

国内电子商务个性化推荐研究进展：核心技术

孙雨生^{1,2,3} 张 晨¹ 任 洁¹ 朱礼军²

(1. 湖北工业大学经济与管理学院, 湖北 武汉 430068;

2. 中国科学技术信息研究所信息技术支持中心, 北京 100038;

3. 湖北工业大学区域产业生态发展协同创新中心, 湖北 武汉 430068)

(摘 要) 本文从用户兴趣建模、推荐机制、信息资源管理三方面阐述了国内电子商务个性化推荐研究进展: 用户兴趣建模研究用户兴趣模型表示、初始化、进化; 推荐机制按实现方式分协同过滤推荐、基于内容推荐、基于规则推荐、基于知识推荐、基于效用推荐、基于人口统计推荐、混合推荐; 信息资源管理研究集中在数据挖掘和存储。

(关键词) 电子商务; 个性化推荐; 研究进展; 用户兴趣建模; 推荐机制

DOI: 10.3969/j.issn.1008-0821.2017.04.024

(中图分类号) TP399; G202 (文献标识码) A (文章编号) 1008-0821(2017)04-0151-07

Research Development of E-Commerce Personalized Recommendation in China: Core Technology

Sun Yusheng^{1,2,3} Zhang Chen¹ Ren Jie¹ Zhu Lijun²

(1. School of Economics and Management, HuBei University of Technology, Wuhan 430068, China;

2. Information Technology Support Center, Institute of Scientific & Technical Information of China, Beijing 100038, China;

3. Collaborative Innovation Center of Regional Industrial Ecology Development, HuBei University of Technology, Wuhan 430068, China)

(Abstract) The authors revealed research development on e-commerce personalized recommendation in china from three aspects: users' interest modeling, recommendation mechanism, information resource management. Users' interest modeling explore the representation, initialization and evolution of users' interest model; According to the way of implementation, recommendation mechanism can be divided into collaborative filtering recommendation, content-based recommendation, rule-based recommendation, knowledge-based recommendation, utility-based recommendation, demographic statistics-based recommendation and hybrid recommendation; research on information resource management was concentrated on data mining and storage.

(Key words) e-commerce; personalized recommendation; research development; user interest modeling; recommendation mechanism

随着互联网应用普及和电子商务迅猛发展, 用户个性化需求日益凸显, 客观需为其个性化推荐商品和服务。由

收稿日期: 2016-10-08

基金项目: 国家自然科学基金资助项目“基于语义网格的数字图书馆个性化推荐模型研究”(项目编号: 71003032); 中国博士后科学基金资助项目“基于领域本体的数字图书馆用户兴趣建模研究”(项目编号: 2014M560107); 湖北省社会科学基金资助项目“语义网格环境下数字图书馆个性化推荐机制及其应用研究”(项目编号: 2015108); 全国大学生创新创业训练项目“美源网生态农场项目方案设计”(项目编号: 201310500035)。

作者简介: 孙雨生(1980-), 男, 副教授, 博士后, 硕士生导师, 研究方向: 计算机信息系统工程、知识管理技术、数字图书馆技术。

此, 电子商务个性化推荐应运而生, 其基于所获取用户需求、兴趣爱好、网络行为等提取特征^[1]并构建用户兴趣模型, 动态发现用户潜在、个性需求并选择推荐机制匹配候选商品^[2], 提供个性化主动信息、商品服务以提高用户满意度、商品销量。

本文以知网、万方的学位论文、期刊论文库及维普的期刊论文库为信息源, 以“电子商务”、“个性化”、“推荐”为关键词组合在题名中检索相关文献(截至 2016 年 8 月 11 日, 从知网获硕博论文 73 篇、期刊论文 91 篇; 从万方获硕博论文 80 篇(新发现 17 篇)、期刊论文 90 篇(新发现 3 篇); 从维普获期刊论文 95 篇(新发现 8 篇), 合计 192 篇文献), 从用户兴趣建模、推荐机制、信息资源管理角度阐述国内电子商务个性化推荐研究进展。

1 电子商务个性化推荐核心技术研究进展

1.1 用户兴趣建模

1.1.1 模型表示

国内用户兴趣模型表示多用基于向量空间模型法^[3](较复杂且精度不够高^[2,4-6])、基于神经网络法(依赖模型学习所用神经网络类别和算法, 适用范围窄且难理解^[3])、基于案例法(反映用户短期兴趣^[4-6])、基于用户-项目评价矩阵法(难反映兴趣变化^[3])、基于本体法(可重用共享知识但无法避免本体设计问题^[4-5,7])、基于决策树法、基于简单贝叶斯概率模型法、基于关键词法、基于语义网络法, 此外还有基于加权向量模型法、基于类型层次结构模型法、基于目录结构法等^[3]。

1.1.2 模型初始化

主要是用户兴趣数据采集及处理, 涉及兴趣学习技术(如 TF-IDF、贝叶斯分类、决策树、神经网络、聚类^[3]等^[3])。

1) 兴趣数据采集

获取与用户特征、兴趣或活动相关信息, 遵循增量式、协作性、简单性原则^[3], 方式有显式、隐式^[5-6]、混合式(融合隐式、显式采集优点^[3])^[4,7]。

显式采集信息由注册信息、评分、文本评价^[3]、显式样本集、目标陈述^[4]等组成, 简单直接、准确^[5,7]、学习速度快^[4]但需用户确知兴趣且参与(多不成立), 无法收集全面真实数据^[5-8], 即使用户提供真实兴趣, 单独显式采集难描述用户对项目复杂感知和反应^[4], 为此, 代其锋^[9]、蒋翀^[10]认为应告知用户个性化推荐概念、作用、原理以激励其参与。

隐式采集信息由服务器端(日志, 记录用户 IP、ID 及访问(时间、URL)、浏览^[4,6](浏览时间)、检索^[3]、购买等行为)、客户端数据(包括用户请求帮助、页面停留时间、访问项目次序、浏览、点击、保存、操作时间^[2,7]等^[4])组成。此外, 代其锋^[9]、蒋翀^[10]认为包括反复行

为、标记行为、操作行为、浏览路径、消极行为。优点是增加用户负担且动态获取海量数据(可能有冗余, 计算成本和复杂度较大)^[4,6]; 缺陷是缺乏控制(随机、不确定)^[5,7], 需解决基于用户导航行为的长、短期兴趣^[4]区分问题。

2) 兴趣数据处理

主要是数据预处理和兴趣度量化。前者先预处理 Web 日志和用户评分^[11](先净化数据并识别用户、会话、路径、事务, 再分词、计算特征词权重及提取特征词集, 特征化、聚类事务集^[11]), 得出用户访问情况以便量化兴趣度。对于后者, 李凤慧^[12]用(浏览时长/总浏览时长)×(网页字节数/总信息数)量化兴趣度(指标分用户访问频率^[13]、持续访问时间^[14]、最近访问时间^[13])^[4]; 安芳^[15]、张黎黎^[16]用阈值法量化兴趣度; 李熠^[17]用购买次数量化兴趣度; 杨永亮^[18]基于时间衰减因子, 用对商品发生行为时间间隔量化兴趣度; 王伟等^[19]预处理文本评价(用 Core-NLP 包词性(POS)标记提取特征并用 WordNet 计算与种子词相似度以同义分组; 给出并综合用户对各意见词的积极性、消极性、客观性值形成意见值)形成特征-意见值对, 量化兴趣度; 邓晓懿^[20]基于支持度、置信度框架量化规则兴趣度; 刘晶等^[21]基于虚拟个体关联、情景关联、行为关联模块量化规则兴趣度。

1.1.3 模型进化

1) 进化内容

用户兴趣模型随用户活跃度、兴趣变化, 基于用户显隐式反馈、兴趣衰减自适应进化^[4,7,9,14], 主要包括兴趣项目和兴趣度: 前者基于用户注册、反馈信息提取兴趣项目(生成并合并显、隐式兴趣项目向量)并排序^[3], 张建伟^[11]基于交易事务链接前后 Web 页特征词变化进化用户兴趣; 后者基于用户评分修改^[3]、阈值比较判断是否进化。

2) 进化方法

现有模型进化方法常按固定频率更新, 动态跟踪、高效获取用户兴趣变化能力有限, 影响模型精度。李晓昀^[4]提出信息增补法(基于用户反馈进化, 分直接增补法(不剔除不再感兴趣项目, 有精度不高、存储空间浪费问题)、权重调整增补法(性能受新信息选择方法、增量影响))、自然进化法(按生态系统优胜劣汰法进化, 需针对性设计适应度函数和染色体编码法)和神经网络法(调整网络连接权重自适应进化, 须以神经网络为学习法); 杨静^[3]、张黎黎^[16]提出主动进化法(更新兴趣项目并调整权重、次序, 维护知识分类体系及资源类型目录)和基于用户反馈自动进化法(胡一^[10]、蒋翀^[10]又根据显式、隐式分直接进化(影响浏览, 参与者有限)和间接进化法(基于浏览行为、日志挖掘等)); 杨永亮^[18]基于用户行为指纹算法(哈希处理用户行为记录并通过记录权重随时间衰减模拟兴趣进化)进化; 蒋翀^[10]提出线性衰减法进化(兴趣度衰

减到零时淘汰项目); 俞正骄^[2] 递进挖掘序列模式间兴趣度进化基于规则推荐中用户兴趣。

1.2 推荐机制

电子商务个性化推荐核心及性能直接决定因素, 由推

荐引擎基于用户会话、推荐机制, 推荐离线挖掘结果和动态内容 (链接、广告、文本、商品、服务等^[8])^[23], 主流推荐机制 (按实现方式分类) 优劣势见表 1。

表 1 主流推荐机制优劣势

机 制	优 势	劣 势
协同过滤推荐 ^[3]	无需领域知识, 实现简单 ^[3] ; 实时性好 ^[3] (用户反馈少、学习速度快); 个性化、自动化 ^[3] 、自适应性强, 奇异发现能力强 (内容不相似信息、用户潜在兴趣 ^[4,26,23]); 推荐质量随用户量增加提高; 共享他人经验, 内容分析全而精 ^[3] ; 可推荐复杂对象; 可解释性好 ^[3]	有数据稀疏 ^[3] 、冷启动 ^[6,30] 、准确性 ^[4,30] 、健壮性、扩展性、灵活性 (难动态变更)、同义词、特殊用户 (难发现相似用户)、时间因素 (难区分用户兴趣变化) 及难区分用户正负兴趣、隐私性等问题; 须有一定用户评价信息且目标用户已评价所推荐项目 ^[3]
基于内容推荐 ^[3]	较成熟, 简单有效, 无需他人数据; 实时性好 (可离线建模、计算商品相似度) ^[3] ; 可推荐新或非流行项目且用户独立 ^[7,17,32] ; 结果可解释性强 (易懂、给出推荐项目内容特征); 部分解决用户多兴趣和产品信息多样问题 ^[3] ; 无特殊用户、数据稀疏、冷启动问题	实现略复杂, 依赖历史数据且需较多分类器; 奇异发现能力不强 ^[10,32] , 仅推荐特征相似项目且仅基于目标用户反馈 ^[3] ; 需内容易抽取成有意义、结构化特征且用户兴趣易用内容特征形式表示; 无法处理品质、风格或观点 ^[3]
基于规则推荐	无需领域知识, 简单且技术成熟; 多离线处理实时性好 ^[2,33] ; 性能、伸缩性较好; 可推荐与用户已喜欢项目相关 (项目关联)、不相关 (用户关联) 项目并动态反映兴趣变化	个性化、自动化度不高 (基于被购商品间关系构建项目关联); 规则难发现 ^[3] (尤其是数据处理繁杂, 数据集高维、稀疏时)、管理 (无法动态更新 ^[4] , 同义词问题)、解释; 新项目问题
基于知识推荐	实时性好; 基于用户知识推荐尽可能广而全项目 (对用户信息需求较少 (可通过商品反映), 可考虑非商品属性); 无数据稀疏、冷启动问题	奇异发现能力有限; 知识难管理 (获取、聚合、知识库构建)
基于效用推荐	全面性、个性化度高; 可考虑非项目属性; 无数据稀疏、冷启动、可扩展问题 ^[2,30,34-35]	奇异发现能力、灵活性弱; 存在效用属性重叠问题; 过程繁琐且不适合临时浏览用户 ^[33]
基于人口统计推荐	操作简单, 无冷启动问题 ^[30,33]	奇异发现能力弱, 有数据稀疏问题, 人口统计资料难获取, 结果不精确

1.2.1 协同过滤推荐

研究最多也最成功^[23-24], 又称社会过滤、协作过滤、合作过滤^[10,14], 算法分析用户对商品评价 (点击、购买历史^[21])、兴趣度, 发现兴趣相近用户、关系相似项目, 基于近邻历史偏好进行推荐^[2,4,7-8,10,17,20-21,23,25-29,32,33], 主要研究近邻方法和基于项目、用户特征评级的潜在因素模型, 分基于记忆 (分基于用户、基于项目)、基于模型两类, 对比见表 2, 代表系统为 Tapestry、Ringo/Firefly、GroupLens/Netperceptions、Movielens^[23]。

1.2.2 基于内容推荐

用机器学习法学习用户信息 (兴趣^[2]、需求等) 和已

评价商品特征^[6] (通过项目属性定义特征并基于数学模型^[27] 用特征值表示候选项目和用户兴趣模型^[2,4,6,8,10,17,21,24-25,28,31-33]), 用自然语言、人工智能、概率统计、数据挖掘、机器学习、邻居函数分析等技术计算用户与项目相似度^[21] 进行推荐^[2] 并基于用户反馈进化用户兴趣模型。常用推荐方法有基于商品特征分类^[3]、基于商品关联、基于文本^[21] 分类、基于语义法^[2], 此外, 王星^[24] 按用户兴趣模型表示分基于向量空间模型、基于关键词分类、基于领域分类和基于潜在语义索引等推荐法; 代表系统有 Personal WebWatcher、LIBRA、LyricTime、SIFTS、Letizia、InfoFinder、Newsweeder、WebMate、WebACE 等。

表 2 协同过滤推荐类型对比

	基于记忆 (内存)		基于模型
	基于用户	基于项目	
比较对象	用 户	项 目	用 户
依 据	启发规则、用户 - 项目评分矩阵		用户行为模型
目 的	预测目标用户未评分项目评分		预测目标用户评分
相似性计算	用户相似性	项目相似性	无
方 法	统计学、相似性计算、最近邻		统计学、机器学习、数据挖掘

表 2 (续)

	基于记忆 (内存)		基于模型
	基于用户	基于项目	
步 骤	初始化数据 (收集数据并建模用户兴趣); 基于用户评分计算相似度; 用 Top - N、KNN、阈值法、聚类法、贝叶斯网络形成用户或项目邻居集; 基于最近邻对候选项目历史评分, 用简单均值、基于用户相似度加权均值、中心加权均值、增量计算加权均值等预测目标用户项目评分 ^[6,8] , 基于最频繁项目、关联规则、Top - N 等推荐 ^[7,14,20,24,29]		学习用户历史评分; 识别复杂模式归纳用户行为模型; 预测目标用户项目评分并推荐 (按评分类型 (类目或数值) 分别采用分类算法, 回归模型和奇异值分解算法)
优 势	易实现、较常用; 使用相似用户知识经验、反馈信息; 奇异发现能力强	易实现、较常用; 实时性好 (离线计算); 可扩展性、解释性好	实时性好 (离线建模); 准确度高
劣 势	实时性、完整性及多样性、扩展性、稀疏性、冷启动、特殊用户问题	奇异发现力不强; 多样化、冷启动问题 ^[29]	离线模型训练耗时且需周期更新

1.2.3 基于规则推荐

基于所选规则前件 (用户已购商品或兴趣^[4,5,17,25,28,33]) 个性化推荐规则后件 (推荐对象)^[2,21,24]。规则由用户定制或基于用户关联关系^[2]挖掘^[4] (基于用户事务数据库, 挖掘满足最小支持度及可信度的关联规则^[17,23,27,33]), 代表系统有 WebSphere、BroadVision、SurfLen、ASARM^[3]。

1.2.4 基于知识推荐

用基于案例推理、数据挖掘^[23,27]技术挖掘项目间关联知识进行推荐, 按所用知识形式 (如功能知识 (满足特定用户需求且可解释需求和推荐间关系^[3])、用户信息 (支持推理知识结构^[4-5,25,33,33]或详细用户需求表示)) 分数据库知识发现、基于案例推理、知识推理^[23]。

1.2.5 基于效用推荐

计算商品对用户效用 (基于商品属性分值和效用函数权值分析内容、K 均值得出效用函数并随用户兴趣动态调整^[37])^[34], 用受限满足技术匹配用户需求和项目^[2,25,30,33]进行 Top - N 推荐并评估效果^[25,35,37], 核心是用效用函数代表用户兴趣模型。

1.2.6 基于人口统计推荐

基于用户显式属性分类用户, 结合人口统计信息进行推荐^[25,27,30,33]。

1.2.7 混合推荐

综合各推荐方法优势提升推荐性能, 较成功是协同过滤推荐与基于内容推荐组合^[17,23], 难点是组合对象选取方法及组合方式^[3], 组合方式有加权^[3]、变换^[28]、混合^[3]、特征组合、层叠、特征扩充^[17,31]、元层次性^[29,32-33,33]。此外, 杨永亮^[38]提出时间敏感物质扩散算法 (计算资源间相似度时考虑资源属性、流行度随时间变化) 并与基于项目协同推荐算法混合; 用最长公共子序列算法计算用户相对次序下行为最长序列以度量用户消费轨迹相似度, 提取并计算基于时间敏感用户间非公共兴趣列表对用户兴趣贡献度, 与基于用户协同推荐算法线性加权混合。

1.3 信息资源管理

主要分数据采集、挖掘、表示、存储、更新, 主要管

理用户^[24]、项目、推荐过程产生信息。

1.3.1 数据采集

现有研究集中在用户特征显式、隐式采集 (通过网络爬虫^[3]采集信息、分析日志, 获取用户浏览内容、行为信息)。对于采集内容, 李婷^[10]、马瑞敏等^[28]认为有用户信息 (注册、评分)、商品属性; 汪亭廷^[23]、辛勤芳^[24]认为有用户背景知识、评论、反馈及行为 (网站访问、购物^[23]) 等; 刘晶等^[3]从移动端流式采集网络、视频、传感等数据; 刘庆华^[33]认为有概念、关联规则、时间序列模式、最频繁购物路径。此外, 商品信息多用人工录入、系统导入、自动设备采集等方式。

1.3.2 数据挖掘

电子商务数据挖掘主要针对 Web 数据, 多通过关联规则、分类^[3]离线挖掘以支持在线推荐, 前沿是基于本体的语义 Web 挖掘^[34]。

1) Web 数据挖掘

基于海量、复杂 Web 信息 (多半结构化^[1,3]、不完整、有噪声)、异构数据库环境^[1,13], 通过分析用户习惯、兴趣, 基于其心理变化, 通过挖掘算法发现蕴涵、未知、有潜在应用价值、新颖、易理解及非平凡模式过程 (用于指定挖掘任务, 如描述所挖掘数据一般特性、预测^[38]趋势和行为、检测偏差、支持决策^[3]), 常用人工智能、计算语言学、信息学、统计学等领域知识^[13,33], 按挖掘对象分 Web 内容 (文本、多媒体) 挖掘、Web 结构 (超链接、内容、URL) 挖掘、Web 使用记录 (日志) 挖掘^[1,8,11,13,22,32,33] (三者可融合)^[3], 挖掘流程见表 3。此外, 宋伟国^[4]用数据仓库处理、集成与装载、分析数据以支持数据挖掘; 刘蓓琳^[6]用降维、稠密化和标准化方法预处理数据。

2) 关联规则挖掘

分频繁项集发现、关联规则生成^[20,33,38], 后者支持商业决策 (广告、商品推荐)、网站商品陈列调整, 算法有 Apriori、改进 Apriori、AprioriTid、DHP、FP - tree、Tree Projection 等^[9,10,14,21,23,29,34]。

表 3 Web 数据挖掘流程^[8,13,32,38]

步 骤	内 容	内 涵
数据采集	服务器端 (日志、查询数据)、客户端 (主要是 Cookie)、代理端周期采集用户访问行为 ^[1] 并采集市场、用户登记、Web 页面信息。	
数据预处理 (形成半结构化事务序列)	清洗	按实现方式与范围分手工实现、专门编程、清洗特定应用领域问题、清洗特定应用领域无关数据等以除噪、补缺、纠正不一致 ^[23]
	变换	将数据转成适合挖掘形式, 方法有平滑 (分箱、聚类、回归)、聚类、概念化分层、规范化 (将属性数据按比例缩放, 落入特定区间) 和最小 - 最大规范化 (线性变换)
	集成	集成多源数据, 涉及实体识别、剔冗等
	归约	将大量数据转为小数据集, 常用策略有数据立方体聚集、维归约、数据压缩、数值压缩、离散化、概念分层等
	用户识别	根据服务日志用跟踪机制 ^[11] 、基于注册信息或 IP 识别用户; 还可利用 Cookie (嵌入 SessionID)、代理软件、浏览器修改
	会话识别	识别日志中同一用户同次访问请求; 方法有面向时间启发式法、基于引用启发式法和时间窗法 ^[11,23] , 可聚类每个用户有意义页面、跟踪其兴趣变化
	路径识别	借助网站拓扑结构、补全用户请求, 获得完整访问路径 ^[11]
	事务识别	语义分组用户会话, 形成适合挖掘 Web 对象并格式化, 方法有引用时长法、最大向前访问路径法和时间窗法
模式识别	用数据挖掘算法 (统计分析、关联规则、分类 (根据用户属性)、聚类 ^[34] (用户聚类 (基于访问信息)、页面聚类)、分类聚类混合、序列模式 ^[34] (挖掘 Web 日志、预测用户访问模式)、频繁访问组、依赖建模等) 发现模式 ^[23]	
模式分析	知识表示 (根据应用以用户易理解形式表示统计结果、规则、模型)、模式评估 (分析、剔除冗余或无关模式, 重新处理无法满足要求模式)、经观察、选择和技术筛选后应用	

3) 分类挖掘

基于机器学习法, 用分类器 (基于训练集产生并随训练集动态变化, 可离线进行) 分类对象^[11], 常见类型见表 4 (还有决策树、SVM 等)。为更好推荐, 需将目标用户分到与其较相似用户所在簇^[3]; 黄莹^[34]、陈晓红^[3] 基于与

簇相关度加权目标用户与所在簇关系强度^[10,13]; 王海超^[34] 提出用聚类离线预处理最近邻算法以缩小计算范围、分布计算、提高寻找相似簇性能、推荐扩展性^[34] 但质量不高。

表 4 分类挖掘类型^[10,13,29,32]

类 型	内 涵	方法及算法	特 点
聚类 ^[20,21]	发现数据分布 (识别密集、稀疏区, 发现全局对象分布模式及数据属性间关系) 和模式 (基于用户信息发现用户群并用购买模式描述), 形成多个簇 ^[34]	划分法、层次法、基于密度法、基于网格法、基于模型法、模糊聚类等 ^[11] ; 算 法 有 SOM、K - means ^[11] 、PSO、Isodata 等	无
贝叶斯网络 ^[23]	基于概率分析、图论 (描述数据实例中关联或因果关系) 的不确定性知识表达 (不分前向、后向推理) 和基于概率推理模型, 结构化描述知识并支持推理: 基于先验信息、样本数据、专家知识, 根据事物间关系确定结构模型, 用其它方法确定各节点 (事件) 条件概率 (确定网络图, 基于因果关系 (连接事件) 计算节点间条件概率) 基于图论 (节点代表用户, 边代表用户相似度, 可跳过中间节点寻找最近邻 (节点间传递相似关系), 构建用户或商品间相似关系图) 表示思想、综合近邻节点观点推荐	随机模拟、多树传播、团树传播、图约减、组合优化等算法	表达语义直观清晰, 推理结论可信易理解, 不适合用户兴趣模型频繁修改情形, 应用时计算量大
Hortling 图		无	简单, 实时性高, 无冷启动问题

1.3.3 数据表示

主要表示用户、商品信息。杨静^[3]、安芳^[34]、张黎黎^[11] 基于商品信息 (分类、编码)、用户信息提取商品属性、用户属性并用特征向量表示。

1.3.4 数据存储

用数据库^[3] 工具构建资源库存储用户及商品、推荐过程信息, 分静态、动态数据库, 核心是电子商务元数据, 含数据存储、分类、被访问信息等。此外, 丁建军^[3] 分别用数据库、分布式文件系统存储显式、隐式采集的用户兴趣。

1) 用户数据库

即存储用户基本、行为信息,包括用户 ID、用户名、密码、姓名、性别、年龄、出生地^[2]、教育背景、兴趣、收入、工作单位、住址、邮箱^[28]、电话、有效期、级别^[39-40],上次登录时间、购物记录、商品评分(含用户 ID、商品 ID、评分、时间戳^[28])^[4]等。此外还有管理员数据库^[34]。

2) 商家数据库

含商品(库存)、商家、订单等信息表^[23],商品及商家数据涉及商品 ID、名称、品牌、简介^[28]、规格、价格、状态、上架时间、更新日期、售后服务、库存、操作员 ID、备注、商家信用度^[9,26-27,35,40]等。陈健^[34]认为库存信息表涉及项目 ID、商品 ID、当前库存量、提醒阈值、最近一次入库量、最近一次入库时间、操作员 ID、备注等;汪亭廷^[23]认为订单信息表涉及订单 ID、用户 ID、商品 ID、名称、价格、数量、总额、评分、姓名、电话、地址、是否支付、当前状态、状态变更时间、操作员 ID、订单结束时间、备注等。

3) 推荐处理类数据库

涉及规则库(存储规则、概念模式等,支持推理)^[40]、推荐场景库(支持任务分解)、推荐算法库(存储推荐算法及组合,支持算法求解^[23])、用户兴趣模型库^[24]、随机数据库(记录过程性信息及获得方式(用户回答(注意次序)、推理(需解释))^[4]。此外,单明^[39]、汪亭廷^[23]认为推荐规则信息表包括推荐内容表(推荐 ID、规则 ID、商品 ID、当前状态、状态变更时间、备注^[39])和推荐规则表(规则 ID、状态、年龄、性别、工作、条件、结果、兴趣、支持度、置信度、备注^[23,39]);刘丽峰^[40]、罗亚^[41]认为需存储专家知识、产生式规则、商品数据和咨询记录(描述为系列参数(规则特性、参数^[40]、函数、上下文特性)并以对象-属性-值(值、取值可信度、参数被跟踪次数)形式存储)。

1.3.5 数据更新

用户兴趣更新主要基于用户兴趣模型进化机制实现,商品信息主要通过数据库自带程序自动更新^[44]。杨静^[3]、安芳^[43]、张黎黎^[44]跟踪用户访问行为以更新用户信息库;辛勤芳^[24]挖掘隐式评分等修改用户评分并存入数据库,触发推荐系统调用推荐算法,基于新信息更新用户兴趣模型库、推荐模型。

2 结束语

综上,本文从用户兴趣建模、推荐机制、信息资源管理角度阐述了国内电子商务个性化推荐核心技术研究进展:现有研究集中在推荐机制,其余两者需加强;用户兴趣建模研究用户兴趣模型表示(多为表示方法)、初始化(多为用户兴趣采集、处理)、进化(多为进化内容、方法);

推荐机制按实现方式分协同过滤推荐、基于内容推荐、基于规则推荐、基于知识推荐、基于效用推荐、基于人口统计推荐和混合推荐;信息资源管理研究集中在数据挖掘(多为关联规则、分类挖掘)、存储(分用户、商家、推荐处理类数据库)。

下一步,笔者将系统分析国内电子商务个性化推荐系统体系架构与技术实现研究成果,以供相关研究与系统实现参考。

参 考 文 献

- [1] 胡一. 基于大数据的电子商务个性化信息推荐服务模式研究 [J]. 长春: 吉林大学, 2015.
- [2] 蔡斌强. 电子商务个性化推荐研究 [J]. 中国电子商务, 2013, (24): 18-19.
- [3] 刘晶, 李妍, 侯会茹. 移动电子商务多源关联个性化推荐架构 [J]. 情报理论与实践, 2014, 37 (4): 98-100.
- [4] 李晓昀. 基于隐性反馈的个性化自适应电子商务推荐系统研究 [J]. 衡阳: 南华大学, 2008.
- [5] 代其锋. 基于电子商务的个性化推荐研究 [D]. 合肥: 中国科学技术大学, 2011.
- [6] 丁建军. 面向电子商务个性化推荐系统的 Web 数据挖掘应用研究 [D]. 杭州: 浙江理工大学, 2015.
- [7] 蒋翀. 电子商务个性化推荐系统研究 [D]. 长沙: 中南大学, 2010.
- [8] 何志敏. 基于数据挖掘的电子商务个性化推荐算法的研究及应用 [D]. 贵阳: 贵州大学, 2011.
- [9] 杨静. 电子商务中个性化推荐模型的研究 [D]. 天津: 天津师范大学, 2007.
- [10] 李婷. 基于用户兴趣与合作的个性化电子商务推荐系统研究 [D]. 哈尔滨: 哈尔滨工业大学, 2007.
- [11] 张建伟. 基于 WEB 文本挖掘的电子商务网站个性化推荐研究 [D]. 淮南: 安徽理工大学, 2014.
- [12] 李凤慧. 基于用户浏览行为挖掘的电子商务个性化推荐系统 [J]. 潍坊学院学报, 2004, 4 (2): 66-67, 71.
- [13] 李明晶. 基于 Web 挖掘的电子商务个性化推荐系统研究 [D]. 北京: 北京交通大学, 2008.
- [14] 宋伟国. 基于 CF 的个性化电子商务推荐系统研究 [D]. 兰州: 兰州大学, 2010.
- [15] 安芳. 电子商务个性化信息推荐服务的研究 [D]. 北京: 对外经济贸易大学, 2006.
- [16] 张黎黎. 电子商务个性化推荐模型研究 [D]. 重庆: 重庆大学, 2009.
- [17] 李熠. 引入信任的二部图电子商务个性化推荐算法改进研究 [D]. 成都: 电子科技大学, 2015.
- [18] 杨永亮. 基于时间敏感的电子商务个性化推荐方法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2015.
- [19] 王伟, 徐平平, 王华君, 等. 基于概率回归模型和 K-最近邻的电子商务个性化推荐方案 [J]. 湘潭大学学报: 自然科学版, 2016, 38 (1): 97-100, 119.

- [20] 邓晓懿. 移动电子商务个性化服务推荐方法研究 [D]. 大连: 大连理工大学, 2012.
- [21] 俞正骄. 面向电子商务个性化推荐的序列关联挖掘研究 [D]. 天津: 河北工业大学, 2015.
- [22] 易明. 基于 Web 挖掘的电子商务个性化推荐机理与方法研究 [D]. 武汉: 华中科技大学, 2008.
- [23] 汪亭廷. 美味网电子商务个性化推荐系统的设计与实现 [D]. 成都: 电子科技大学, 2014.
- [24] 王星. 个性化推荐系统在电子商务网站中的应用研究 [J]. 商, 2013, (7Z): 156-156.
- [25] 张晓敏. 电子商务个性化推荐系统中协同过滤技术及应用研究 [D]. 重庆: 重庆大学, 2007.
- [26] 辛勤芳. 电子商务个性化推荐系统研究 [D]. 泉州: 华侨大学, 2012.
- [27] 翟烁. 基于协同过滤的电子商务个性化推荐的研究与分析 [D]. 北京: 北京工业大学, 2015.
- [28] 马瑞敏, 卞艺杰, 陈超, 等. 基于 Hadoop 的电子商务个性化推荐算法——以电影推荐为例 [J]. 计算机系统应用, 2015, 24 (5): 111-117.
- [29] 黄莹. 经济昆虫电子商务系统的实现与个性化推荐技术研究 [D]. 昆明: 昆明理工大学, 2007.
- [30] 刘蓓琳. 电子商务用户个性化推荐技术接受影响因素研究 [D]. 徐州: 中国矿业大学, 2011.
- [31] 朱岩, 林泽楠. 电子商务中的个性化推荐方法评述 [J]. 中

国软科学, 2009, (2): 183-192.

- [32] 陈晓红. 面向电子商务的个性化推荐技术研究 [D]. 扬州: 扬州大学, 2007.
- [33] 吴昕方. 基于特征的个性化电子商务网站推荐系统的研究与实现 [D]. 上海: 华东师范大学, 2006.
- [34] 邹晶晶. 基于大数据的电子商务个性化商品推荐方法 [J]. 电子制作, 2014, (10): 97-97, 96.
- [35] 刘庆华. 个性化推荐技术及其在电子商务中的应用 [D]. 南昌: 南昌大学, 2008.
- [36] 陈健. 基于个性化推荐的旅游移动电子商务平台设计研究 [D]. 武汉: 湖北工业大学, 2015.
- [37] 鲍丽倩, 陈思璇. 基于效用理论的电子商务个性化推荐 [J]. 科技情报开发与经济, 2011, 21 (24): 150-153.
- [38] 王海超. 基于 Web 挖掘的电子商务个性化推荐技术研究 [D]. 赣州: 江西理工大学, 2010.
- [39] 单明. 基于个性化推荐的电子商务推荐系统的设计与实现 [D]. 长春: 吉林大学, 2014.
- [40] 刘丽峰. 刍议基于专家系统的电子商务个性化推荐技术 [J]. 电子技术与软件工程, 2015, (10): 191-191.
- [41] 罗亚. 基于专家系统的电子商务个性化推荐技术 [J]. 制造业自动化, 2012, 34 (10): 100-102.

(本文责任编辑: 马 卓)

(上接第 107 页)

者少数的专家团队、资源评估委员会, 是资源建设应该秉持的重要原则。评价这些读者的需求是否真实可靠, 是否具有相当的学术价值, 单纯依靠读者的点击是缺乏可靠性的。资源建设也不局限于图书文献, 还有大量的电子文献和数据库, 如果仅仅依靠事先设定的读者浏览情况而进行资源建设^[4], 恐怕也是不合适的。

本文提出的资源建设与资源评价联动模式不涉及到具体的资金运作, 只是逻辑上对学科服务视角下的资源评价与资源建设应该如何联动, 应该注重文献资源的学术价值与学科建设, 而不是简单地以读者的需求来进行资源建设的构想。不仅理论上可行, 模型也不随文献资源采购政策的变化而受到影响。模型只是把图书馆的服务与资源的评价、资源的建设当作一个整体来看待, 充分利用已购资源尤其是数据库系统提供的文献计量与学术计量指标, 指导图书馆的资源建设。

参 考 文 献

- [1] 张甲, 胡小菁. 读者决策的图书馆藏书采购藏书建设 2.0 [J]. 中国图书馆学报, 2011, (2): 36-39.
- [2] 唐吉深. 我国读者决策采购 (PDA) 研究述评 [J]. 图书馆学研究, 2015, (2): 22-28.
- [3] 彭骏, 陆敏, 杨发毅, 等. 基于“二次选择”算法的高校图书馆图书采购决策系统构建 [J]. 情报理论与实践, 2009, 32 (6): 74-77.
- [4] 林笠, 马颖慧. 遗传算法在图书采购决策中的应用 [J]. 广

东工业大学学报, 2003, (3): 23-26.

- [5] 张必兰, 吴诗贤. 基于粒子群算法的图书采购复本量优化决策 [J]. 情报杂志, 2007, (8): 137-139.
- [6] 高校图书馆数字资源采购联盟 (DRAA) 简介 [EB/OL]. <http://www.libconsortia.edu.cn/Space/view.action?pagecode=gylm>, 2016-09-26.
- [7] 张云, 茆意宏. 社会化阅读平台的用户互动关系探析 [J]. 情报理论与实践, 2014, 37 (12): 99-103.
- [8] 李慧美. 从豆瓣网看图书馆学科信息服务的改进 [J]. 图书馆杂志, 2009, 28 (8): 35-38.
- [9] 席剑霄. OAuth2.0 协议及其应用模式研究 [J]. 软件导刊, 2015, 14 (7): 23-26.
- [10] Google Books. Developer's Guide [EB/OL]. https://developers.google.com/books/docs/viewer/developers_guide, 2016-09-26.
- [11] 豆瓣 douban 开发者. 图书 Api V2 [EB/OL]. https://developers.douban.com/wiki/?title=book_v2, 2016-09-26.
- [12] Google. OAuth 2.0 Scopes for Google APIs [EB/OL]. <https://developers.google.com/identity/protocols/googlescopes>, 2016-09-26.
- [13] ELSEVIER Developers. Serial Title API [EB/OL]. <http://api.elsevier.com/documentation/SerialTitleAPI.wadl>, 2016-09-26.
- [14] 中国科学院文献情报中心世界科学前沿分析中心. 在线数据平台开放数据访问接口 [EB/OL]. <http://www.fenqubiao.com/Core/Page.aspx?id=5>, 2016-09-26.

(本文责任编辑: 郭沫含)