈(shilling attack),并设计算法监控评分的时序模式,去除不正常的对象与用户.

### 4) 相关反馈研究(relevance feedback)与侵袭性问题

一般的推荐系统大多需要用户依据自身对推荐对象的喜好程度,提供适当的评比反馈信息,这种评比方式称为相关反馈(relevance feedback).相关反馈可以分为显性反馈(explicit feedback)和隐性反馈(implicit feedback).目前部分推荐系统主要使用显性反馈的方式,即需要用户每次查看对象都进行评分或者其他操作.一些研究<sup>[20,35,64]</sup>中使用的用户反馈的方法也属于显性反馈.使用这种方式来取得评价的方法对于用户不是很友好,具有很大的侵袭性.隐性反馈使用无须用户评价的用户数据采集方法.在文献[65]的研究中列举了12种用户在网络上的浏览行为,Oard和Kim<sup>[66]</sup>延伸了文献[65]的研究,进一步将这些行为分为3类,并增加了一种新的行为类型.文献[67]针对隐性反馈的可靠性进行了实验,结果显示“使用者的浏览时间”和“使用者是否喜欢该文章”有正向潜在关系.解决侵袭性问题的另一个研究方向是研究用户能够接受的评价方式是什么,比如能够有耐心进行几次评分.MovieLens系统<sup>[44]</sup>利用固定负担模型这种最简单的算法来计量用户评价的负担,将侵袭性问题转化为最优化问题来研究.

### 5) 推荐算法的评价准则<sup>[7]</sup>

有效性和时间消耗作为推荐系统的重要指标,其评价准则的设计一直是一个热点,但是没有统一的结论.文献[55,68-70]都研究过评测推荐系统性能的准则,包括查全率(coverage)和查准率(accuracy)的评测.另外,信息检索领域使用的F-measure和ROC(receiver operating characteristic)指标也可以用于评测.第4.2节给出了部分研究工作中使用过的推荐效果度量方法.然而这些基于用户评分的数据和评测的方法存在固有的缺陷,比如用户的评分常常是针对他喜欢的对象,而其他对象被访问的概率则很小.评价准则必须考虑到人类行为的特点.同时,如果召集若干志愿者来做实时的实验,不仅样本量小,而且耗费的时间、人力也很昂贵.这是一个很有研究意义的方向.

### 6) 基于复杂网络理论及图方法的推荐系统

近年来,机器学习、理论物理和复杂网络系统的研究者从图和动态复杂网络(graph and dynamic network)的视角,开始关注推荐系统中的大规模网络智能挖掘.文献[71]将网络视频推荐问题转化为热量散播平衡态网络上的谱图分割(spectral graph partitioning)问题,通过设计长尾发现(long tail discovering)的推荐策略引导用户发现潜在的感兴趣的网络视频.

### 7) 社会分化和推荐逻辑空间的巴尔干化现象(balkanization):

以Internet为代表的信息和通信技术消除了地理壁垒,即传统的地理空间巴尔干(即地域性限制).然而却不可避免地生成了逻辑空间巴尔干<sup>[72]</sup>.在推荐系统中根据个体兴趣、学科专业、社会地位以及观点进行推荐时尤其容易导致这种地域性限制和社会分化的出现.本质上,个体偏好关系是导致生成逻辑空间巴尔干的原因.推荐和个性化技术所产生的一系列社会学问题,也是很重要的一个研究方向.网络推荐中体现的其他社会学问题还包括从众心理和行为研究,推荐与主流兴趣形成的相互作用等.

## 5 结 论

在互联网的迅猛发展下,随着信息过载问题的逐年升温,互联网用户对信息需求的日益膨胀,推荐系统在

个领域的数字化进程中扮演着越来越重要的角色.在过去的数十年中,推荐系统在学术研究、工业界各种应用上取得了长足的进步.然而,现有的推荐算法仍然存在特征提取、冷启动、过拟合、稀疏问题,需要不断完善和解决.同时,多维度推荐、相关反馈、评价准则、安全性以及推荐社会学等仍然是当前进行深入研究和扩展的热点方向.伴随着这些问题的逐渐解决,推荐系统将在互联网领域为用户提供更加便捷、有效的用户信息获取体验.

## References:

- [1] Shardanand U, Maes P. Social information filtering: Algorithms for automating "Word of Mouth". In: Proc. of the Conf. on Human Factors in Computing Systems. New York: ACM Press, 1995. 210–217.
- [2] Hill W, Stead L, Rosenstein M, Furnas G. Recommending and evaluating choices in a virtual community of use. In: Proc. of the Conf. on Human Factors in Computing Systems. New York: ACM Press, 1995. 194–201.
- [3] Resnick P, Iakovou N, Sushak M, Bergstrom P, Riedl J. GroupLens: An open architecture for collaborative filtering of netnews. In: Proc. of the Computer Supported Cooperative Work Conf. New York: ACM Press, 1994. 175–186.
- [4] Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval. New York: Addison-Wesley Publishing Co., 1999.
- [5] Murthi BPS, Sarkar S. The role of the management sciences in research on personalization. *Management Science*, 2003,49(10): 1344–1362.
- [6] Smith SM, Swinyard WR. Introduction to marketing models. 1999. <http://marketing.byu.edu/htmlpages/courses/693r/modelsbook/preface.html>
- [7] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowledge and Data Engineering*, 2005,17(6):734–749.
- [8] Resnick P, Varian HR. Recommender systems. *Communications of the ACM*, 1997,40(3):56–58.
- [9] Balabanovic M, Shoham Y. Fab: Content-Based, collaborative recommendation. *Communications of the ACM*, 1997,40(3):66–72.
- [10] Schafer JB, Konstan J, Riedl J. Recommender systems in e-commerce. In: Proc. of the 1st ACM Conf. on Electronic Commerce. New York: ACM Press, 1999. 158–166.
- [11] Terveen L, Hill W. Beyond recommender systems: Helping people help each other. In: Carroll J, ed. HCI in The New Millennium. New York: Addison-Wesley Publishing Co., 2001. 1–21.
- [12] Mooney RJ, Bennett PN, Roy L. Book recommending using text categorization with extracted information. In: Proc. of the AAAI'98/ICML'98 Workshop on Learning for Text Categorization. Madison: AAAI Press, 1998. 49–54.
- [13] Pazzani M, Billsus D. Learning and revising user profiles: The identification of interesting Web sites. *Machine Learning*, 1997, 27(3):313–331.
- [14] Somlo G, Howe A. Adaptive lightweight text filtering. In: Proc. of the 4th Int'l Symp. on Intelligent Data Analysis. Berlin, Heidelberg: Springer-Verlag, 2001. 319–329.
- [15] Zhang Y, Callan J, Minka T. Novelty and redundancy detection in adaptive filtering. In: Proc. of the 25th Annual Int'l ACM SIGIR Conf. New York: ACM Press, 2002. 81–88.
- [16] Robertson S. Threshold setting and performance optimization in adaptive filtering. *Information Retrieval*, 2002,5(2-3):239–256.
- [17] Zhang Y, Callan J. Maximum likelihood estimation for filtering thresholds. In: Proc. of the 24th Annual Int'l ACM SIGIR Conf. New York: ACM Press, 2001. 294–302.
- [18] Rich E. User modeling via stereotypes. *Cognitive Science*, 1979,3(4):329–354.
- [19] Goldberg D, Nichols D, Oki BM, Terry D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992,35(12):61–70.
- [20] Konstan JA, Miller BN, Maltz D, Herlocker JL, Gordon LR, Riedl J. GroupLens: Applying collaborative filtering to usenet news. *Communications of the ACM*, 1997,40(3):77–87.
- [21] Terveen L, Hill W, Amento B, McDonald D, Creter J. PHOAKS: A system for sharing recommendations. *Communications of the ACM*, 1997,40(3):59–62.
- [22] Goldberg K, Roeder T, Gupta D, Perkins C. Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 2001,4(2):133–151.

- [23] Breese JS, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. Technical Report, MSR-TR-98-12, Redmond: Microsoft Research, 1998.
- [24] Delgado J, Ishii N. Memory-Based weighted-majority prediction for recommender systems. In: Proc. of the ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation. New York: ACM Press, 1999.
- [25] Sarwar B, Karypis G, Konstan J, Riedl J. Item-Based collaborative filtering recommendation algorithms. In: Proc. of the 10th Int'l WWW Conf. New York: ACM Press, 2001. 285–295.
- [26] Billsus D, Pazzani M. Learning collaborative information filters. In: Proc. of Int'l Conf. on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 1998. 46–54.
- [27] Getoor L, Sahami M. Using probabilistic relational models for collaborative filtering. In: Proc. of the Workshop Web Usage Analysis and User Profiling. 1999.
- [28] Hofmann T. Collaborative filtering via Gaussian probabilistic latent semantic analysis. In: Proc. of the 26th Int'l ACM SIGIR Conf. New York: ACM Press, 2003. 259–266.
- [29] Marlin B. Modeling user rating profiles for collaborative filtering. In: Proc. of the 17th Annual Conf. on Neural Information Processing Systems. Cambridge: MIT Press, 2003. 627–634.
- [30] Ungar LH, Foster DP. Clustering methods for collaborative filtering. In: Proc. of the Workshop on Recommendation Systems. Menlo Park: AAAI Press, 1998. 112–125.
- [31] Chien YH, George EI. A Bayesian model for collaborative filtering. In: Proc. of the 7th Int'l Workshop on Artificial Intelligence and Statistics. San Francisco: Morgan Kaufmann, 1999.
- [32] Pavlov D, Pennock D. A maximum entropy approach to collaborative filtering in dynamic, sparse, high-dimensional domains. In: Proc. of the 16th Annual Conf. on Neural Information Processing Systems. 2002.
- [33] Shani G, Brafman R, Heckerman D. An MDP-based recommender system. The Journal of Machine Learning Research, 2005,6:1265–1295.
- [34] Burke R. Knowledge-Based recommender systems. Encyclopedia of Library and Information Systems, 2000,69(32):180–200.
- [35] Middleton SE, Shadbolt NR, de Roure DC. Ontological user profiling in recommender systems. ACM Trans. on Information Systems, 2004,22(1):54–88.
- [36] Basu C, Hirsh H, Cohen W. Recommendation as classification: Using social and content-based information in recommendation. In: Proc. of the AAAI'98. Menlo Park: AAAI Press. 1998. 714–720.
- [37] Claypool M, Gokhale A, Miranda T, Murnikov P, Netes D, Sartin M. Combining content-based and collaborative filters in an online newspaper. In: Proc. of the ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation. New York: ACM Press, 1999.
- [38] Pazzani M. A framework for collaborative, content-based, and demographic filtering. Artificial Intelligence Review, 1999,13(5-6): 393–408.
- [39] Billsus D, Pazzani M. User modeling for adaptive news access. User Modeling and User-Adapted Interaction, 2000,10(2-3): 147–180.
- [40] Soboroff I, Nicholas C. Combining content and collaboration in text filtering. In: Proc. of the Int'l Joint Conf. on Artificial Intelligence Workshop: Machine Learning for Information Filtering. Stockholm, 1999. 86–91.
- [41] Good N, Schafer JB, Konstan JA, Borchers A, Sarwar B, Herlocker JL, Riedl J. Combining collaborative filtering with personal agents for better recommendations. In: Proc. of the 16th National Conf. on Artificial Intelligence. Menlo Park: AAAI Press, 1999. 439–446.
- [42] Melville P, Mooney RJ, Nagarajan R. Content-Boosted collaborative filtering for improved recommendations. In: Proc. of the 18th National Conf. on Artificial Intelligence. Menlo Park: American Association for Artificial Intelligence, 2002. 187–192.
- [43] Ansari A, Essegai S, Kohli R. Internet recommendations systems. Journal of Marketing Research, 2000,37(3):363–375.
- [44] Miller BN, Albert I, Lam SK, Konstan JA, Riedl J. MovieLens unplugged: Experiences with an occasionally connected recommender system. In: Proc. of the Int'l Conf. on Intelligent User Interfaces. New York: ACM Press, 2003. 263–266.
- [45] EachMovie. EachMovie collaborative filtering data set. 1997. <http://research.compaq.com/SRC/eachmovie>
- [46] <http://www.bookcrossing.com/>

- [47] <http://www.netflix.com/>
- [48] <http://www.newsgroups.com/>
- [49] Ding C, Jin R, Li T, Simon HD. A learning framework using Green's function and kernel regularization with application to recommender system. In: Proc. of the 13th ACM SIGKDD National Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2007. 260–269.
- [50] <http://archive.ics.uci.edu/ml/>
- [51] Sheth B, Maes P. Evolving agents for personalized information filtering. In: Proc. of the 9th IEEE Conf. on Artificial Intelligence for Applications. 1993. 345–352.
- [52] Yu K, Schwaighofer A, Tresp V, Xu X, Kriegel HP. Probabilistic memory-based collaborative filtering. IEEE Trans. on Knowledge and Data Engineering, 2004,16(1):56–69.
- [53] Rashid AM, Albert I, Cosley D, Lam SK, McNee SM, Konstan JA, Riedl J. Getting to know you: Learning new user preferences in recommender systems. In: Gil Y, Leake DB, eds. Proc. of the Int'l Conf. on Intelligent User Interfaces. New York: ACM Press, 2002. 127–134.
- [54] Sarwar B, Karypis G, Konstan J, Riedl J. Application of dimensionality reduction in recommender systems—A case study. In: Proc. of the ACM WebKDD Workshop. 2000.
- [55] Mooney RJ, Roy L. Content-Based book recommending using learning for text categorization. In: Proc. of the ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation. New York: ACM, 1999. 195–204.
- [56] Li J, Zaiane OR. Combining usage, content, and structure data to improve Web site recommendation. In: Proc. of the 5th Int'l Conf. on Electronic Commerce and Web Technologies. Berlin: Springer-Verlag, 2004. 305–315.
- [57] Adomavicius G, Tuzhilin A. Multidimensional recommender systems: A data warehousing approach. In: Proc. of the 2nd Int'l Workshop on Electronic Commerce. LNCS 2232, London: Springer-Verlag, 2001. 180–192.
- [58] Adomavicius G, Sankaranarayanan R, Sen S, Tuzhilin A. Incorporating contextual information in recommender systems using a multidimensional approach. ACM Trans. on Information Systems, 2005,23(1):103–145.
- [59] Canny J. Collaborative filtering with privacy via factor analysis. In: Proc. of the 25th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 2002. 238–245.
- [60] Polat H, Du WL. Privacy-Preserving collaborative filtering using randomized perturbation techniques. In: Proc. of the 3rd IEEE Int'l Conf. on Data Mining. Washington: IEEE Computer Society, 2003. 625–628.
- [61] O'Donovan J, Smyth B. Trust in recommender systems. In: Proc. of the 10th Int'l Conf. on Intelligent User Interfaces. New York: ACM Press, 2005. 167–174.
- [62] Chirita PA, Nejdl W, Zamfir C. Preventing shilling attacks in online recommender systems. In: Proc. of the 7th Annual ACM Int'l Workshop on Web Information and Data Management. New York: ACM Press, 2005. 67–74.
- [63] Zhang S, Chakrabarti A, Ford J, Makedon F. Attack detection in time series for recommender systems. In: Proc. of the 12th ACM Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2006. 809–814.
- [64] Caglayan A, Snorrason M, Jacoby J, Mazzu J, Jones R, Kumar K. Learn sesame—A learning agent engine. Applied Artificial Intelligence, 1997,11:393–412.
- [65] Nichols DM. Implicit rating and filtering. In: Proc. of the 5th DELOS Workshop on Filtering and Collaborative Filtering. Budapest: ERCIM, 1997. 31–36.
- [66] Oard DW, Kim J. Implicit feedback for recommender systems. In: Kautz H, ed. Proc. of the AAAI Workshop on Recommender Systems. Madison: AAAI Press, 1998. 81–83.
- [67] Morita M, Shinoda Y. Information filtering based on user behavior analysis and best match text retrieval. In: Proc. of the 17th Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: Springer-Verlag, 1994. 272–281.
- [68] Herlocker JL, Konstan JA, Borchers A, Riedl J. An algorithmic framework for performing collaborative filtering. In: Proc. of the 22nd Annual Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM Press, 1999. 230–237.

- [69] Herlocker JL, Konstan JA, Terveen LG, Riedl J. Evaluating collaborative filtering recommender systems. *ACM Trans. on Information Systems*, 2004,22(1):5-53.
- [70] Yang Y, Padmanabhan B. Evaluation of online personalization systems: A survey of evaluation schemes and knowledge-based approach. *Journal of Electronic Commerce Research*, 2005,6(2):112-122.
- [71] Wu X, Zhang YD, Guo JB, Li JT. Web video recommendation and long tail discovering. In: *Proc. of the 2008 IEEE Int'l Conf. on Multimedia & Expo*. Hoboken: Wiley-IEEE Press, 2008.
- [72] Sunstein C. *Republic.com*. Princeton: Princeton University Press, 2002.



许海玲(1982-),女,天津人,硕士,主要研究领域为互联网寻址系统,计算机网络应用.



吴满(1982-),男,博士生,主要研究领域为网络多媒体数据挖掘,计算机视觉,机器学习.



李晓东(1976-),男,博士,副研究员,主要研究领域为计算机网络与通信.



阎保平(1950-),女,博士,研究员,博士生导师,CCF 高级会员,主要研究领域为计算机网络,信息系统工程,大型数据库技术.