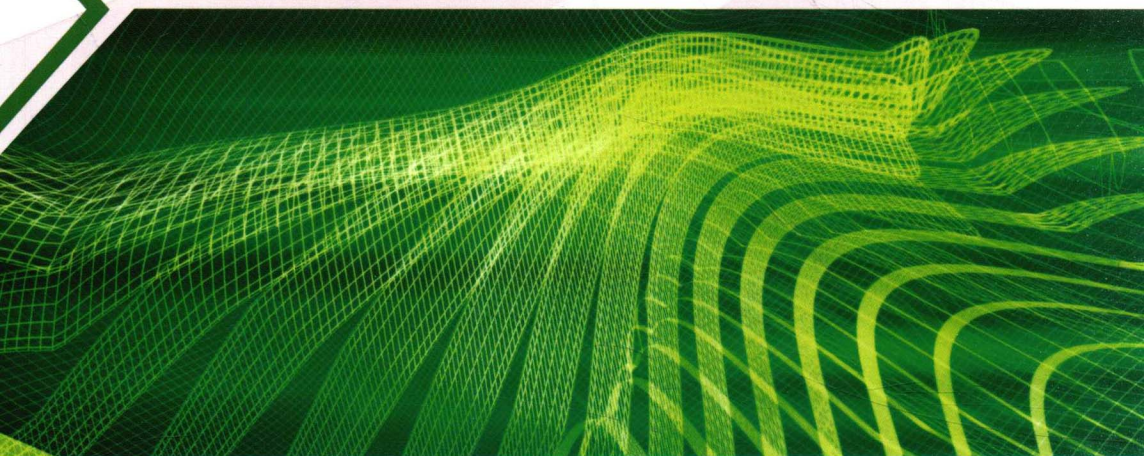




知识管理与知识服务研究

王伟军 主编

基于屏幕视觉热区的 用户偏好提取及个性化推荐



刘 凯 著



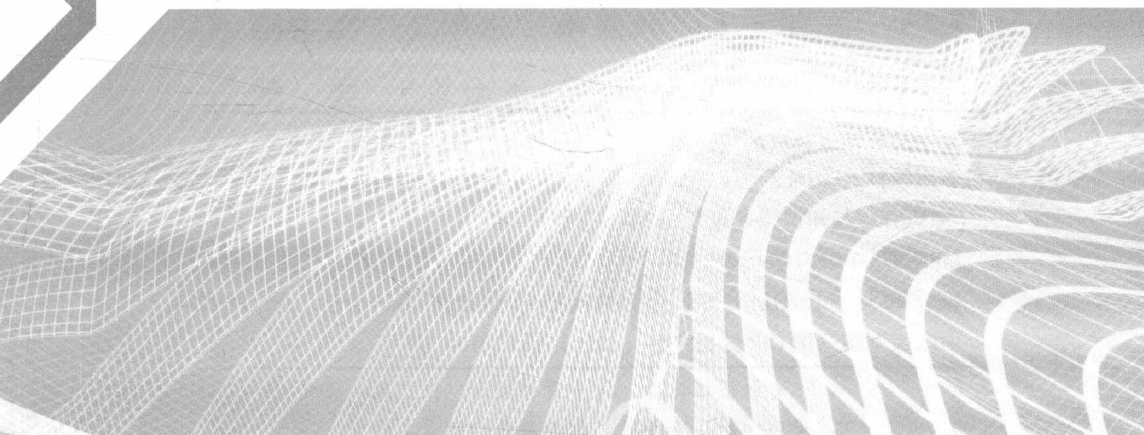
科学出版社



知识管理与知识服务研究

王伟军 主编

基于屏幕视觉热区的 用户偏好提取及个性化推荐



刘凯 著

科学出版社

北京

内 容 简 介

个性化推荐成为当前学术界和实业界的研究热点,然而用户偏好提取的粒度和精度却成为制约个性化推荐效果的主要问题之一。本书在用户为中心的理論视角下,通过心理学眼动实验发现并证实了屏幕视觉热区的存在,从而将用户行为、偏好提取及分析单元细化至用户注视过的信息层面。借助屏幕视觉热区,本书研究了用户实时注视的短文本关键词提取方法,以及利用提取的关键词构建长期、短期、即时的用户偏好复合模型,并最终提供具有推荐解释功能的交互收敛式个性化推荐服务。

本书适用于高等院校管理科学、信息科学、计算机科学和心理学相关专业的师生阅读,也可供自然科学、工程技术乃至电子商务领域的开发人员参考。

图书在版编目(CIP)数据

基于屏幕视觉热区的用户偏好提取及个性化推荐 / 刘凯著. —北京: 科学出版社, 2016. 2

(知识管理与知识服务研究/王伟军主编)

ISBN 978-7-03-047305-9

I. ①基… II. ①刘… III. ①互联网络-数据处理-算法分析 IV. ①TP274

中国版本图书馆 CIP 数据核字 (2016) 第 020164 号

责任编辑: 林 剑 / 责任校对: 张凤琴

责任印制: 徐晓晨 / 封面设计: 耕者工作室

科学出版社 出版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京中石油彩色印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 2 月第 一 版 开本: 720×1000 B5

2016 年 2 月第一次印刷 印张: 15 3/4

字数: 320 000

定价: 110.00 元

(如有印装质量问题, 我社负责调换)

《知识管理与知识服务研究》丛书 编委会

主 编 王伟军

副主编 王学东 娄策群 夏立新

编 委 (以姓氏汉语拼音为序)

段 钊 李进华 李延晖 李玉海

刘 华 刘可静 刘清堂 卢新元

田 鹏 吴建华 易 明 张自然

总 序

知识，作为社会经济活动的基本要素，已成为社会经济发展的基本资源和根本动力，人类因此进入知识经济和知识社会的新时代。但是，新的知识环境在促进社会发展和人类进步的同时，也让我们置身于知识生态的重重矛盾之中：一方面知识存量激增，并呈爆炸性增长；另一方面知识稀缺严重，人们生活在知识的海洋中，却难以获得所需要的知识。一方面知识产生速度加快，新知识源源不断；另一方面知识老化加速，知识更新周期缩短。一方面知识广泛传播，互联网络提供了知识传播的新途径，跨越了知识扩散的时空障碍；另一方面数字鸿沟日趋明显，城乡差距、地区差异、人群差别影响知识的扩散。因此，如何有效地管理和开发利用知识资源，更好地满足人们日益增长和迫切的知识需求，是人类自我完善和自我发展的需要，更是推动知识创新与知识经济发展的前提和基础，是社会全面协调和科学发展的关键。

知识管理与知识服务诞生于知识经济逐渐兴起、信息技术飞速发展、商业竞争日益加剧的环境中，广泛融合了信息科学、管理学、图书情报学等多学科理论与方法，形成了以“知识”为核心和研究对象的一个新的跨学科研究领域。从管理学视角，知识管理是将组织可获得的各种来源的信息转化为知识，并将知识与人联系起来的过程，强调对显性知识和隐性知识的管理与共享，利用集体的智慧提高组织的应变和创新能力；而知识服务是知识管理领域的演变进化，是随知识管理发展而延伸的概念，是新兴的服务科学、管理和工程学科（SSME）的重要分支。从图书情报视角，知识管理是信息管理的进一步发展，知识服务是信息服务的深化与拓展，知识服务的功能应建立在信息管理和知识管理的基础之上，以满足用户的知识需求和实现知识增值为目标。因此，知识管理是知识服务的基础，知识服务是知识管理的延伸，也是知识管理实现知识创新目标的有效途径。知识管理与知识服务也逐渐成为图书情报学、管理学和信息科学等多学科关注的重要领域和研究热点。

华中师范大学信息管理系及其相关院所的部分教师,长期以来围绕“信息—信息资源—知识的组织与管理、服务与开发利用”等方面,展开积极的探索,从人、环境、信息及其交互关系的视角,运用图书情报学、心理学、管理学、信息科学等多学科的理论和研究方法,开展知识管理与知识服务基础理论、知识组织与检索、知识管理评价与优化、知识管理与知识服务系统及其关键技术、知识转移与知识创新等方面的研究。先后承担或参与了国家“863”计划、国家“十一五”科技攻关计划、教育部高等学校学科创新引智计划、教育部新世纪优秀人才支持计划、国家自然科学基金和国家社会科学基金等多个国家级项目和省部级课题,取得了一系列的研究成果,产生了一定的社会和学术影响,并有多位教师入选教育部新世纪优秀人才支持计划。通过这些重要项目的引领和驱动,华中师范大学逐渐显现出知识管理与知识服务方面的研究特色与发展潜力,基本形成了以信息管理系部分教师为主体的充满激情和活力的研究队伍。为了进一步凝聚学科发展方向,提升学科发展的核心竞争力,学校特成立知识管理与知识服务研究中心,定位于跨学科、创新性的研究平台,以更好地团结和组织相关研究人员开展跨学科联合攻关,服务于国家战略和区域经济社会发展。

知识管理与知识服务研究中心的一项重要工作就是搭建一个开放式的学术交流平台,经常性地开展学术讲座、专题研讨和学术沙龙等活动,并及时精选研究团队中有价值的研究成果予以发展。现在将首次呈现在读者面前的《知识管理与知识服务研究》丛书共有10部著作:《Web 2.0 信息资源管理》(王伟军等),《XML 文档全文检索的理论和方法》(夏立新),《网格知识管理与服务》(李进华),《基于 Web 挖掘的个性化信息推荐》(易明),《供应链中的知识转移与知识协同》(李延晖),《区域产业集群中的知识转移研究》(段钊),《知识交流中的版权保护与利益平衡研究》(刘可静),《数字图书馆评价方法》(吴建华),《知识流程服务外包》(王伟军、卢新元等),《IT 外包服务中的知识转移风险研究》(卢新元)。这些著作都是从国家级项目的研究成果或博士学位论文中精选出来,经过进一步补充与完善而写成的学术专著。

以上选题涉猎虽广,但都聚焦于“知识”或“知识流”这一核心,置之于一代代互联网环境,关注知识的组织、交流与共享、转移与创新、评价与服务,分别立足于宏观基础、中观产业和微观组织层面展开相关研究。例如,宏观层面的基于 Web 2.0 的信息资源与知识管理变革、网格知识管理与服务的实现、知

识交流中的知识产权保护与利益平衡研究；中观产业层面的区域产业集群中的知识转移与知识创新、供应链中的知识转移与知识协同、知识流程服务外包研究；微观组织或具体应用层面的 XML 文档全文检索的理论与方法、基于 Web 挖掘的个性化信息服务、数字图书馆评价方法等。从中我们不难发现，这些研究都是针对现实中具体的理论与应用问题展开的积极探索，具有很强的跨学科性，显著的创新性和前沿性。

知识管理与知识服务仍是一个新兴的跨学科领域，需要我们大胆地探索。丛书是开放性的学术平台，今后还会不断推出优秀的研究成果，旨在促进我国知识管理与知识服务的理论创新与应用研究，形成有中国特色的知识管理与知识服务理论和方法体系，指导我国知识管理与知识服务的应用实践，为促进我国知识经济的发展和创新型国家建设做出积极的贡献。

本套丛书的出版得到了华中师范大学研究生处、社科处、科技与产业处和信息管理系的大力支持，也得到了科学出版社的鼎力相助，在此表示衷心的感谢！

王伟军

武汉桂子山

2009 年 3 月 28 日

前 言

千禧之年，革故鼎新，信息技术浪潮席卷全世界。时至今日，人类社会已经随之发生了三次深刻变革：起初，中央处理器（CPU）、图形处理器（GPU）计算性能的大幅提升拉开信息化的帷幕，摩尔定律令人叹为观止，数字自然界应运而生。随后，网络技术推动社交网络迅速蹿红，人类的社会关系映射到数字自然界中进行更为复杂的延展。时下，存储技术的突飞猛进宣告了大数据时代的到来，海量数据存储实现效率与效益兼顾，令数字自然界生机盎然。从PC机到平板电脑，从固定电话到智能手机，从闭路电视到虚拟现实头盔，人们无时无刻不被数据浸润着，物理自然与人类社会已经悄然融入数字自然界之中。数据将像土地、石油和资本一样，成为经济运行中的根本性资源。然而，人类认知水平和认知能力的提高却极为有限，远不及技术发展和数据膨胀的速度，不断增长的数据与有限的认知能力之间形成尖锐矛盾，信息过载问题越来越受到关注和重视。

作为继搜索引擎之后兴起的新星，个性化推荐系统通过向用户提供更具针对性的服务，从而有效缓解了信息过载问题的影响。从被动等待用户输入检索词，到主动了解用户需求，个性化推荐系统在解决用户信息过载问题上具有与生俱来的巨大优势，被人们广为认可。但是，也必须清醒地看到当前个性化推荐系统仍然存在着准确性不高、实时性差、个性化程度低、用户互动性弱等诸多不足，尤其是广泛使用的协同过滤算法先天具有冷启动、数据稀疏性等问题。如果无法有效地解决这些问题，个性化推荐也就无从获得更大的提升。因此，本书以用户为中心，着眼于解决用户信息过载这一本质性问题，通过解决个性化推荐系统现存顽疾而实现推荐系统中用户个性化更为精准识别的突破，进而采用更加实时和互动性的算法为用户进行准确而多样的推荐。本书首先介绍了研究背景、主要概念、研究目的、研究内容等问题。其次，较为细致地分析了个性化推荐系统的研究现状。按照认知论、方法论和矛盾论的逻辑线索对个性化推荐系统的研究现状

进行了梳理,介绍了个性化推荐系统满意化研究及外围相关研究的进展,并指出个性化推荐系统面对的五大根本矛盾。再次,提出用户为中心的个性化推荐系统理论体系,主要包括个性化推荐系统的历史背景及相关理论、目的、类型、本质、特征维度及其发展瓶颈。在此基础上,利用眼动实验证明屏幕视觉热区的存在,在双中线降噪的网页自动分类算法及短文本关键词实时提取方法的基础上,构建了较为完整的基于屏幕视觉热区的用户偏好提取方法。接下来,以提取屏幕视觉热区的用户偏好数据为基础,以商品自组织层次聚类方法和基于兴趣的用户会话切分算法为方法,构建融合即时、短期和长期偏好的用户偏好复合模型。继而,基于评价介入理论构筑分析体系,利用话语标记理论构建语料库,探索了个性化推荐系统中的推荐解释项目及风格。最后,以用户即时偏好为基础,以加权的用户短期和长期偏好以及其他情境因子为约束条件,设计具有推荐解释功能的、人机实时互动的个性化推荐方法,通过不断叠加约束条件而迅速收敛到用户满意的结果范围。

本书的创新之处主要体现在:

理论上,明确指出个性化推荐系统的本质不是算法而是认知助手,构建了推荐内容(what)、推荐策略(how)和推荐时机(when)的“WHW”个性化推荐系统理论体系,以及综合即时偏好、短期偏好和长期偏好的用户复合偏好模型。

方法上,首先,借助心理学实验发现了屏幕视觉热区,并据此从用户实时浏览行为中提取用户即时偏好,在基于双中线法消除噪声的网页自动分类算法和短文本关键词实时提取方法的基础上,实现了用户偏好的实时提取;其次,通过基于属性的商品自组织层次聚类方法和用户会话切分算法,从海量即时偏好数据中提炼用户短期和长期偏好;再次,引入功能语言学中的评价介入理论构筑分析体系,利用话语标记理论构建语料库,确定在线商品评分修正方案及推荐解释风格;最后,以用户历史行为、偏好复合模型、即时交互行为为约束条件,开发原型系统测试并验证交互收敛式个性化推荐方法,与传统推荐方法相比本书提出的方法不仅具有更高的精确性和用户满意度,还能有效解决推荐算法实时性差、未登录用户偏好提取难等问题。

特别的,本书方法具有显著的应用价值,是各类中小型电子商务网站(或

中小型新闻、论坛网站等) 建设高精度、高性能个性化推荐系统的福音。通常现有个性化推荐算法的应用门槛颇高: 不仅需要足够的软硬件资源投入, 还要有丰富的注册用户以及用户标记数据作为支撑。然而, 绝大多数的中小型网站都处于起步或发展阶段, 往往并不具备开展传统个性化推荐所需要的高性能计算资源、丰富的用户数量和数据等门槛条件。本书的推荐方法不仅对大型电子商务网站非常有效, 对于各类中小型电子商务网站而言则更为适合。不论是基于屏幕视觉热区的偏好获取方法, 还是交互收敛式个性化推荐方法, 都属于计算资源消耗低但效果明显, 且能满足实时性要求。对于数量浩繁、种类各异的中小型网站来说, 书中提出的原创思想、观点、解决方案等都是其实施个性化推荐的理想选择。

本书只是在“人”的视角下对个性化推荐系统理论和方法的突破性尝试, 加之作者水平有限, 书中难免有不足之处, 恳请读者批评指正。意见请发表在作者的技术博客 (lk. cublog. cn) 或新浪微博 (@五-岳-之-巅) 上, 亦可直接通过 E-mail (ccnulk@gmail. com) 进行联系, 真诚欢迎和期待更多同行的加入和交流。

刘 凯
2015 年 11 月

目 录

前言

1 绪论	1
1.1 研究背景	1
1.2 研究范围及相关概念	4
1.3 研究目的	6
1.4 研究内容与方法	8
1.5 创新及意义	12
2 相关理论与研究现状	15
2.1 综述体系说明	15
2.2 精确性研究	15
2.3 满意化研究	31
2.4 其他相关研究	36
2.5 理论评述——矛盾论	37
3 用户为中心的个性化推荐系统理论体系	40
3.1 历史背景	40
3.2 学科基础	41
3.3 个性化推荐系统理论核心体系	48
3.4 个性化推荐系统的理论瓶颈	54
3.5 总结	57
4 基于屏幕视觉热区的用户偏好提取方法	58
4.1 用户浏览行为中的屏幕视觉热区	58
4.2 面向用户偏好识别的网页机能分类及其判别方法	69
4.3 基于屏幕视觉热区的中文短文本关键词实时提取方法	78
4.4 总结	90

5 基于屏幕视觉热区的用户偏好建模	92
5.1 基于商品特征自组织层次聚类的用户偏好模型.....	92
5.2 基于商品特征自组织层次聚类的网络用户会话切分研究	113
5.3 基于屏幕视觉热区的用户偏好复合模型	124
5.4 总结	134
6 基于在线商品评分修正的推荐解释	135
6.1 研究背景	135
6.2 相关研究	136
6.3 评价介入理论	141
6.4 话语标记	153
6.5 基于评价介入理论和话语标记的在线评论修正方法	158
6.6 基于在线商品评论的推荐解释风格	170
6.7 总结	172
7 基于用户偏好复合模型的交互收敛式个性化推荐方法	173
7.1 研究背景	173
7.2 相关研究	174
7.3 基本原理	176
7.4 交互行为与约束条件	179
7.5 交互收敛式的实时个性化推荐方法	181
7.6 原型系统构建	186
7.7 评价指标	193
7.8 实验验证	195
7.9 总结	203
8 总结与展望	204
8.1 总结	204
8.2 展望	206
参考文献	209
索引	233

1 绪 论

1.1 研究背景

1.1.1 挑战：世界范围内网络用户信息过载问题严重

就在此刻，远在大洋彼岸的互联网实时数据统计网站 Internet Lives Stats 展示了如下惊人的即时数据。一方面，仅仅 1 秒内，新增 7894 条 Tweet 信息、1414 张 Instagram 图片、1549 篇 Tumblr 博客、1568 次 Skype 通话，产生 46 027 次 Google 搜索、90 368 次 YouTube 视频观看，并发送 2 343 302 封电子邮件，同时消耗 23 950 GB 网络流量。另一方面，此时全世界网民数量共计 2 994 642 899 人，虽然基数巨大但增幅却明显减缓，如表 1-1 所示。从 1994 年的年均增长率 79.7% 到 2004 年的 16.9%，再到 2014 年的不足 8%，而人口增长率却在轻微递减，说明在人口增长率维持基本稳定的情况下新增网民数量已经向临界值靠拢，网民总数也将处于稳定状态。

表 1-1 世界网络用户人口统计表

年份	网民总人数/人	网民增长率/%	世界人口总数/人	人口增长率/%	渗透率/%
2014	2 925 249 355	7.9	7 243 784 121	1.14	40.4
2013	2 712 239 573	8.0	7 162 119 430	1.16	37.9
2012	2 511 615 523	10.5	7 080 072 420	1.17	35.5
2011	2 272 463 038	11.7	6 997 998 760	1.18	32.5
2010	2 034 259 368	16.1	6 916 183 480	1.19	29.4
2009	1 752 333 178	12.2	6 834 721 930	1.20	25.6
2008	1 562 067 594	13.8	6 753 649 230	1.21	23.1
2007	1 373 040 542	18.6	6 673 105 940	1.21	20.6
2006	1 157 500 065	12.4	6 593 227 980	1.21	17.6

续表

年份	网民总人数/人	网民增长率/%	世界人口总数/人	人口增长率/%	渗透率/%
2005	1 029 717 906	13. 1	6 514 094 610	1. 22	15. 8
2004	910 060 180	16. 9	6 435 705 600	1. 22	14. 1
2003	778 555 680	17. 5	6 357 991 750	1. 23	12. 2
2002	662 663 600	32. 4	6 280 853 820	1. 24	10. 6
2001	500 609 240	21. 1	6 204 147 030	1. 25	8. 1
2000	413 425 190	47. 2	6 127 700 430	1. 26	6. 7

资料来源：Internet Live Stats（国际电信联盟及美国人口统计局数据的合并）

因此，网络数据量的迅猛增长与网络用户数量的基本稳定，使得每位网络用户要面对数量更多、类型更复杂且生成速度越来越快的海量数据，从而必然导致个体用户的平均信息负荷急剧升高，信息过载问题由此产生。

从全球来看，世界网民半数位于亚洲，如图 1-1 所示。亚洲占全球网民人数的 48.4%，其中中国以 22.0% 高居榜首，印度和日本则分别以 8.3% 和 3.7% 紧随其后。尽管中国和日本是 IT 领域潜力巨大的消费市场，但其分别为 4% 和 8% 的年新用户增长量却对发展模式发出了预警：用户规模很难在短期内有量级的突破，依托人口红利维持快速发展的信息服务模式即将成为历史，充分了解用户、提升服务精度和效率才是未来之选，也是必由之路。总之，一条由量向质的转变之门已经打开，除了适应与接受外我们别无选择。

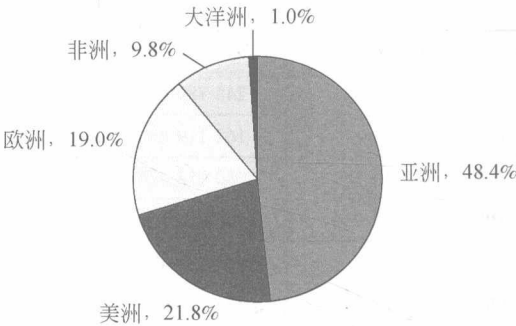


图 1-1 世界网民人数洲际分布图

资料来源：Internet Live Stats，2013-07-01

1.1.2 机遇：大数据分析——点“数”成金

大数据概念虽新，可“大数据”却早已存在。虽然海量的数据规模是最近几年内才形成，但大数据的概念却早已在科学界和医学界中萌芽。科学家们对大数据集进行研究和分析后得出一个结论——数据多多益善。也就是说，数据越多，分析越深入，所得的结论就越全面。于是，研究者们开始在分析过程中引入相关数据集、非结构化数据、归档数据和实时数据，而这反过来又催生出我们今日所说的“大数据”。在商业领域中，大数据暗藏商机。据 IBM 公司称，全球每天产生 2500 拍字节的数据，当今世界 90% 的数据都是近些年产生的。这些数据来源广泛，有的来自收集气候信息的传感器，有的来自社交媒体网站，还有的是网络上传的数字照片和视频、电子商务交易记录甚至手机 GPS 信号等。这些都是大数据的“催化剂”，所有的数据都蕴含内在价值，可以利用数据挖掘、机器学习和可视化等技术将内在价值进行提炼并为用户提供更具个性的推荐服务。

餐饮业中的健康餐饮分析推荐系统具有在线健康餐饮分析的功能，通过记录用户的饮食行为数据，推荐可以改善健康状况的食品，同时降低发生疾病的风险；家具制造企业利用眼动追踪设备捕获用户对商品的生理感知数据，借助数据挖掘技术建立家具商品风格分类数据库，随后将风格统一的商品推荐给潜在用户，用户可以在虚拟的立体环境中感受商品，从而促进家具的网络销售；新闻行业中，《纽约时报》利用大数据工具进行文本分析和 Web 挖掘，迪士尼公司则分析了旗下所有店铺、主题公园和网站的数据，试图发现数据间的关联性，进而理解用户行为；在传统农业中，科学家对匈牙利 1960 ~ 2000 年农田施肥实验数据进行深入挖掘，设计出精准施肥推荐系统，能够实现在不同土壤中降低氮磷钾损失率而达到产量和单位面积收益的最大化。这样的例子不胜枚举，大数据的应用已经相当广泛，在电子商务、金融业、制造业、医药保健业及政府机关中都能见到它的身影。

但是也必须看到，在大数据时代，信息过载问题变得更为棘手、更为严峻。经过多年努力，人们提出并实践了从搜索引擎到推荐系统再到个性化推荐系统的多种方案，个性化推荐系统的发展成果彰显了其实际应用价值，并逐渐成为缓解信息过载问题的生力军和有效武器而得到了实业界和学术界的双重垂青。虽然个性化推荐系统前进的步伐势不可挡，但不可否认其自身也存在着诸多不足和顽疾。本书便是针对当前个性化推荐系统存在的若干根本问题进行反思与突破，力图在构建个性化推荐理论体系的基础上，提出一整套新颖而高效的实时个性化推

荐方法来有效解决这些问题,从而为网上信息服务质量和速度乃至用户综合体验的提升作出理论性和应用性两个层面的贡献。

1.2 研究范围及相关概念

1.2.1 研究范围界定

个性化推荐系统发展至今已然形成了一个成熟而庞大的体系。理论上,对个性化推荐系统的研究必定涉及多学科的交叉与融合,在实践中又涉及诸多应用领域,不同领域有着不同的知识结构并对应着不同的学科基础。因此,为了使本书的研究既具有理论深度又做到言之有物和切实可行,则必须对研究范围进行限定,不能泛泛谈之、流于形式。

理论上,个性化推荐系统主要的研究问题多局限于用户偏好获取手段、推荐具体算法上,鲜有文章以整体视角对其涉及的各个理论问题进行梳理与整合。本书将从以下几个方面进行个性化推荐的理论构筑:个性化推荐系统产生的历史背景及各个主要的发展阶段、理论来源、本质、目标、分类、结构以及个性化推荐系统在信息处理和心理学两种视角下的系统功能。

方法上,本书的研究涉及推荐系统进行推荐的主要过程(但有不同侧重):用户数据采集(显性数据和隐性数据)、用户偏好建模、选取推荐算法、生成推荐结果、推荐反馈五个基本步骤。首先,数据采集步骤中来源数据的质和量是后续推荐结果准确性和多样性的基础,但这一步也是极易被忽视或忽略的,于是如何发现和挖掘出更为适当的数据源便是本书所要解决的一个重要问题;其次,利用丰富的用户数据构建出合适的偏好模型,通过引入即时偏好改进现有的长期、短期用户偏好模型,以求实现更为动态更加实时的用户兴趣表达;最后,对个性化推荐算法进行革新,通过与用户的互动充分挖掘用户实时意图而提出一种精确度更高的实时推荐算法,实现推荐结果准确性和多样性的兼顾。

实践中,个性化推荐的算法验证及实施效果必须融入特定的场景才有意义,本书选择的是电子商务情境,如未做特别说明,所有构想和验证的都是基于电子商务情境下的个性化推荐。但是,需要说明的是,本书构建的理论体系和一系列新颖的方法不仅适用于推荐产品和服务的电子商务网站,也同样适用于门户、新闻以及各类社区站点的主题推荐、话题推荐、文章推荐等。

1.2.2 相关概念界定

1. 推荐系统

推荐系统 (recommender system, RS) 指的是主要用于电子商务网站或其他在线服务的一类软件, 帮助在线顾客快速发现最为相关的商品项目或信息条目, 如今已广泛应用在图书、电影、音乐、旅社、餐馆及新闻等领域中。推荐系统概念在英文中有多种表达方式, 即 recommender system、advisory system 或 recommendation system, 而基于语义的推荐系统是其中非常重要的一个分支, 称为 semantic-based recommender system, 与此类似的概念还有 social recommender system、tag-based recommendation 和 web 2.0 recommender system 等。

2. 个性化推荐系统

个性化推荐系统 (personalized recommender system, PRS) 是指根据用户行为和用户社会关系网络中的偏好为其提供结果更为准确与可信的推荐系统。

需要特别说明的是, 国外文献中个性化推荐系统的通用写法为 personalized recommender system, 也有写为 individual recommender system 及 customized recommender system, 但后者较少使用。而在国内, 更为常见的则是 personalized recommendation system 的形式, 其实国内的这种写法的合理性有待商榷, recommender 在推荐系统情境中的含义是 “interface for providing recommendations”, 即推荐者是提供推荐内容的接口 (或界面)。由此可见 recommender 本质上是 recommendation 的施动者, 而 recommendation 则为 recommender 的具体内容。因此, 在本书中推荐系统全部采用词组 “recommender system” 来指代, 而推荐 (指具体内容和项目) 则使用单词 “recommendation” 进行标识。

3. 个性化推荐

个性化推荐 (personalized recommendation, PR) 概念的正式表达方式为 personalized recommendation, 也有 individual recommendation 或 customized recommendation 的写法, 但后两种并非主流。个性化推荐概念在国内存在着混淆的情况。例如, 余肖生、孙珊和王雨果等认为通过分析网络用户对网络信息 (主要是电子商务行为) 的搜索、浏览、购买、评价等行为, 抽取出用户可能感兴趣的有用信息, 并将其主动推荐至用户。可以看出, 上述概念将个性化推荐与个性化推荐系