

## 推荐系统研究进展\*

朱扬勇<sup>1,2</sup> 孙 婧<sup>1,2+</sup>

1. 上海市数据科学重点实验室(复旦大学),上海 201203
2. 复旦大学 计算机科学技术学院,上海 201203

### Recommender System: Up to Now\*

ZHU Yangyong<sup>1,2</sup>, SUN Jing<sup>1,2+</sup>

1. Shanghai Key Laboratory of Data Science, Fudan University, Shanghai 201203, China
  2. School of Computer Science, Fudan University, Shanghai 201203, China
- + Corresponding author: E-mail: [jingsun@fudan.edu.cn](mailto:jingsun@fudan.edu.cn)

ZHU Yangyong, SUN Jing. Recommender system: up to now. Journal of Frontiers of Computer Science and Technology, 2015, 9(5) :513-525.

**Abstract:** Recommender system is the product of cyber age today. There have been many achievements in research and application. This paper makes a comprehensive survey of the recommender system. It proposes three research phases, and points out the milestone events in each stage of recommender system development. In the age of big data, exploiting recommendation in the perspective of data, this paper classifies the recommender system into seven main classes according to the different data used in recommendation, and analyzes and comments the recommended models used in each classification and their advantages and disadvantages. Exploiting big data in the perspective of recommendation, this paper proposes that making recommendation based on big data is one of the promising research directions. Finally, this paper compares the evaluation metrics of recommendation, and gives future research directions.

**Key words:** recommender system; personalization; collaborative filtering; big data

**摘 要:** 推荐系统(recommender system, RS)是当今网络时代的产物,在技术研究和应用方面取得了许多成果。综述了推荐系统领域的研究状况和进展,提出了3个研究阶段,并指出了每个阶段标志性意义的事件。在当前大数据环境下,从数据的角度看推荐,提出了推荐系统新的分类方法,即根据推荐时所使用的数据不同分

---

\* The National Natural Science Foundation of China under Grant Nos. 61170096, 71331005 (国家自然科学基金); the Shanghai Foundation for Development of Science and Technology under Grant Nos. 13dz2260200, 13511504300, 14511107302 (上海市科学技术发展基金).

Received 2014-12, Accepted 2015-03.

CNKI网络优先出版 2015-03-31, <http://www.cnki.net/kcms/detail/11.5602.TP.20150331.1426.001.html>

为7种类别,同时指出了每个类别使用了哪些推荐模型及其优缺点。提出了在大数据环境下进行推荐是未来推荐系统研究的一个大方向,分析了推荐视角下的大数据机制。最后比较和总结了推荐系统的评价指标,给出了未来的主要研究方向和可能的突破点。

关键词 推荐系统;个性化;协同过滤;大数据

文献标志码:A 中图分类号:TP301

## 1 引言

随着互联网、移动技术的发展,推荐系统已经渗透到人们的衣食住行中。将哪件衣服推荐给消费者被购买的可能性最大?该到附近哪家餐馆吃饭最合口味?该买哪个地段的房子是最佳的投资?当天走哪条路线不会堵车?类似的例子还有很多。如何对未来做出正确的决策,为用户提供便捷访问的高质量推荐,正是推荐系统研究领域的主要目标。近年来,推荐系统被广泛应用到很多领域,例如电子商务、音频视频网站、音乐电台、社交网络、个性化阅读、个性化广告、基于位置的服务和移动推荐等,并催生了很多新的推荐技术,涌现出一些著名的推荐系统,如Amazon的个性化产品推荐、Netflix的视频推荐、Pandora的音乐推荐、Facebook的好友推荐和Google Reader的个性化阅读等。推荐系统广泛应用在电子商务、大规模零售业和各种知识管理应用中,不仅给运营商带来了利益,也给用户带来了诸多便利。

推荐系统是主动地从大量信息中找到用户可能感兴趣的信息的工具。构建支持用户在线决策的系统,推荐个性化、匹配度高的产品或项目是推荐系统的核心问题。它最早可追溯到认知科学、近似理论、信息检索、预测理论、管理科学和市场中的客户选择模型等<sup>[1]</sup>。鉴于推荐系统的理论和实际应用价值,本文综述了推荐系统的研究进展,试图为进一步深入研究推荐系统理论和拓展其应用领域奠定一定的基础。

本文组织结构如下:第2章提出了推荐系统的3个研究阶段及其发展过程;第3章详细评述了推荐系统的各种分类算法;第4章给出了推荐系统的评测指标;第5章指出了推荐系统未来的研究方向。

## 2 推荐系统研究阶段

社交网络和电子商务的发展使得推荐系统的研

究和应用越来越广泛。根据谷歌学术搜索关键字“recommender system”得到的统计结果如图1所示,可以直观地看出,从1992年至今,以发表论文数目来说,与推荐系统相关的研究成果显著增长。

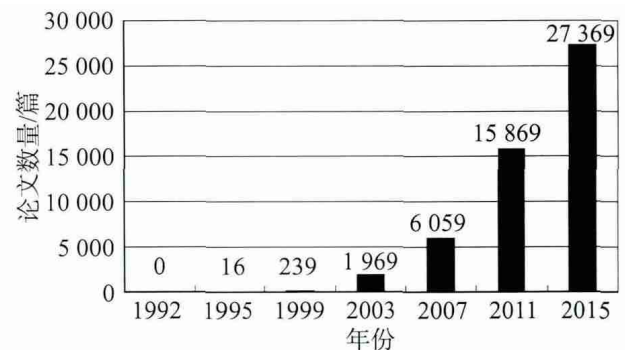


Fig.1 Number of published papers about “recommender system”

图1 推荐系统 论文数量

纵观推荐系统的研究发展进程,将推荐系统分为如下3个阶段,每个阶段又有其标志性意义的事件。

第一阶段是推荐系统形成的初期阶段。这一时期属于面向系统的探索阶段,不仅有基于协同过滤的系统,还有基于知识的系统(比如FindMe系统),表明了推荐系统的可行性和有效性,极大地激发了人们推动该领域在科研及商业实践方面不断发展的积极性。这一时期有以下3个标志性意义的事件:

(1)协同过滤。1992年Xerox Palo Alto研究中心开发了实验系统Tapestry<sup>[2]</sup>,该系统是基于当时相对新颖的利用其他用户显式反馈(评分和注释)的思想,帮助用户过滤邮件,解决邮件过载问题。这篇论文也是最早使用“协同过滤”(collaborative filtering, CF)一词的,自首次在文章中提出后被广泛引用。尽管Tapestry能够提供较好的推荐,但它有一个主要的缺陷,即需要用户书写复杂的查询语句。

(2)自动推荐。1994年产生了第一个能够自动推荐的系统GroupLens<sup>[3]</sup>。该系统也是为文本文档

(Usenet 新闻消息)过滤而开发的,和 Tapestry 不同之处在于,Tapestry 专指一个点(比如一个网站内、一个系统内)的过滤机制,GroupLens 则是既能跨网计算又能自动完成。

(3)推荐系统。1997 年 Resnick 等人<sup>[4]</sup>首次提出“推荐系统”(recommender system,RS)一词,认为该词比“协同过滤”更合适来描述推荐技术。原因有二:第一,推荐人可能不显式地和被推荐者合作,他们相互之间可能不知道对方。第二,推荐除了指出哪些应该被过滤掉,还可能建议特别感兴趣的项目。自此,“推荐系统”一词被广泛引用,并且推荐系统开始成为一个重要的研究领域。

第二阶段是推荐系统商业应用的出现。这一时期推荐系统快速商业化,效果显著。MIT 的 Pattie Maes 研究组于 1995 年创立了 Agents 公司(后更名为 Firefly Networks)。明尼苏达州的 GroupLens 研究组于 1996 年创立了 NetPerceptions。这一时期工作主要解决在大大超越实验室规模的情况下运行带来的技术挑战,开发新算法以降低在线计算时间等。这一时期标志性事件有:

(4)电子商务推荐系统。最著名的电子商务推荐系统是亚马逊(Amazon.com),顾客选择一个感兴趣的商品后,页面下方就会出现“通常一起购买的商品”和“购买此商品的顾客同时购买”的商品列表。Linden 等人公布了在 Amazon.com 中使用的基于物品内容的协同过滤方法,该方法能处理大规模的评分数据(当时有 2 900 万客户和几百万的商品目录),并能产生质量良好的推荐,大大提高了 Amazon 的营业额,据统计,推荐系统的贡献率在 20%~30%之间<sup>[5]</sup>。另一个成功的应用是 Facebook 的广告,系统根据个人资料、用户朋友感兴趣的广告等对个人提供广告推销。

第三阶段是研究大爆发,新型算法不断涌现阶段。2000 年至今,随着应用的深入和各个学科研究人员的参与,推荐系统得到迅猛发展。来自数据挖掘、人工智能、信息检索、安全与隐私以及商业与营销等各个领域的研究,都为推荐系统提供了新的分析和方法。又因为可以获得海量数据,算法研究方面取得了很大进步,更是在 2006 年被 Netflix 的 100 万美元大奖推上了高峰。这一时期标志性事件有:

(5)推荐分类。2005 年 Adomavicius 等人的综述论文<sup>[1]</sup>将推荐系统分为 3 个主要类别,即基于内容的、协同的和混合的推荐方法,并提出了未来可能的研究方向。到目前为止,这篇文章引用率高达 5 181 次,此文对推荐系统领域的研究有承上启下的作用。

(6)Netflix 竞赛。2006 年 10 月北美最大的在线视频服务提供商 Netflix 宣布了一项竞赛,任何人只要能够将它现有电影推荐算法 Cinematch 的预测准确度提高 10%,就能获得 100 万美元的奖金。该比赛在学术界和工业界引起了较大的关注,参赛者提出了若干推荐算法,提高推荐准确度,降低了推荐系统的预测误差,极大地推动了推荐系统的发展。

(7)推荐系统大会 RecSys。2007 年第一届 ACM 推荐系统大会在美国举行,到 2014 年已经是第 8 届。这是推荐系统领域的顶级会议,主要是提供一个重要的国际论坛来展示推荐系统在较广领域的新的研究成果、系统和方法。

迄今为止,推荐算法的准确度和有效性方面得到了诸多改进,极大完善了推荐效果,并可满足更多的应用需求。然而随着 Web2.0 的发展,不同的用户需求以及越来越大的数据规模对推荐系统算法的研究提出了更高的要求,因此相关研究还有大量工作要做。

### 3 推荐系统算法

推荐系统针对不同的服务对象,可以分为基于全体用户的推荐和基于单个用户的推荐(即个性化推荐)。以电子商务为例,基于全体用户的推荐,是对所有消费者给出同样的推荐,如根据物品流行度计算出的 top- $k$  个项目,推荐给所有用户。基于单个用户的推荐,是指针对不同的用户给出不同的推荐结果,如根据单个用户的行为记录,模型化用户的喜好,从而做出一对一的推荐。

#### 3.1 推荐系统定义

推荐技术也即个性化信息过滤技术,是用来预测一个给定的用户是否将会喜欢一个特定的项目(预测问题)或者是来识别一个给定用户感兴趣的  $N$  项集(top- $N$  推荐问题)。Adomavicius 等人给出了推荐系统的形式化定义<sup>[1]</sup>:设  $C$  表示所有用户集合; $S$  表示所有可能被推荐的项目集合(如书籍、电影、餐



馆等);  $u$  是一个效用函数, 用来度量项目  $s$  对用户  $c$  的有用性, 如  $u: C \times S \rightarrow R$ ;  $R$  是一个全序集合(如一定范围内的非负整数或实数)。那么, 对每一个用户  $c \in C$ , 想要找到使用户的效用函数  $u$  最大的项目  $s' \in S$ , 即:

$$\forall c \in C, S'_c = \arg \max_{s \in S} u(c, s)$$

推荐系统主动向用户提供可能感兴趣的项目, 本质是通过一定的方式将用户和项目联系起来。图2给出了推荐系统的工作原理, 从左到右依次是输入数据源, 采用某种推荐算法, 产生推荐结果进行个性化推荐。不同的推荐系统使用了不同的推荐算法, 因此推荐系统的核心在于根据不同的数据源采用不同的推荐算法。

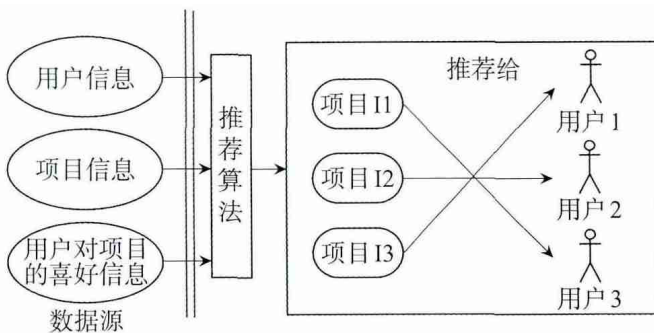


Fig.2 Operating principle of recommender system

图2 推荐系统工作原理图

### 3.2 推荐算法分类

推荐算法层出不穷, 按照不同的分类标准可以得到不同的分类结果。主流的推荐算法分为3类<sup>[1]</sup>: 基于内容的推荐、协同过滤推荐和混合推荐。随着推荐系统研究的不断深入和发展, 涌现出越来越多的算法和模型。按照模型分为最近邻模型、隐因素模型(latent factor model, LFM)和图模型等。还可根据应用领域的不同分为电子商务领域的推荐、社交网络领域的推荐、多媒体领域的推荐、移动应用领域的推荐、跨领域推荐等。文献[6]提出了推荐系统分类框架, 如图3所示, 框架图左分支是推荐系统的应用领域, 如书籍、文本、图片、电影、音乐、购物、电视节目及其他, 右分支是推荐系统用到的数据挖掘技术, 如关联规则、聚类、决策树、K-NN(K-nearest neighbor)、链路分析、神经网络、回归和启发式方法等。

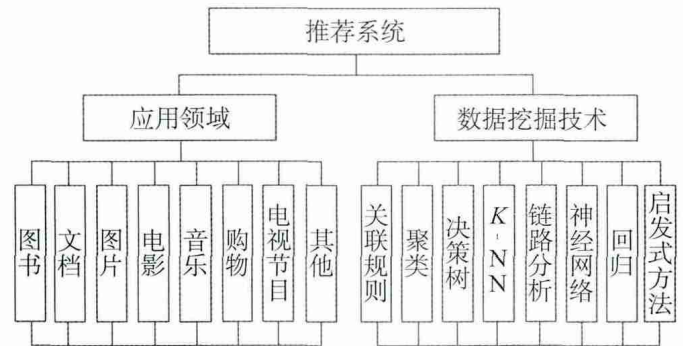


Fig.3 Classification framework of recommender system

图3 推荐系统分类框架

数据是推荐系统的基础。从图2也可以看出, 良好的推荐效果一定是来自于丰富而准确的数据(用户信息、项目信息和偏好信息)。因此, 不同于已有的推荐分类方法, 本文根据推荐系统使用的数据源不同, 将推荐系统分为以下几类: 基于用户行为数据的推荐、基于项目内容数据的推荐、基于社交网络数据的推荐、基于语境感知数据的推荐、基于人口统计学数据的推荐、基于心理学数据的推荐和基于大数据的推荐。下面详细论述这7种推荐算法, 并对每一类别的推荐方法比较分析优缺点及各自常用的推荐模型。

#### 3.2.1 基于用户行为数据的推荐

基于用户行为数据的推荐(user behavior-based recommendation)算法, 也称为协同过滤算法, 是推荐系统领域应用最广泛的算法。利用某兴趣相投, 拥有共同经验的群体喜好来给使用者推荐可能感兴趣的项目。协同过滤, 顾名思义, 个人通过合作机制给予项目相当程度的回应(如评分)并记录下来, 以达到过滤的目的, 进而帮助别人筛选项目<sup>[7]</sup>。该算法不需要预先获得用户或项目的特征, 仅依赖于用户的历史行为(显式反馈或隐式反馈)给用户兴趣建模, 从而为用户进行推荐。这里, 显式反馈是指用户对项目的打分或报告, 隐式反馈包括用户购买历史、浏览历史、搜索模式, 甚至鼠标移动等。协同过滤方法主要分为3种类型, 基于用户的协同过滤、基于项目的协同过滤和基于模型的协同过滤。

##### (1) 基于用户的协同过滤算法

基于用户的协同过滤算法(user-based CF)基于这样的假设: 一个用户喜欢和他具有相似喜好的用户喜欢的项目, 两个用户喜欢的项目交集越大, 这两

个用户越相似。两个用户兴趣相似度的计算可以有多种方法,常见的如 Pearson 相关相似性和余弦相似度计算。给定两个用户  $u, v$ , 令  $I(u)$  是  $u$  感兴趣的项目集,  $I(v)$  是用户  $v$  感兴趣的项目集, 计算出相似性用户  $u$  和  $v$  的相似性  $\text{sim}(u, v)$  后, 可通过式(1)计算用户  $u$  对项目  $i$  的喜好程度:

$$r_{ui} = \sum_{v \in S(u, K) \cap N(i)} \text{sim}(u, v) r_{vi} \quad (1)$$

其中,  $S(u, K)$  包含和用户  $u$  兴趣最接近的  $K$  个用户;  $N(i)$  是对项目  $i$  有过行为的用户集合;  $r_{vi}$  代表用户  $v$  对项目  $i$  的兴趣度。因为用户的喜好通常随时间而变化, 所以需要周期性地对用户之间的相似性进行一次更新。

### (2) 基于项目的协同过滤推荐

基于项目的协同过滤推荐(item-based CF)基于这样的假设<sup>[8]</sup>: 一个用户会喜欢与他之前喜欢的项目相似的项目。因此, 基于项目的协同过滤推荐关键在于计算物品之间的相似度。Item-based CF 认为两个项目被越多的用户同时喜欢, 两个项目就越相似。通过式(2)计算用户  $u$  对一个物品  $j$  的兴趣度:

$$r_{uj} = \sum_{i \in N(u) \cap S(j, K)} \text{sim}(j, i) r_{ui} \quad (2)$$

这里,  $N(u)$  是用户喜欢的项目的集合;  $S(j, K)$  是和项目  $j$  最相似的  $K$  个项目的集合;  $\text{sim}(j, i)$  是项目  $j$  和  $i$  的相似度;  $r_{ui}$  是用户  $u$  对项目  $i$  的兴趣度。

基于用户的协同过滤和基于项目的协同过滤统称为基于邻域的推荐(nearest neighbor recommendation), 也称作基于记忆的推荐算法(memory-based recommendation)。基于邻域的推荐算法需要维护一个用户相似度矩阵或项目相似度矩阵, 因此对于项目的数目更新速度远远小于用户数目的增长速度的情况, 宜采用基于项目的推荐算法, 如 Amazon 建立的推荐系统正是基于项目的协同过滤推荐算法<sup>[5]</sup>, 还有移动应用产品的推荐<sup>[9]</sup>。另外, 有研究表明, 基于项目的算法一般在性能上要优于基于用户的算法<sup>[10]</sup>。Wang 等人<sup>[11]</sup>通过相似性融合, 统一了基于用户和基于项目的协同过滤方法, 这样就避免了单纯基于用户或基于项目协同过滤的局限性。基于领域的推荐算法不足之处在于数据稀疏性等问题, 难以处理大数据量下的即时结果。因此提出了基于模型的协同过滤推荐算法。

### (3) 基于模型的协同过滤推荐

基于模型的协同过滤推荐(model-based CF)是采用机器学习或数据挖掘等算法, 用训练数据来学习识别复杂模式, 从而得到学习模型, 然后基于学习模型在数据集上进行智能预测<sup>[12]</sup>。文献[13]提出了一个概率方法的协同过滤, 未知打分的计算如式(3)所示:

$$r_{us} = E(r_{us}) = \sum_{i=0}^n i \times \Pr(i | r_{us}, s' \in S_c) \quad (3)$$

这里, 假设打分值是在 0 到  $n$  之间的整数, 概率表示为: 给定已有的被打分项目的用户的打分, 用户  $u$  将对项目  $s$  给出一个特定的打分。

比较有名的基于模型的协同过滤推荐算法还有: 贝叶斯信念网协同过滤模型(Bayesian belief nets CF models)<sup>[14]</sup>、聚类协同过滤模型(clustering CF models)<sup>[15]</sup>、概率因素模型(probabilistic factor models)<sup>[16]</sup>、隐语义模型(latent semantic CF models)<sup>[10]</sup>、隐因素模型(latent factor model)<sup>[17]</sup>也称为矩阵分解模型(matrix factorization), 采用 SVD(singular value decomposition)、主成分分析<sup>[18]</sup>、SVD++<sup>[19]</sup>等矩阵分解方法来推荐。文献[20]结合传统的协同过滤和概率主题模型的优点, 为用户和项目提供一个可解释的隐语义结构, 用来给某一个在线社区的用户推荐科技文章。此外, 还有基于图模型的推荐<sup>[21-22]</sup>, 将推荐问题转化为图模型中的节点选择问题, 通过分析节点之间的路径, 来提高协同过滤算法推荐结果给用户带来的惊喜度。

### 3.2.2 基于内容数据的推荐

基于内容数据的推荐(content-based recommendation)的基本思想<sup>[23]</sup>是对一个给定的用户, 推荐与他之前喜欢的项目在内容上有相似性的其他项目。这种推荐仅需要得到两类信息: 项目特征的描述和用户过去的喜好信息。它不需要一个很大的用户社区或打分历史, 对单个用户就可以产生推荐列表。关键技术是对项目建模和对用户的喜好建模, 并计算它们的相似性。建模的方法有多种, Salton 等人提出的向量空间模型<sup>[24]</sup>是最常用的基于内容数据的推荐方法。其他建模方法还有很多, 如改进的向量空间模型、显式决策模型、线性分类和机器学习等。

基于内容的推荐有两个主要缺陷: (1) 推荐给用户的项目与该用户已经消费过的项目很相似, 这使得不易发现用户不熟悉但是潜在感兴趣的项目种

类,惊喜度不高〔2〕需要预处理项目以得到能够代表它们的项目特性,但这种预处理在实际问题中往往非常困难,特别是在多媒体领域的数据(图像、音频、视频等)。针对以上不足之处,提出利用领域专家给项目打标签的方法<sup>[25]</sup>,也即传统的分类系统(Taxonomy),另一种是用户给项目打标签,也即大众分类系统(Folksolomy)。前一种比较科学和权威,但是不能很好地表达用户的个人观点,是单向无互动的。后一种比较多元化,不够准确,但能及时反馈用户观点,有助于兴趣挖掘。

### 3.2.3 基于社会网络数据的推荐

社会推荐(social recommender)主要是针对用户和他们的过去行为的社会关系进行系统化平衡。它基于这样的信念:在社会网络中,与其他用户进行链接的用户,表示愿意分享一定共同的兴趣爱好,或者他们之间有相似的品味(同质性原理),这里的相似性被期望来帮助改进推荐准确率和质量。

基于社会网络数据的推荐(social network-based recommendation)早先大部分是基于领域的方法。首先探索打分者的社会网络,聚集打分者的打分来计算预测打分,然后找到打分者的邻居。除了简单的基于邻域的方法,还有许多是基于模型的推荐方法,如采用矩阵分解方法<sup>[26]</sup>、图模型<sup>[27]</sup>,将用户的社会网络和用户物品的喜好关系建模到一张图中,联合基于项目的推荐和基于信任(trust-based)的推荐,然后利用随机游走算法给用户做推荐。采用异构数据源对用户兴趣进行建模,从异构社会网络中学习相关性<sup>[28]</sup>。这里,基于信任的推荐是将信任度引入推荐系统,用信任度代替相似度。有许多方法可以计算用户的top-*N*信任邻居。用一个信任矩阵(如Eigentrust<sup>[29]</sup>)来计算top-*N*被信任的用户,或者激活扩散模型(spreading activation model)<sup>[30]</sup>通过他们接收到的能量来给结点排序。更进一步,考虑到一个用户可能在不同的领域信任不同的朋友子集,文献[31]通过引进一个“推测的朋友圈”概念,来决定用户朋友的最佳子集合,即将用户的朋友分到与所属类别相关的社会信任环,提出基于环的在线社会网络推荐。

### 3.2.4 基于语境感知数据的推荐

20世纪90年代,“语境感知计算”概念被提出,语

境信息类型包括<sup>[32]</sup>时间、信息、外界物理环境(如天气、温度等)、设备类型、周围人员、活动状态、目的/意图等。还有些系统考虑了情绪、计算平台、网络条件、社会网络等更为广泛的语境。上下文/语境在一个消费者购买的行为中会引起很重要的变化,在用户行为模型中引入语境信息能改进预测的能力<sup>[33]</sup>。基于语境感知数据的推荐(context aware-based recommendation)系统一般分为两组:(1)通过语境驱动的查询和搜索推荐;(2)通过语境喜好的启发和评估推荐。基于语境驱动的查询和搜索已经被用在移动和游客推荐系统<sup>[34-35]</sup>,典型的是使用语境信息来查询或搜索一个特定的资源应答(如餐馆),并给出最好的匹配资源推荐给用户(如离用户最近的当前营业着的餐馆)。基于语境喜好的启发和评估,包括不同用户提供的各种项目打分的启发和评估<sup>[36-38]</sup>。这个方法可追溯到文献[37],使用多维方法来整合语境信息到推荐系统,这里传统的二维用户/项目范例被延伸来支持额外的语境维度,如时间、位置和公司。此后,许多针对语境喜好的启发和评估方法<sup>[38-39]</sup>被提出,所有这些都强调模型化和学习用户的语境感知喜好。

### 3.2.5 基于人口统计学数据的推荐

人口统计学(demographic)由人口统计资料搜集方法、人口统计资料的汇总与整理和人口分析方法3个互相联系的部分组成(<http://en.wikipedia.org/wiki/Demographics>)。通常人口统计学数据包括人的年龄、性别、国籍、民族、工作、学历、出生地等。基于人口统计学数据的推荐(demographic-based recommendation)是根据人口统计学数据对每个用户建立一个用户剖面(user profile),系统根据用户的剖面图,计算用户间相似度,得到当前用户的最近邻集,最后系统会把基于“邻居”用户群喜好的项目推荐给当前用户。

目前已有应用概率模型或智能引擎来进行基于人口统计学数据推荐<sup>[40-41]</sup>。文献[40]考虑用户模式(stereotypes),并应用基于概率的模型推荐书籍。还有利用智能引擎与互联网上的用户进行交互,基于用户的人口统计学剖面(profiles)来推荐网页。如文献[41]通过研究如何利用大量的用户人口统计数据 and 用户行为数据来构建用户的兴趣模型(LifeStyle Finder)。首先根据用户看电视、购物的行为,将用户



按照他们的人口统计学特征分成62个簇,然后对任意一个用户,先找到他所属的簇,再给他推荐这个簇里其他用户喜欢的项目。

### 3.2.6 基于心理学数据的推荐

已有研究表明,人的心理特征和情感因素在用户做决策时非常重要<sup>[42-43]</sup>。人类情感、个性及其模型已经被广泛地在计算机上实施<sup>[44]</sup>。基于心理学数据的推荐(psychology-based recommendation)方法有情感智能(emotional intelligence)<sup>[45]</sup>、满意度(satisfaction)<sup>[46]</sup>、心理作用(psychological effects)<sup>[47]</sup>、个性特质(personality-based)<sup>[48]</sup>。文献[45]提出使用用户情感智能(emotional intelligence)来改进推荐性能,这是最早把情感元素加入用户剖面来进行个性化推荐的研究。文献[49]开发了一个适应性强的智能用户模型(smart user mode SUM),引入一个兴奋属性概念,该属性是基于它们所连接的情感状态,被映射到 $[-1, 1]$ 价区间的一个加权图。接近-1价意味着压抑(inhibited)的属性,此时推荐系统忽视它。价趋近于1意味着一个激活的(activated)属性,此时推荐系统应考虑它。值得一提的是,Neflix竞赛获奖者之一就是学心理学的。

### 3.2.7 基于大数据的推荐

目前,大数据通常被认为是当前传统技术难以处理的,这里的传统技术也包括传统推荐技术。基于大数据的推荐(big data-based recommendation)有两个内涵:一个是大数据使得传统推荐技术更加准确;另一个是传统的技术已经不能满足需求,大数据需要新的推荐技术。大数据具有4V(volume, variety, value, velocity)特性:Volume指数据量大,PB级别;Variety指数据多样性形式来源;Value指价值巨大,但价值密度低;Velocity指时效性,规定时间分析。大数据分析就是从具有4V特征的数据集寻找价值的技术。如何有效分析和处理这些数据已经成为人类社会所面临的越来越严峻的挑战。

大数据的来源多样,对推荐系统而言,“大数据”包含系统可触及到的数据,如用户行为数据、社会网络数据、人口统计学数据、语境感知数据等。人们可以合理利用系统能够触及到的所有数据进行推荐。相对于前面3.2.1到3.2.6节基于单一数据的推荐,大数据环境下的推荐是至少基于两种类型的数据而进

行的推荐。研究表明,基于大数据的各种混合推荐算法的推荐效果要优于单纯的基于一种数据的推荐,如将基于社会网络的推荐和协同过滤推荐结合<sup>[50]</sup>,将基于内容的推荐和协同过滤方法结合<sup>[51]</sup>,跨领域推荐<sup>[52]</sup>。文献[53]提出了如下组合思路:加权(weighted)、变换(switching)、混合(mixed)、特征组合(feature combination)、级联(cascade)、特征扩充(feature augmentation)和元级别(meta-level),以上7种组合思路都可用来进行基于大数据的推荐。

基于大数据的推荐系统以对海量数据进行分析、挖掘为基础,收集系统涉及到的数据,通过多种算法进行精准推荐,不断提高推荐质量。基于大数据的推荐本质上是一种个性化的排序,根据不同的场景,使用不同的数据和不同的推荐算法策略,不仅帮助用户发现感兴趣的信息,而且使用推荐系统能够防止用户流失,提高了用户的忠诚度,同时一定程度上也提高了网站的点击率。此外,传统的推荐系统,架构在关系型数据库之上,关系型数据库本身对于大数据的分析与统计存在计算与存储的性能瓶颈与扩容不易等问题。基于大数据的推荐,可利用分布式技术和云计算的优势,采集并存储大量的用户行为日志,并针对不同的用户产生个性化的推荐,从数据采集、分析到推荐结果集的呈现完全实现自动化。同时,基于大数据的推荐系统具有横向扩容的特性,使得随着数据的增长,能够随时扩充推荐系统的计算与存储能力。另外,基于大数据推荐系统的性能优化方面,可以使用倒排索引和cache机制等。

## 3.3 比较与分析

每种推荐算法都有它的优势和不足,如基于用户行为数据的推荐不需要领域知识,只需用户处于一个社区中,收集用户行为数据进行推荐;基于内容的推荐算法不需要一个很大的用户社区或打分历史,对单个用户就可以产生推荐列表等。本文总结了7类推荐算法的优缺点,如表1所示。

以基于人口统计学的推荐为例,其优点是不需要当前用户对项目的历史行为数据,因此很好地解决了“冷启动”(cold start)问题。另外该方法不依赖于项目本身,因此在不同项目领域可以使用,是领域独立的(domain-independent)。不足之处在于年龄、

Table 1 Tradeoffs between recommendation algorithms

表1 权衡推荐算法

| 算法           | 优点  | 缺点                            |
|--------------|---|-------------------------------|
| 基于用户行为数据的推荐  | 不需要项目内容信息<br>不需要领域知识<br>随着时间推移,预测精度的质量会越高<br>有足够的隐式反馈                       | 新用户问题<br>新项目问题<br>对历史数据质量的依赖性 |
| 基于内容数据的推荐    | 不需要领域知识<br>随着时间推移,预测精度的质量会越高<br>有足够的隐式反馈<br>不需要用户打分或用户的邻居信息                 | 新用户问题<br>对历史数据质量的依赖性          |
| 基于社交网络数据的推荐  | 随着时间推移,预测精度的质量会越高<br>有足够的隐式反馈<br>融合社交网络信息的积极影响                              | 新用户问题<br>新项目问题                |
| 基于语境感知数据的推荐  | 不需要领域知识<br>随着时间推移,预测精度的质量会越高<br>有足够的隐式反馈<br>结合语境感知数据有利于提高推荐准确度              | 需要收集语境信息                      |
| 基于人口统计学数据的推荐 | 不需要项目内容信息<br>不需要领域知识<br>随着时间推移,预测精度的质量会越高                                   | 需要收集人口统计信息                    |
| 基于心理学数据的推荐   | 不需要项目内容信息<br>不需要领域知识<br>随着时间推移,预测精度的质量会越高                                   | 需要收集心理学方面的信息                  |
| 基于大数据的推荐     | 不需要项目内容信息<br>不需要领域知识<br>随着时间推移,预测精度的质量会越高<br>可以包括非项目的特征<br>能够在用户需求到项目之间建立映射 | 对历史数据质量的依赖性                   |

性别等人口统计数据的获取较难。由于用户对隐私的考虑,可能不愿意提供这些信息,或者不愿意提供真实的信息,从而导致推荐系统在准确度方面有误差。此外基于人口统计数据的推荐是粗粒度的,使用户得不到很好的推荐结果,没有实现完全的个性化。

除了上述已经分类好的推荐方法外,还有其他一些推荐方法:组推荐(group recommendation)<sup>[54]</sup>和基于知识的推荐(knowledge-based recommendation)<sup>[55]</sup>。基于知识的推荐使用关于用户和项目的知识来推理什么项目满足用户的需求,然后进行推荐。还有研究将经济学中的理论应用到推荐系统中,如投资组合理论<sup>[56]</sup>、边际效用递减法则(law of diminishing marginal utility),也有研究引入了购买时间间隔信息来改进推荐效果<sup>[57]</sup>。

#### 4 推荐系统的评测

随着推荐系统应用的深入和发展,基于这些基本方法的改进算法不断被提出。目前大多数推荐系统的评价指标有预测准确度<sup>[58]</sup>、覆盖率、召回率、多样性<sup>[59]</sup>、ROC(receiver operating characteristic)等<sup>[60]</sup>,文献[61]从准确度、多样性、新颖性及覆盖率等方面进行多角度阐述,细分为27个评价指标,并对各自的优缺点以及适用环境进行了深入的分析,如表2所示。其中准确度是最重要也是最常用的推荐系统评测指标,用来度量推荐系统预测的能力。对于top-*N*推荐,用准确率(precision)和召回率(recall)来评测推荐系统。

以上都是一些实验性指标,有一定的局限性,并且都是在用户选择打分的测试数据上,如很可能用户只选择那些他们喜欢的项目进行打分,是一个倾斜样本,则实验在无偏见随机样本测试中推荐质量



Table 2 Evaluation index of recommender system

表2 推荐系统评价指标简表

| 评价指标      | 名称                   | 符号          | 偏好 | 是否依赖于推荐列表长度 | 备注   |  |  |
|-----------|----------------------|-------------|----|-------------|--|--|--|
| 预测评分准确度   | 平均绝对误差               | MAE         | 小  | 否           | 适用于比较关注精确的预测评分的系统                                |  |  |
|           | 平均平方误差               | MSE         |    |             |  |  |  |
|           | 均方根误差                | RMSE        |    |             |  |  |  |
|           | 标准平均绝对误差             | NMAE        |    |             |  |  |  |
| 预测评分关联    | Pearson 关联           | PCC         | 大  | 否           | 适用于不关注精确的预测评分的系统，其中NDMP 适用于弱排序                   |  |  |
|           | Spearman 关联          | $\rho$      | 大  |             |  |  |  |
|           | Kendall 'sTau        | $\tau$      | 大  |             |  |  |  |
|           | 基于距离的标准指标            | NDMP        | 小  |             |  |  |  |
| 分类准确度     | 准确率                  | $P(L)$      | 大  | 是           | 除 AUC 外 ,其他不适用于没有明确二分喜好的系统                       |  |  |
|           | 召回率                  | $R(L)$      |    | 是           |  |  |  |
|           | 准确率提高率               | $e_p(L)$    |    | 是           |  |  |  |
|           | 召回率提高率               | $e_R(L)$    |    | 是           |  |  |  |
|           | F1 指标                | $F_1(L)$    |    | 是           |  |  |  |
|           | ROC 曲线面积             | AUC         |    | 否           |  |  |  |
| 排序准确度     | 平均分排序                | RS          | 小  | 否           | 适用于对推荐排序要求严格的系统                                  |  |  |
| 基于排序加权的指标 | 半衰期效用指标              | $HL(L)$     | 大  | 是           | 考虑了具体的推荐排序值 ,更合理些                                |  |  |
|           | 折扣累计利润               | $DCG(b, L)$ |    |             |  |  |  |
|           | 排序偏差准确率              | $RBP(p, L)$ |    |             |  |  |  |
| 覆盖率       | 预测覆盖率                | $COV_p$     | 大  | 否           | 这些指标单独使用没有意义 ,应与准确度指标一起考虑。欲计算种类覆盖率指标需要先对商品种类进行分类 |  |  |
|           | 排序覆盖率                | $COV_p(L)$  |    | 是           |  |  |  |
|           | 种类覆盖率                | $COV_c$     |    | 是           |  |  |  |
| 多样性       | Inter-user diversity | $H(L)$      | 大  | 是           |  |  |  |
|           | Intra-user diversity | $I(L)$      | 小  |             |  |  |  |
| 新颖性       | 推荐商品平均度              | $N(L)$      | 小  | 是           |  |  |  |
|           | 系统的自信息量              | $U(L)$      | 大  |             |  |  |  |
|           | 推荐的新颖率               | UE          | 大  |             |  |  |  |
|           | 考虑排序的推荐新颖率           | UER         | 大  |             |  |  |  |

是很低的<sup>[62]</sup>。为了真正理解推荐方法的利弊,高质量的实验是很有必要的。因此,可以开发面向经济的衡量指标来捕获推荐的业务价值,如投资回报率(return on investments ROI)和客户生命周期(customer lifetime value, LTV)度量<sup>[63-64]</sup>。除了以上提到的评测指标,还有一些未量化的指标:用户满意度、实时性(real-time)、健壮性(robust)、自适应性(adaptivity)<sup>[60]</sup>、隐私性(privacy)<sup>[65]</sup>等。

5 总结和展望

本文首先提出了推荐系统发展的3个阶段以及

每个阶段的标志性事件,接着从使用数据的不同对推荐系统进行分类,分析比较了每种推荐算法的优缺点,同时从推荐的角度来分析大数据,然后总结了已有的推荐系统评价指标。目前各种应用领域的推荐系统研究方兴未艾,许多问题有待进一步解决。本文认为在大数据环境下进行推荐是未来推荐系统研究的一个大方向,从大数据的4V角度看,主要的挑战及未来研究方向有以下几个方面:

(1) Volume(数据规模)。数据量巨大加剧了数据稀疏性问题和长尾(long tail)问题。在推荐系统中,可获得的已打分数目通常远小于需要预测的打

分数目。如常用的数据集都非常稀疏,当评分矩阵达到某种程度之后,相比标准的协同过滤技术,推荐质量会有所下降,而且距离关系的计算代价很高,很难实际应用到大规模评分数据上。长尾是指那些原来不受到重视的销量小但种类多的产品或服务由于总量巨大,累积起来的总收益超过主流产品的现象。大数据背景下数据稀疏性和长尾问题是一个有趣并将得到持续研究的问题。

(2) Variety(数据类型多样)。推荐系统可使用的数据复杂繁多,如社交网络里面的信息、地点位置信息和其他上下文感知信息都考虑进来,不但数据量增加,计算复杂度亦会成倍增加。另外推荐系统研究涉及到隐私保护问题,如何既保证个性化推荐又保护用户的隐私,这是一个对立的问题,给研究人员与开发者带来了很大挑战。

(3) Value(价值)。大数据本身的价值密度低,但价值巨大。对推荐系统而言,对用户兴趣建模,并将用户可能感兴趣的项目推荐给他,这里的项目相对用户而言,是有价值的项目(数据)。由于目前大多数推荐系统效率不高,如何从大数据中找出用户真正感兴趣的这些有价值的项目(数据)是个挑战。因此,更大程度地挖掘使用数据的潜在价值,并且深入挖掘用户行为模式或在复杂场景下进行推荐是未来持续研究的问题之一。

(4) Velocity(时效性)。推荐系统对时效性要求较高,想真正捕获最优的推荐机会,时效性非常重要。如何将海量的用户数据应用到实时的用户交互中以提高用户体验,这就涉及到推荐系统可扩展性(scalability)问题。目前很多推荐模型虽然可以获得很好的预测精度,但因它们有较高的计算复杂度,很难应用于实际的大规模推荐问题。未来推荐算法的可扩展性仍是一个值得研究的问题。

综上所述,推荐系统是一种联系用户和项目的信息服务系统:一方面它能够帮助用户发现潜在的感兴趣的项目;另一方面它能够帮助项目提供者将项目投放给对它感兴趣的用户。推荐系统是一个有力的系统,能够对公司或业务产生增值效应,未来必将得到持续研究与发展,给用户带来更好的体验。

## References:

- [1] Adomavicius G, Tuzhilin A. 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(6): 734-749.
- [2] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry[J]. Communications of the ACM, 1992, 35(12): 61-70.
- [3] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, Oct 22-26, 1994. New York, NY, USA: ACM, 1994: 175-186.
- [4] Resnick P, Varian H R. Recommender systems[J]. Communications of the ACM, 1997, 40(3): 56-58.
- [5] Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering[J]. IEEE Internet Computing, 2003, 7(1): 76-80.
- [6] Park D H, Kim H K, Choi I Y, et al. A literature review and classification of recommender systems research[J]. Expert Systems with Applications, 2012, 39(11): 10059-10072.
- [7] Collaborative filtering[EB/OL]. Wikipedia [2014-10-27]. [http://en.wikipedia.org/wiki/Collaborative\\_filtering](http://en.wikipedia.org/wiki/Collaborative_filtering).
- [8] Deng Ailin, Zhu Yangyong, Shi Bole. A collaborative filtering recommendation algorithm based on item rating prediction[J]. Journal of Software, 2003, 14(9): 1621-1628.
- [9] Meng Xiangwu, Hu Xun, Wang Licai, et al. Mobile recommender systems and their applications[J]. Journal of Software, 2013, 24(1): 91-108.
- [10] Deshpande M, Karypis G. Item-based top-*n* recommendation algorithms[J]. ACM Transactions on Information Systems, 2004, 22(1): 143-177.
- [11] Wang Jun, De Vries A P, Reinders M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion[C]//Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, USA, Aug 6-10, 2006. New York, NY, USA: ACM, 2006: 501-508.
- [12] Su Xiaoyuan, Khoshgoftaar T M. A survey of collaborative filtering techniques[J]. Advances in Artificial Intelligence, 2009. doi:10.1155/2009/421425.
- [13] Breese J S, Heckerman D, Kadie C. Empirical analysis of pre-

- dictive algorithms for collaborative filtering[C]//Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, Madison, USA, Jul 24-26, 1998. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc, 1998: 43-52.
- [14] Ungar L H, Foster D P. Clustering methods for collaborative filtering[C]//Proceedings of the AAAI Workshop on Recommendation Systems, Madison, USA, Jul 26-27, 1998. Menlo Park, CA, USA: AAAI, 1998: 114-129.
- [15] Ma H, Liu C, King I, et al. Probabilistic factor models for Web site recommendation[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, Jul 24-28, 2011. New York, NY, USA: ACM, 2011: 265-274.
- [16] Hofmann T. Latent semantic models for collaborative filtering[J]. *ACM Transactions on Information Systems*, 2004, 22(1): 89-115.
- [17] Bell R, Koren Y, Volinsky C. Modeling relationships at multiple scales to improve accuracy of large recommender systems[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, USA, Aug 12-15, 2007. New York, NY, USA: ACM, 2007: 95-104.
- [18] Goldberg K, Roeder T, Gupta D, et al. Eigentaste: a constant time collaborative filtering algorithm[J]. *Information Retrieval*, 2001, 4(2): 133-151.
- [19] Koren Y. Factorization meets the neighborhood: a multi-faceted collaborative filtering model[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, USA, Aug 24-27, 2008. New York, NY, USA: ACM, 2008: 426-434.
- [20] Wang Chong, Blei D M. Collaborative topic modeling for recommending scientific articles[C]//Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, USA, Aug 21-24, 2011. New York, NY, USA: ACM, 2011: 448-456.
- [21] Onuma K, Tong H, Faloutsos C. TANGENT: a novel, surprise me, recommendation algorithm[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, Jun 28-Jul 1, 2009. New York, NY, USA: ACM, 2009: 657-666.
- [22] Xiang Liang, Yuan Quan, Zhao Shiwan, et al. Temporal recommendation on graphs via long- and short-term preference fusion[C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, USA, Jul 25-28, 2010. New York, NY, USA: ACM, 2010: 723-732.
- [23] Balabanović M, Shoham Y. Fab: content-based, collaborative recommendation[J]. *Communications of the ACM*, 1997, 40(3): 66-72.
- [24] Salton G, Wong A, Yang C S. A vector space model for automatic indexing[J]. *Communications of the ACM*, 1975, 18(11): 613-620.
- [25] Guy I, Zwerdling N, Ronen I, et al. Social media recommendation based on people and tags[C]//Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, Jul 19-23, 2010. New York, NY, USA: ACM, 2010: 194-201.
- [26] Ma Hao, Yang Haixuan, Lyu M R, et al. SoRec: social recommendation using probabilistic matrix factorization[C]//Proceedings of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, USA, Oct 26-30, 2008. New York, NY, USA: ACM, 2008: 931-940.
- [27] Jamali M, Ester M. Trustwalker: a random walk model for combining trust-based and item-based recommendation[C]//Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, Jun 28-Jul 1, 2009. New York, NY, USA: ACM, 2009: 397-406.
- [28] Wang Chi, Raina R, Fong D, et al. Learning relevance from heterogeneous social network and its application in online targeting[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, Jul 24-28, 2011. New York, NY, USA: ACM, 2011: 655-664.
- [29] Kamvar S D, Schlosser M T, Garcia-Molina H. The eigentrust algorithm for reputation management in P2P networks[C]//Proceedings of the 12th International Conference on World Wide Web, New York, USA, May 17-22, 2003. New York, NY, USA: ACM, 2003: 640-651.
- [30] Minsky M L, Minsky M. Semantic information processing[M]. Cambridge, MA, USA: MIT Press, 1968, 227-270.
- [31] Yang Xiwang, Steck H, Liu Yong. Circle-based recommendation in online social networks[C]//Proceedings of the



- 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, Aug 12-16, 2012. New York, NY, USA: ACM, 2012: 1267-1275.
- [32] Wang Licai, Meng Xiangwu, Zhang Yujie. Context-aware recommender systems[J]. *Journal of Software*, 2012, 23(1): 1-20.
- [33] Bettman J R, Luce M F, Payne J W. Constructive consumer choice processes[J]. *Journal of Consumer Research*, 1998, 25(3): 187-217.
- [34] Cena F, Console L, Gena C, et al. Integrating heterogeneous adaptation techniques to build a flexible and usable mobile tourist guide[J]. *AI Communications*, 2006, 19(4): 369-384.
- [35] Van Setten M, Pokraev S, Koolwaaij J. Context-aware recommendations in the mobile tourist application COMPASS[C]//LNCS 3137: Proceedings of the 3rd International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Eindhoven, Netherlands, Aug 23-26, 2004. Berlin, Heidelberg: Springer, 2004: 235-244.
- [36] Adomavicius G, Sankaranarayanan R, Sen S, et al. Incorporating contextual information in recommender systems using a multidimensional approach[J]. *ACM Transactions on Information Systems*, 2005, 23(1): 103-145.
- [37] Adomavicius G, Tuzhilin A. Multidimensional recommender systems: a data warehousing approach[M]//Electronic Commerce. Berlin, Heidelberg: Springer, 2001: 180-192.
- [38] Anand S S, Mobasher B. Contextual recommendation[M]. Berlin, Heidelberg: Springer, 2007: 142-160.
- [39] Rendle S, Gantner Z, Freudenthaler C, et al. Fast context-aware recommendations with factorization machines[C]//Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, Beijing, China, Jul 24-28, 2011. New York, NY, USA: ACM, 2011: 635-644.
- [40] Rich E. User modeling via stereotypes[J]. *Cognitive Science*, 1979, 3(4): 329-354.
- [41] Krulwich B. Lifestyle finder: intelligent user profiling using large-scale demographic data[J]. *AI Magazine*, 1997, 18(2): 37-45.
- [42] Damasio A R, Marg E. Descartes' error: emotion, reason, and the human brain[J]. *Optometry and Vision Science*, 1995, 72(11): 847-847.
- [43] Thagard P, Kroon F. Hot thought: mechanisms and applications of emotional cognition[M]. Cambridge, MA, USA: MIT Press, 2008.
- [44] Nunes M. Psychological aspects in lifelike synthetic agents: towards the personality markup language (a brief survey)[J]. *Novas Tecnologias na Educação*, 2009, 7(3): 1-11.
- [45] Gonzalez G, de la Rosa J L, Montaner M, et al. Embedding emotional context in recommender systems[C]//Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop, Istanbul, Turkey, Apr 17-20, 2007. Piscataway, NJ, USA: IEEE, 2007: 845-852.
- [46] Masthoff J, Gatt A. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems[J]. *User Modeling and User-Adapted Interaction*, 2006, 16(3/4): 281-319.
- [47] Saari T, Ravaja N, Laarni J, et al. Psychologically targeted persuasive advertising and product information in e-commerce[C]//Proceedings of the 6th International Conference on Electronic Commerce, Delft, Netherlands, Oct 25-27, 2004. New York, NY, USA: ACM, 2004: 245-254.
- [48] Burke R. Hybrid recommender systems: survey and experiments[J]. *User Modeling and User-Adapted Interaction*, 2002, 12(4): 331-370.
- [49] González G, López B, de la Rosa J L. A multi-agent smart user model for cross-domain recommender systems[C]//Proceedings of the 2005 International Conference on Intelligent User Interfaces: Beyond Personalization, San Diego, USA, Jan 9, 2005.
- [50] Konstas I, Stathopoulos V, Jose J M. On social networks and collaborative recommendation[C]//Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval, Boston, USA, Jul 19-23, 2009. New York, NY, USA: ACM, 2009: 195-202.
- [51] Melville P, Mooney R J, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C]//Proceedings of the 18th National Conference on Artificial Intelligence, Edmonton, Canada, Jul 2002. Menlo Park, CA, USA: AAAI, 2002: 187-192.
- [52] Tang Jie, Wu Sen, Sun Jimeng, et al. Cross-domain collaboration recommendation[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, Aug 12-16, 2012. New York, NY, USA: ACM, 2012: 1285-1293.
- [53] Burke R. Hybrid recommender systems: survey and experi-

- ments[J]. User Modeling and User-Adapted Interaction, 2002, 12(4): 331-370.
- [54] Amer-Yahia S, Roy S B, Chawlat A, et al. Group recommendation: semantics and efficiency[J]. Proceedings of the VLDB Endowment, 2009, 2(1): 754-765.
- [55] Burke R. Knowledge-based recommender systems[J]. Encyclopedia of Library and Information Systems, 2000, 69(S): 175-186.
- [56] Shi Yue, Zhao Xiaoxue, Wang Jun, et al. Adaptive diversification of recommendation results via latent factor portfolio[C]//Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, USA, Aug 12-16, 2012. New York, NY, USA: ACM, 2012: 175-184.
- [57] Zhao Gang, Lee M L, Hsu W, et al. Increasing temporal diversity with purchase intervals[C]//Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, USA, Aug 12-16, 2012. New York, NY, USA: ACM, 2012: 165-174.
- [58] Massa P, Avesani P. Trust-aware recommender systems[C]//Proceedings of the 2007 ACM Conference Recommender Systems, Minneapolis, USA, Oct 19-20, 2007. New York, NY, USA: ACM, 2007: 17-24.
- [59] Smyth B, McClave P. Similarity vs. diversity[M]//Case-Based Reasoning Research and Development. Berlin, Heidelberg: Springer, 2001: 347-361.
- [60] Xiang Liang. Recommendation system practice[M] Beijing: The People's Posts and Telecommunications Press, China, 2012.
- [61] Zhu Yuxiao, Lv Lingyuan. Evaluation metrics for recommender systems[J]. Journal of University of Electronic Science and Technology of China, 2012, 41(2): 163-175.
- [62] Mooney R J, Roy L. Content-based book recommending using learning for text categorization[C]//Proceedings of the 5th ACM Conference on Digital Libraries, San Antonio, USA, Jun 2-7, 2000. New York, NY, USA: ACM, 2000: 195-204.
- [63] Dwyer F R. Customer lifetime valuation to support marketing decision making[J]. Journal of Direct Marketing, 1989, 3(4): 8-15.
- [64] Rosset S, Neumann E, Eick U, et al. Customer lifetime value modeling and its use for customer retention planning[C]//Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, Jul 23-25, 2002. New York, NY, USA: ACM, 2002: 332-340.
- [65] Kobza A. Privacy-enhanced personalization[J]. Communications of the ACM, 2007, 50(8): 24-33.

#### 附中文参考文献:

- [8] 邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628.
- [9] 孟祥武, 胡勋, 王立才, 等. 移动推荐系统及其应用[J]. 软件学报, 2013, 24(1): 91-108.
- [32] 王立才, 孟祥武, 张玉洁. 上下文感知推荐系统[J]. 软件学报, 2012, 23(1): 1-20.
- [60] 项亮. 推荐系统实践[M]. 北京: 人民邮电出版社, 2012.
- [61] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. 电子科技大学学报, 2012, 41(2): 163-175.



ZHU Yangyong was born in 1963. He received the Ph.D. degree in computer software and theory from Fudan University in 1994. Now he is a professor and Ph.D. supervisor at Fudan University. His research interests include dataology, data science, data mining, big data analysis and data cloud, etc.

朱扬勇(1963—),男,浙江武义人,1994年于复旦大学计算机软件与理论专业获得博士学位,现为复旦大学计算机科学技术学院教授、博士生导师,上海市数据科学重点实验室主任,主要研究领域为数据学、数据科学、数据挖掘、大数据分析、数据云等。



SUN Jing was born in 1985. She is a Ph.D. candidate at Fudan University, and the member of CCF. Her research interests include data science, data mining and recommender system, etc.

孙婧(1985—),女,江苏盐城人,复旦大学博士研究生,CCF会员,主要研究领域为数据科学、数据挖掘、推荐系统等。