

计 算 机 科 学 丛 书



Springer

# 推荐系统

## 技术、评估及高效算法

弗朗西斯科·里奇 (Francesco Ricci)

利奥·罗卡奇 (Lior Rokach)

[美] 布拉哈·夏皮拉 (Bracha Shapira) 编

保罗 B. 坎特 (Paul B. Kantor)

李艳民 胡聪 吴宾 王雪丽 丁彬钊 译 蒋凡 审核

李艳民 胡聪 吴宾 王雪丽 丁彬钊 译 蒋凡 审核

### Recommender Systems Handbook

FRANCESCO RICCI  
LIOR ROKACH  
BRACHA SHAPIRA  
PAUL B. KANTOR EDITORS

### RECOMMENDER SYSTEMS HANDBOOK

Springer

机械工业出版社  
China Machine Press

计 算 机 科 学 丛 书

# 推荐系统

技术、评估及高效算法

弗朗西斯科·里奇 (Francesco Ricci)

利奥·罗卡奇 (Lior Rokach)

[美] 布拉哈·夏皮拉 (Bracha Shapira) 编

保罗 B. 坎特 (Paul B. Kantor)

李艳民 胡聪 吴宾 王雪丽 丁彬钊 译 蒋凡 审核

Recommender  
Systems Handbook

FRANCESCO RICCI  
LIOR ROKACH  
BRACHA SHAPIRA  
PAUL B. KANTOR EDITORS

RECOMMENDER  
SYSTEMS  
HANDBOOK

 Springer



机械工业出版社  
China Machine Press

## 图书在版编目 (CIP) 数据

推荐系统: 技术、评估及高效算法 / (美) 里奇 (Ricci, F.) 等编; 李艳民等译. —北京: 机械工业出版社, 2015.6

(计算机科学丛书)

书名原文: Recommender Systems Handbook

ISBN 978-7-111-50393-4

I. 推… II. ①里… ②李… III. 计算机网络 IV. TP393

中国版本图书馆 CIP 数据核字 (2015) 第 115324 号

本书版权登记号: 图字: 01-2013-7856

Translation from the English language edition:

*Recommender Systems Handbook* by Francesco Ricci, Lior Rokach, Bracha Shapira, Paul B. Kantor.

Copyright © 2011 Springer-Verlag New York, Inc.

Springer is a part of Springer Science+ Business Media.

All Rights Reserved.

本书中文简体字版由 Springer Science+ Business Media 授权机械工业出版社独家出版。未经出版者书面许可, 不得以任何方式复制或抄袭本书内容。

本书汇聚不同领域专家学者的理论成果和实践经验, 全面介绍推荐系统的主要概念、理论、趋势、挑战和应用, 详细阐释如何支持用户决策、计划和购买过程。书中既详细讲解了经典方法, 又介绍了一些新研究成果, 内容涵盖人工智能、人机交互、信息技术、数据挖掘、统计学、自适应用户界面、决策支持系统、市场和客户行为等领域, 无论是从事技术开发, 还是从事产品营销的读者, 都能从中受益。

本书可分成五部分, 共 25 章。第 1 章是概述, 系统介绍推荐系统的概念、功能、应用领域以及当前应用过程中遇到的问题与挑战。第一部分 (第 2~7 章) 介绍当前构建推荐系统最普遍使用的技术, 如协同过滤、基于内容的数据挖掘方法、上下文相关的方法等。第二部分 (第 8~12 章) 概述已用于评价推荐质量的技术和方法, 涉及推荐系统设计与实践方面, 描述设计和实施推荐系统的注意事项, 为选择更合适的算法提供准则, 以及评估用于开发推荐系统的方法、挑战和评测指标。第三部分 (第 13~17 章) 讨论推荐系统如何呈现、浏览、解释和可视化等若干问题, 这一部分讨论的技术使推荐过程更加结构化以及具有可交互性。第四部分 (第 18~21 章) 讨论利用各类用户生成内容 (UGC, 如标签、搜索查询、信任评价等) 产生类型新颖且更可信的推荐结果。第五部分 (第 22~25 章) 讨论推荐系统的高级课题, 如探索用主动学习的原则来指导获取新知识; 防止推荐系统受恶意用户攻击的合适技术; 如何整合多种类型的用户反馈和用户偏好信息来构造更可靠的推荐系统。

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 秦 健

责任校对: 殷 虹

印 刷: 北京瑞德印刷有限公司

版 次: 2015 年 7 月第 1 版第 1 次印刷

开 本: 185mm×260mm 1/16

印 张: 36.25

书 号: ISBN 978-7-111-50393-4

定 价: 139.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzsjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的优势，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自1998年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson 等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力相助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作；而原书的作者也相当关注其作品在中国的传播，有的还专门为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：[www.hzbook.com](http://www.hzbook.com)

电子邮件：[hzsj@hzbook.com](mailto:hzsj@hzbook.com)

联系电话：(010)88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



面对互联网上的海量信息，普通用户往往无所适从。为了满足用户的信息需求，计算机科学家发明了两项意义深远的技术：一是搜索技术，它帮助计算机系统被动地应回答用户的查询，将与之最相关、最重要的信息返回给用户；二是推荐系统技术，它根据用户的兴趣、行为、情景等信息，把用户最可能感兴趣的内容主动推送给用户。近年来，推荐系统技术得到了长足的发展，不但成为学术研究的热点之一，而且在电子商务、在线广告、社交网络等重要的互联网应用中尽显身手。

推荐系统是一个相对而言比较新的技术领域，市面上全面介绍其相关技术的中文书籍并不多见。我和蒋凡在微博上“神交”已久，这次听闻他在翻译整理《推荐系统：技术、评估及高效算法》一书，非常欣慰，也相信他的译本可以为对推荐系统感兴趣的广大读者带来福音。因此，欣然允诺写序文一篇，来“推荐”这本书。

本书是一本结集，由来自不同领域的资深研究人员联合撰写，其内容涵盖了推荐系统的各个方面，既有实战价值，又有理论深度，是一本受众面很广的好书。它的英文版有非常好的销售记录，也被很多学校和公司作为参考书目。将这样一本书翻译成中文有很大的价值，同时也有不小的难度。比如，由于本书篇幅很长，翻译过程需要付出巨大的时间成本；由于本书各章节的多样性，翻译过程需要很宽的知识面；由于推荐系统的技术比较新，某些技术词汇的翻译尚未达成广泛共识；由于本书兼备技术和理论，对译者的算法、数学等方面都提出了很高的要求。翻阅了蒋凡的翻译稿，能够清晰地感受到他在本书的翻译过程中所做的巨大努力。多年来蒋凡对推荐系统技术的深入研究，以及在互联网公司的亲身实践，能为本书的翻译保驾护航。总体而言，本书的翻译稿用词准确、语言流畅，并且在原书交代不够清晰的地方还加入了译者注解，更方便读者阅读和理解。相信蒋凡参与翻译的这本书会对国内推荐系统方面的研究与实践产生积极的推动作用，并且成为该技术领域的重要中文参考文献。

刘铁岩

首席研究员/主任研究员

微软亚洲研究院

2015年6月

随着互联网尤其是移动互联网的兴起，在海量的信息和数据中挖掘有价值的信息呈现给用户，成为电商、社交、新闻、影音等各大主流应用的核心功能。

推荐系统正是这样一项技术，它集数据挖掘、机器学习、用户行为学、人机交互等多个领域的知识为一体，结合日渐普及的大规模并行数据处理框架，为每位用户提供精准的、及时的乃至带有惊喜感的个性化信息服务。

本书以推荐系统领域的欧美知名学者为主要发起人，结合实际应用，为推荐系统的初学者和从业人员提供了一份权威参考资料。

对在校学生及初学者而言，若时间允许，建议比较全面地浏览本书主要章节，边读边思考，重在构建知识的体系、深入思考问题和解决问题的习惯。例如深刻理解协同过滤中 Item-based、User-based(同好)、矩阵分解等算法各自的特性，独立进行代码实现，在开放的、带有语义信息的数据集上，根据多种指标分析推荐的结果并思考算法为何会产生此类结果。通过这个过程，将书本知识转化为自己的体悟。

对推荐系统的从业者而言，本书是一本很好的工作手册，遇到问题时可以随时翻阅，但其作用不仅限于此，同时帮助大家开拓思路、提升系统性解决问题的能力。举例而言，本书第 8 章提到的推荐系统的若干评测指标，如新颖度、多样性、惊喜度、健壮性等，给我们启发良多，在手机淘宝首页逛场景推荐算法和系统中均有应用。对健壮性的考虑，可以预防频频发生的第三方攻击；针对用户惊喜度的设计及多目标优化，可以大大提升产品体验和用户留存率等核心指标。

某些章节介绍的内容较生僻，但在某些垂直行业的推荐系统中可能起着关键作用。如第 21 章中介绍的组推荐算法和策略是多用户参与的旅游活动、家庭娱乐系统的关键；第 20 章谈到的信任推荐问题是社交网络、金融服务领域开发推荐系统的核心要素。

推荐系统是一个年轻的领域，相关技术正在蓬勃发展，学有余力的朋友可以多关注推荐和相关领域顶级会议(SIGKDD、ACM RecSys、WWW、WSDM 等)及国内外业界公司的最新成果。同时，推荐系统的创新应用层出不穷，在移动终端、智能机器人、互联网金融等，只要和用户、消费者交互的地方，都可以看到它的身影。一个更智能、更懂你的互联网离不开智能推荐技术。

本书是一个开始，祝大家早日在推荐的知识海洋中自在遨游，探索未知新世界。

袁全

阿里资深技术专家，淘宝推荐负责人

**胡聪(胡户主)**:这本书着实是一本厚书,在刚开始着手翻译的时候,反复浏览目录里25章的内容,心潮澎湃,激动不已,希望放下手中一切来研习它。在大数据浪潮中,推荐系统作为一个主要分支,逐渐被国内同行所关注,但一直苦于可用资料不多。想必很多刚入此行的同学都在问“有没有全套推荐系统学习资源”,“行业内推荐系统算法有哪些”,等等问题,这本书能够带来一份很好的答案。这本书由概述到细节,由算法理论到行业应用,层层剖析了推荐系统中的技术细节和应用方向,书中知识涵盖较广却又不失细节,相信会成为推荐系统爱好者和从业人员手中必备的一本工具书。

**吴宾**:Netflix 大赛有力地推进了对推荐系统的广泛而深入的研究,促成新的研究热点。然而,国内当前还缺少比较完整的中文科普图书。因此,本书的出现可有效地填补这一空白,进一步推动推荐系统的研究和应用。综合来看,本书既可以作为推荐系统专业研究人员的入门书籍,也适用于一般应用推荐系统技术的工程人员。对推荐系统的热爱和研究是我有幸参与此书翻译工作的一个重要原因,也因此结识了许多志同道合的小伙伴。我非常感谢他们对我的帮助和指导,与他们的讨论和交流让我受益匪浅。同时,也要感谢艳民和蒋帆老师,他们不惜放弃宝贵的休息时间来审核和校对此书,严谨的工作作风和态度令人敬佩。

**丁彬钊(于宁)**:研究生期间偶然结缘推荐系统,坚信推荐必将成为继搜索之后解决数据过载的重要方法。有机会参与本书的翻译倍感荣幸,虽然期间经历实习、求职、毕业等系列事件,但是在朋友们的支持和帮助下,还是坚持完成了自己负责的那部分翻译工作。通过对此书的翻译,不仅仅提升了自己对推荐系统的理解,也增强了对推荐系统在大数据时代所起作用的信心!特别感谢李艳民对本书翻译提供的各种帮助,以及蒋凡老师细致负责的审核和校对工作!由于翻译水平和时间限制,本书翻译工作难免存在不足之处,欢迎读者朋友批评指正,谢谢!

**王雪丽**:这是一本“特别”而又“伟大”的书,“特别”在于,它集合了推荐系统所涉及领域的知识,并将这些知识有条不紊地展示出来,成为第一本专属于推荐系统领域的书;而“伟大”在于,它凝结了很多人的智慧,比如说每一章后面那长长的参考目录就体现了这点。这本书对于初学者来说是一个很好的起点,不仅可以系统地了解推荐系统的组成部分以及怎样着手设计一个推荐系统,让初学者不会觉得无从下手,另外本书也描述了推荐系统所面临的挑战,也许能在阅读的过程中突发灵感,找到自己的研究方向。另外,非常幸运能够参与本书的翻译,非常感谢一起完成本书翻译的朋友们,也希望读者朋友对我们的工作提出宝贵意见,非常感谢!

**李艳民**:本书内容和重要程度不言而喻。在长达15个月的翻译过程中共有49位同学参与翻译和审核,期间有晋升为爸爸的、有毕业参加工作的、有结婚步入幸福天堂的,也有开始创业的,在这里由衷感谢他们及其家人能参与和支持我们的翻译。同时特别感谢蒋凡老师能帮我一起审阅。希望本书能给读者带来帮助。读者有疑问或建议可在本书论坛([www.rec-sys.net](http://www.rec-sys.net))讨论。

**蒋凡**:相信翻译这本书应该是每个推荐系统爱好者的心愿,它的博大精深令人醉心不已,但它的卷帙浩繁也令人望而却步。很幸运能和不计辛劳、只求学问的小伙伴们一起完成这个心愿,翻过高山,收入眼帘的就是一马平川的美景;架起云梯,就能帮助更多的小伙伴攻城拔寨、无往不利。这本书终于面世了,我们用专注和坚韧将它唤醒,是因为我们预感它的魔力能够召唤来更多有志于此的工程师,让技术的力量改变世界。

推荐系统是为用户推荐所需物品的软件工具和技术。提供的推荐旨在通过各种决策过程来支持用户，例如，买什么物品、听什么歌或者读什么新闻。推荐系统对于在线用户处理信息过载是一个非常有价值的方法，并成为电子商务领域最强大和流行的工具。因此，人们提出了各种各样的推荐技术，并在过去的 10 年中将其中很多方法成功地运用在商务领域。

推荐系统的发展需要多学科的支持，涉及来自各个领域的专家知识，如人工智能、人机交互、信息检索、数据挖掘、数据统计、自适应用户界面、决策支持系统、市场营销或消费者行为等。本书旨在基于这种多样性，通过展示推荐系统的主要概念、理论、方法论、趋势、挑战和应用等连贯而又统一的知识体系，帮助读者从差异之中梳理出头绪。这是第一本全面阐述推荐系统的书，其中覆盖了主要技术的多个方面。本书中的丰富信息和实践内容为研究人员、学生和行业中的实践者提供了一个有关推荐系统的全面但简洁方便的参考源。本书不仅详细介绍了经典方法，而且介绍了最近引进的新方法及其扩展。本书由五部分组成：技术、推荐系统的应用和评估、推荐系统的交互、推荐系统和社区及高级算法。第一部分展示了如今构建推荐系统的最流行和最基础的技术，如协同过滤、基于内容的过滤、数据挖掘方法和基于情境感知的方法。第二部分首先介绍用来评估推荐质量的研究技术和方法；其次说明了设计推荐系统的实际方面，如设计和实现的考虑，选择更合适算法的环境指南；再次讨论了可能影响设计的相关方面；最后探讨了应用在已成型系统评估上的方法、挑战和估量。第三部分包括了探讨一系列问题的文章，这些问题包括推荐的展示、浏览、解释和视觉化，以及使得推荐过程更结构化和方便的技术等。

第四部分完全聚焦于一个全新的话题，但该话题却基于过滤推荐的主要思想，例如利用用户产生的各种类型的内容来构建具有新类型并更加可信的推荐系统。

第五部分搜集了一些关于高阶话题的文章，例如利用主动学习技术来引导新知识的学习，构建能够抵挡恶意用户攻击的健壮推荐系统的合适技术，以及结合多种用户反馈和偏好来生成更加可信的推荐系统。

我们要感谢所有为本书做出贡献的作者。感谢所有审阅人员提出的慷慨意见及建议。特别感谢 Susan Lagerstrom Fife 和 Springer 的成员，感谢他们在写这本书过程中的合作。最后我们希望这本手册有助于这一学科的发展，为新手提供一个卓有成效的学习方案，能够激起更多专业人士有兴趣参与本书所讨论的主题，使这个具有挑战性的领域能够硕果累累，长足进展。

Francesco Ricci

Lior Rokach

Bracha Shapira

Paul B. Kantor

2010 年 5 月



# 目 录

Recommender Systems Handbook

出版者的话	
推荐序一	
推荐序二	
译者序	
前言	
<b>第 1 章 概述</b> .....	1
1.1 简介 .....	1
1.2 推荐系统的功能 .....	3
1.3 数据和知识资源 .....	5
1.4 推荐技术 .....	7
1.5 应用与评价 .....	10
1.6 推荐系统与人机交互 .....	12
1.6.1 信任、解释和说服力 .....	13
1.6.2 会话系统 .....	13
1.6.3 可视化 .....	14
1.7 推荐系统是个交叉学科领域 .....	15
1.8 出现的问题和挑战 .....	16
1.8.1 本书对出现的问题的讨论 .....	16
1.8.2 挑战 .....	18
参考文献 .....	20
<b>第一部分 基础技术</b>	
<b>第 2 章 推荐系统中的数据挖掘方法</b> .....	28
2.1 简介 .....	28
2.2 数据预处理 .....	29
2.2.1 相似度度量方法 .....	29
2.2.2 抽样 .....	30
2.2.3 降维 .....	31
2.2.4 去噪 .....	33
2.3 分类 .....	34
2.3.1 最近邻 .....	34
2.3.2 决策树 .....	35
2.3.3 基于规则的分类 .....	36
2.3.4 贝叶斯分类器 .....	36
2.3.5 人工神经网络 .....	38
2.3.6 支持向量机 .....	39
2.3.7 分类器的集成 .....	40
2.3.8 评估分类器 .....	41
2.4 聚类分析 .....	42
2.4.1 k-means .....	43
2.4.2 改进的 k-means .....	44
2.5 关联规则挖掘 .....	44
2.6 总结 .....	46
致谢 .....	47
参考文献 .....	47
<b>第 3 章 基于内容的推荐系统：前沿和趋势</b> .....	51
3.1 简介 .....	51
3.2 基于内容的推荐系统的基础 .....	52
3.2.1 基于内容的推荐系统的高层次结构 .....	52
3.2.2 基于内容过滤的优缺点 .....	54
3.3 基于内容的推荐系统的现状 .....	55
3.3.1 物品表示 .....	56
3.3.2 学习用户特征的方法 .....	62
3.4 趋势和未来研究 .....	65
3.4.1 推荐过程中用户产生内容的作用 .....	65
3.4.2 超越特化：惊喜度 .....	66
3.5 总结 .....	68
参考文献 .....	68
<b>第 4 章 基于近邻推荐方法综述</b> .....	74
4.1 简介 .....	74
4.1.1 问题公式化定义 .....	75
4.1.2 推荐方法概要 .....	76
4.1.3 基于近邻方法的优势 .....	77
4.1.4 目标和概要 .....	78
4.2 基于近邻推荐 .....	78

4.2.1 基于用户评分 .....	79	5.6 基于邻域的模型和因子分解 模型的比较 .....	127
4.2.2 基于用户分类 .....	80	参考文献 .....	129
4.2.3 回归与分类 .....	80	<b>第6章 开发基于约束的推荐 系统</b> .....	131
4.2.4 基于物品推荐 .....	81	6.1 简介 .....	131
4.2.5 基于用户和基于物品推荐的 对比 .....	81	6.2 推荐知识库的开发 .....	133
4.3 近邻方法的要素 .....	83	6.3 推荐过程中的用户导向 .....	137
4.3.1 评分标准化 .....	83	6.4 计算推荐结果 .....	142
4.3.2 相似度权重计算 .....	85	6.5 项目和案例研究的经验 .....	143
4.3.3 近邻的选择 .....	89	6.6 未来的研究方法 .....	144
4.4 高级进阶技术 .....	90	6.7 总结 .....	147
4.4.1 降维方法 .....	90	参考文献 .....	147
4.4.2 基于图方法 .....	92	<b>第7章 情境感知推荐系统</b> .....	151
4.5 总结 .....	95	7.1 简介 .....	151
参考文献 .....	96	7.2 推荐系统中的情境 .....	152
<b>第5章 协同过滤算法的高级 课题</b> .....	100	7.2.1 什么是情境 .....	152
5.1 简介 .....	100	7.2.2 在推荐系统实现情境信息的 建模 .....	155
5.2 预备知识 .....	101	7.2.3 获取情境信息 .....	158
5.2.1 基准预测 .....	102	7.3 结合情境的推荐系统形式 .....	159
5.2.2 Netflix 数据 .....	103	7.3.1 情境预过滤 .....	161
5.2.3 隐式反馈 .....	103	7.3.2 情境后过滤 .....	163
5.3 因子分解模型 .....	104	7.3.3 情境建模 .....	164
5.3.1 SVD .....	104	7.4 多种方法结合 .....	167
5.3.2 SVD++ .....	105	7.4.1 组合预过滤器案例研究: 算法 .....	168
5.3.3 时间敏感的因子模型 .....	106	7.4.2 组合预过滤器案例研究: 实验结果 .....	168
5.3.4 比较 .....	111	7.5 情境感知推荐系统的其他 问题 .....	170
5.3.5 总结 .....	112	7.6 总结 .....	171
5.4 基于邻域的模型 .....	112	致谢 .....	171
5.4.1 相似度度量 .....	113	参考文献 .....	172
5.4.2 基于相似度的插值 .....	113	<b>第二部分 推荐系统的应用与评估</b>	
5.4.3 联合派生插值权重 .....	115	<b>第8章 推荐系统评估</b> .....	176
5.4.4 总结 .....	117	8.1 简介 .....	176
5.5 增强的基于邻域的模型 .....	117		
5.5.1 全局化的邻域模型 .....	118		
5.5.2 因式分解的邻域模型 .....	122		
5.5.3 基于邻域的模型的动态 时序 .....	126		
5.5.4 总结 .....	127		

8.2 实验设置 .....	177	9.6.1 离线分析 .....	218
8.2.1 离线实验 .....	178	9.6.2 在线分析 .....	220
8.2.2 用户调查 .....	180	9.7 总结 .....	223
8.2.3 在线评估 .....	182	参考文献 .....	223
8.2.4 得出可靠结论 .....	182		
8.3 推荐系统属性 .....	185	<b>第 10 章 走出实验室的推荐</b>	
8.3.1 用户偏好 .....	185	<b>系统</b> .....	225
8.3.2 预测准确度 .....	186	10.1 简介 .....	225
8.3.3 覆盖率 .....	191	10.2 设计现实环境中的推荐	
8.3.4 置信度 .....	192	系统 .....	225
8.3.5 信任度 .....	193	10.3 理解推荐系统的环境 .....	226
8.3.6 新颖度 .....	194	10.3.1 应用模型 .....	226
8.3.7 惊喜度 .....	195	10.3.2 用户建模 .....	230
8.3.8 多样性 .....	195	10.3.3 数据模型 .....	233
8.3.9 效用 .....	196	10.3.4 一个使用环境模型的	
8.3.10 风险 .....	197	方法 .....	235
8.3.11 健壮性 .....	197	10.4 在迭代设计过程中理解推荐	
8.3.12 隐私 .....	198	验证步骤 .....	236
8.3.13 适应性 .....	198	10.4.1 算法的验证 .....	236
8.3.14 可扩展性 .....	199	10.4.2 推荐结果的验证 .....	237
8.4 总结 .....	199	10.5 应用实例：一个语义新闻推荐	
参考文献 .....	199	系统 .....	240
		10.5.1 背景：MESH 工程 .....	240
<b>第 9 章 IPTV 服务提供商推荐系统：</b>		10.5.2 MESH 的环境模型 .....	240
<b>一个大规模真实产品环境的</b>		10.5.3 实践：模型的迭代	
<b>应用</b> .....	203	实例化 .....	243
9.1 简介 .....	203	10.6 总结 .....	244
9.2 IPTV 架构 .....	204	参考文献 .....	244
9.3 推荐系统架构 .....	206		
9.3.1 数据搜集 .....	206	<b>第 11 章 匹配推荐系统的技术与</b>	
9.3.2 批处理和实时阶段 .....	207	<b>领域</b> .....	247
9.4 推荐算法 .....	208	11.1 简介 .....	247
9.4.1 推荐算法概述 .....	209	11.2 相关工作 .....	247
9.4.2 基于内容隐语义分析		11.3 知识源 .....	248
算法 .....	210	11.4 领域 .....	250
9.4.3 基于物品的协同过滤		11.4.1 异构性 .....	250
算法 .....	213	11.4.2 风险性 .....	251
9.4.4 基于降维的协同过滤		11.4.3 变动性 .....	251
算法 .....	214	11.4.4 交互风格 .....	251
9.5 推荐服务 .....	215	11.4.5 偏好稳定性 .....	251
9.6 系统评价 .....	216	11.4.6 可理解性 .....	252

11.5 知识源 .....	252	13.4 评价平台中的交互研究 .....	293
11.5.1 社群知识 .....	252	13.4.1 扩展到其他评价平台 .....	294
11.5.2 个人知识 .....	253	13.4.2 用户直接操作与限制用户 控制的比较 .....	295
11.5.3 基于内容的知识 .....	253	13.4.3 支持性解释、置信和 信任 .....	296
11.6 从领域到技术 .....	254	13.4.4 可视化、自适应性和分区 动态性 .....	297
11.6.1 算法 .....	255	13.4.5 关于多文化的适用性的 差异 .....	298
11.6.2 抽样推荐领域 .....	256	13.5 评价的评估: 资源、方法和 标准 .....	298
11.7 总结 .....	257	13.5.1 资源和方法 .....	298
致谢 .....	257	13.5.2 评估标准 .....	299
参考文献 .....	257	13.6 总结与展望 .....	300
<b>第 12 章 用于技术强化学习的   推荐系统</b> .....	<b>261</b>	参考文献 .....	<b>301</b>
12.1 简介 .....	261	<b>第 14 章 构建更值得信任和具有   说服力的推荐系统:   特性对评估推荐系统   的影响</b> .....	<b>305</b>
12.2 背景 .....	262	14.1 简介 .....	305
12.2.1 TEL 作为上下文 .....	262	14.2 推荐系统作为社交角色 .....	306
12.2.2 TEL 推荐的目标 .....	263	14.3 来源可信度 .....	306
12.3 相关工作 .....	264	14.3.1 可信度 .....	306
12.3.1 自适应教育超媒体 .....	264	14.3.2 专业能力 .....	307
12.3.2 学习网络 .....	265	14.3.3 对来源可信度的影响 .....	307
12.3.3 相同点与不同点 .....	267	14.4 人际交互中信息特性的 研究 .....	307
12.4 TEL 推荐系统调查 .....	268	14.4.1 相似度 .....	307
12.5 TEL 推荐系统的评估 .....	271	14.4.2 喜好度 .....	308
12.5.1 对组件的评估 .....	272	14.4.3 权威的象征 .....	308
12.5.2 评估 TEL 推荐系统时需要 考虑的问题 .....	273	14.4.4 演讲的风格 .....	308
12.6 总结与展望 .....	274	14.4.5 外在吸引力 .....	308
致谢 .....	274	14.4.6 幽默 .....	309
参考文献 .....	275	14.5 人机交互中的特性 .....	309
<b>第三部分 推荐系统的影响</b>		14.6 用户与推荐系统交互的 特性 .....	309
<b>第 13 章 基于评价推荐系统的   进展</b> .....	<b>282</b>	14.6.1 推荐系统类型 .....	310
13.1 简介 .....	282	14.6.2 输入特性 .....	310
13.2 早期: 评价系统/已得益处 .....	282	14.6.3 过程特性 .....	311
13.3 评价系统的表述与检索 挑战 .....	283		
13.3.1 评价表述的方式 .....	283		
13.3.2 基于评价的推荐系统中的 检索挑战 .....	289		

14.6.4 输出特性 .....	311	15.8 总结与展望 .....	336
14.6.5 内嵌的智能体特性 .....	312	参考文献 .....	337
14.7 讨论 .....	312	<b>第 16 章 基于实例评价研究的产品</b>	
14.8 影响 .....	313	<b>推荐系统的可用性</b>	
14.9 未来研究方向 .....	314	<b>准则 .....</b>	<b>340</b>
参考文献 .....	314	16.1 简介 .....	340
<b>第 15 章 设计和评估推荐系统的</b>		16.2 预备知识 .....	341
<b>解释 .....</b>	<b>321</b>	16.2.1 交互模型 .....	341
15.1 简介 .....	321	16.2.2 基于效用的推荐系统 .....	342
15.2 指引 .....	322	16.2.3 准确率、信任度和代价的	
15.3 专家系统的说明 .....	322	框架 .....	344
15.4 定义的目标 .....	322	16.2.4 本章结构 .....	344
15.4.1 系统如何工作：透明性 .....	324	16.3 相关工作 .....	345
15.4.2 允许用户告诉系统它是		16.3.1 推荐系统分类 .....	345
错误的：被理解 .....	324	16.3.2 基于评分的推荐系统 .....	345
15.4.3 增加用户对系统上的		16.3.3 基于案例的推荐系统 .....	345
信任：信任度 .....	325	16.3.4 基于效用的推荐系统 .....	345
15.4.4 说服用户尝试或购买：		16.3.5 基于评价的推荐系统 .....	346
说服力 .....	326	16.3.6 其他设计指导准则 .....	346
15.4.5 帮助用户充分地决策：		16.4 初始偏好提取 .....	347
有效性 .....	327	16.5 通过实例激励用户表示	
15.4.6 帮助用户快速制定决策：		偏好 .....	349
效率 .....	328	16.5.1 需要多少实例 .....	350
15.4.7 使系统的应用愉悦：		16.5.2 需要哪些实例 .....	350
满意度 .....	328	16.6 偏好修正 .....	352
15.5 评估解释在推荐系统的		16.6.1 偏好冲突和部分满足 .....	352
作用 .....	329	16.6.2 权衡辅助 .....	353
15.5.1 精准度 .....	329	16.7 展示策略 .....	354
15.5.2 学习效率 .....	329	16.7.1 一次推荐一项物品 .....	354
15.5.3 覆盖度 .....	330	16.7.2 推荐 $k$ 项最匹配的物品 .....	355
15.5.4 接受度 .....	330	16.7.3 解释界面 .....	355
15.6 用推荐设计展示与互动 .....	330	16.8 准则验证模型 .....	357
15.6.1 展示推荐 .....	330	16.9 总结 .....	359
15.6.2 与推荐系统交互 .....	331	参考文献 .....	359
15.7 解释风格 .....	332	<b>第 17 章 基于示意图的产品目录</b>	
15.7.1 基于协同风格 .....	333	<b>可视化 .....</b>	<b>363</b>
15.7.2 基于内容风格 .....	334	17.1 简介 .....	363
15.7.3 基于案例风格 .....	334	17.2 基于图的可视化方法 .....	364
15.7.4 基于知识/自然语言风格 .....	335	17.2.1 自组织映射 .....	364
15.7.5 基于人口统计风格 .....	335	17.2.2 树图 .....	365

17.2.3	多维缩放	366
17.2.4	非线性主成分分析	367
17.3	产品目录图	367
17.3.1	多维缩放	368
17.3.2	非线性主成分分析	369
17.4	通过点击流分析决定属性权重	370
17.4.1	泊松回归模型	370
17.4.2	处理缺失值	371
17.4.3	使用泊松回归选择权值	371
17.4.4	阶梯式泊松回归模型	371
17.5	图像购物界面	372
17.6	电子商务应用	373
17.6.1	使用属性权值的基于MDS的产品目录图	373
17.6.2	基于NL-PCA的产品目录图	375
17.6.3	图像购物界面	377
17.7	总结与展望	379
	致谢	380
	参考文献	380

## 第四部分 推荐系统与群体

### 第18章 个性化Web搜索中的群体、协作与推荐系统

18.1	简介	384
18.2	网络搜索历史简介	385
18.3	网络搜索的未来	387
18.3.1	个性化网络搜索	387
18.3.2	协同信息检索	390
18.3.3	向社交搜索前进	392
18.4	案例研究1: 基于群体的网络搜索	392
18.4.1	搜索群体中的重复性和规律性	392
18.4.2	协同网络搜索系统	393
18.4.3	评估	395
18.4.4	讨论	396
18.5	案例研究2: 网络搜索共享	396

18.5.1	HeyStaks系统	397
18.5.2	HeyStaks推荐引擎	399
18.5.3	评估	400
18.5.4	讨论	402
18.6	总结	402
	致谢	403
	参考文献	403

### 第19章 社会化标签推荐系统

19.1	简介	409
19.2	社会化标签推荐系统	410
19.2.1	大众分类法	410
19.2.2	传统推荐系统范式	411
19.2.3	多模式推荐	412
19.3	现实社会化标签推荐系统	413
19.3.1	有哪些挑战	413
19.3.2	案例 BibSonomy	413
19.3.3	标签获取	415
19.4	社会化标签系统的推荐算法	416
19.4.1	协同过滤	416
19.4.2	基于排序的推荐	418
19.4.3	基于内容的社会化标签推荐系统	421
19.4.4	评估方案和评估度量	423
19.5	算法比较	424
19.6	总结与展望	426
	参考文献	427

### 第20章 信任和推荐

20.1	简介	430
20.2	信任的表示与计算	431
20.2.1	信任表示	431
20.2.2	信任计算	433
20.3	信任增强推荐系统	436
20.3.1	动机	436
20.3.2	进展	437
20.3.3	实验比较	441
20.4	进展和开放性挑战	445
20.5	总结	446
	参考文献	446

<b>第 21 章 组推荐系统</b> .....	449	22.2.3 CB 与 UB 的配置文件 构建 .....	470
21.1 简介 .....	449	22.2.4 物品和用户相似度以及 邻居的形成 .....	471
21.2 应用场景和群组推荐系统 分类 .....	450	22.2.5 基于实例推理的连接词 在推荐系统中的应用 .....	472
21.2.1 交互式电视 .....	450	22.2.6 加权混合系统 .....	472
21.2.2 环绕智能 .....	450	22.3 聚合函数概论 .....	472
21.2.3 基于场景的推荐系统 .....	451	22.3.1 定义和属性 .....	472
21.2.4 基于分类的群组推荐 .....	451	22.3.2 聚合成员 .....	475
21.3 合并策略 .....	452	22.4 聚合函数的构建 .....	479
21.3.1 合并策略概览 .....	452	22.4.1 数据收集和处理 .....	479
21.3.2 合并策略在相关工作中的 应用 .....	453	22.4.2 期望属性、语义、解释 .....	480
21.3.3 哪种策略效果最好 .....	454	22.4.3 函数表现的复杂度及其 理解 .....	481
21.4 序列顺序的影响 .....	455	22.4.4 权重和参数的确定 .....	482
21.5 对情感状态建模 .....	456	22.5 推荐系统中的复杂聚合过程： 为特定应用定制 .....	482
21.5.1 对个人的满意度进行 建模 .....	457	22.6 总结 .....	485
21.5.2 个人满意度对群组的 影响 .....	458	22.7 进阶阅读 .....	485
21.6 情感状态在合并策略中的 使用 .....	459	致谢 .....	486
21.7 对单个用户进行组推荐 .....	460	参考文献 .....	486
21.7.1 多准则 .....	460	<b>第 23 章 推荐系统中的主动 学习</b> .....	488
21.7.2 冷启动问题 .....	461	23.1 简介 .....	488
21.7.3 虚拟组成员 .....	462	23.1.1 推荐系统中主动学习的 目标 .....	489
21.8 总结与挑战 .....	462	23.1.2 例证 .....	490
21.8.1 提出的主要问题 .....	463	23.1.3 主动学习的类型 .....	490
21.8.2 警告：组建模型 .....	463	23.2 数据集的属性 .....	491
21.8.3 面临的挑战 .....	464	23.3 主动学习在推荐系统中的 应用 .....	492
致谢 .....	464	23.4 主动学习公式 .....	493
参考文献 .....	465	23.5 基于不确定性的主动学习 .....	495
		23.5.1 输出不确定性 .....	495
		23.5.2 决策边界不确定性 .....	496
		23.5.3 模型不确定性 .....	497
		23.6 基于误差的主动学习 .....	498
		23.6.1 基于实例的方法 .....	498
		23.6.2 基于模型的方法 .....	500
<b>第五部分 高级算法</b>			
<b>第 22 章 推荐系统中的偏好 聚合</b> .....	468		
22.1 简介 .....	468		
22.2 推荐系统中的聚合类型 .....	468		
22.2.1 协同过滤中的偏好聚合 .....	470		
22.2.2 CB 与 UB 推荐中的 特性聚合 .....	470		

23.7 基于组合的主动学习 .....	501	第 25 章 具有健壮性的协同推荐 ...	533
23.7.1 基于模型的方法 .....	501	25.1 简介 .....	533
23.7.2 基于候选的方法 .....	502	25.2 问题定义 .....	534
23.8 基于会话的主动学习 .....	504	25.3 攻击分类 .....	536
23.8.1 基于实例的评论 .....	504	25.3.1 基础攻击 .....	536
23.8.2 基于多样性的方法 .....	504	25.3.2 非充分信息攻击 .....	537
23.8.3 基于查询编辑的方法 .....	505	25.3.3 打压攻击模型 .....	537
23.9 计算因素考虑 .....	505	25.3.4 知情攻击模型 .....	538
23.10 总结 .....	505	25.4 检测系统健壮性 .....	539
致谢 .....	506	25.4.1 评估矩阵 .....	539
参考文献 .....	506	25.4.2 推举攻击 .....	540
<b>第 24 章 多准则推荐系统 .....</b>	<b>510</b>	25.4.3 打压攻击 .....	541
24.1 简介 .....	510	25.4.4 知情攻击 .....	542
24.2 推荐作为多准则决策问题 .....	511	25.4.5 攻击效果 .....	543
24.2.1 决策目标 .....	512	25.5 攻击检测 .....	543
24.2.2 准则簇 .....	512	25.5.1 评估矩阵 .....	544
24.2.3 全局偏好模型 .....	513	25.5.2 单用户检测 .....	544
24.2.4 决策支持流程 .....	513	25.5.3 用户组检测 .....	545
24.3 推荐系统的 MCDM 框架:		25.5.4 检测结果 .....	548
经验教训 .....	515	25.6 健壮的推荐算法 .....	548
24.4 多准则评分推荐 .....	517	25.6.1 基于模型的推荐 .....	548
24.4.1 传统的单值评分推荐		25.6.2 健壮的矩阵分解算法 .....	549
问题 .....	517	25.6.3 其他具有健壮性的推荐	
24.4.2 引入多准则评分来扩展		算法 .....	549
传统推荐系统 .....	518	25.6.4 影响力限制器和基于信誉	
24.5 多准则评分推荐算法综述 .....	519	的推荐 .....	550
24.5.1 预测中使用多准则评分 ...	519	25.7 总结 .....	550
24.5.2 推荐中使用多准则评分 ...	524	致谢 .....	551
24.6 讨论及未来工作 .....	526	参考文献 .....	551
24.7 总结 .....	527	<b>本书贡献者名单 .....</b>	<b>554</b>
致谢 .....	528	<b>翻译团队名单 .....</b>	<b>560</b>
参考文献 .....	528		



## 概 述

Francesco Ricci、Lior Rokach 和 Bracha Shapira

**摘要** 推荐系统(Recommender System, RS)是向用户建议有用物品的软件工具和技术。在本章中,我们会简要地介绍推荐系统的基本思想和概念。主要目标是用连贯和结构化的方式描述这本手册的章节内容,以此来帮助读者理解这本手册提供的极其丰富和详细的内容。

## 1.1 简介

推荐系统(RS)是一种软件工具和技术方法,它可以向用户建议有用的物品[60, 85, 25],这种建议适用于多种决策过程,如购买什么物品、听什么音乐、在网上浏览什么新闻等。

“物品”是用来表示系统向用户推荐内容的总称。一个推荐系统通常专注于一个特定类型的物品(如 CD 或新闻),因此它的设计、图形用户界面以及用于生成建议的核心的推荐技术都是为特定类型的物品提供有用和有效的建议而定制的。

推荐系统主要针对的是那些缺乏足够的个人经验和能力的人,他们无法评估潜在的大量可供选择的物品,比如,某个网站提供的商品[85]。一个典型的例子是图书推荐系统,它帮助用户挑选一本书来读。在亚马逊,网站采用个性化推荐系统为每个客户进行推荐[47]。由于推荐通常是个性化的,不同的用户或用户组接收的建议是不同的。当然也存在非个性化推荐。它们都是非常简单的,通常出现在报纸或杂志上。典型的例子包括书籍和 CD 等的 top 10 推荐(最热销的前 10 名)。虽然在某些情况下它们可能是有用和有效的,但这些类型的非个性化推荐通常不是推荐系统研究要解决的问题。

个性化推荐最简单的形式是提供一个排好序的物品列表。通过这个排序列表,推荐系统试图根据用户的偏好和其他约束条件来预测最合适的产品或服务。为了完成这样的计算任务,推荐系统收集用户的喜好,这种喜好是显式的,如为产品打分,或通过解释用户的行为做出推断。例如,推荐系统可能会把访问某个特定商品详情页的行为作为该用户喜爱这个主页上的商品的隐式信号。

推荐系统的发展源于一个很简单的现象:人们在做日常工作和日常决策时总是依赖于其他人提供的建议[60, 70]。例如,要选择一本书的时候,通常依靠朋友的推荐;雇主依靠推荐信做招聘的决定;当选择观看的影片时,人们倾向于阅读并且依赖影评家写在报纸

---

Francesco Ricci, Faculty of Computer Science, Free University of Bozen-Bolzano, Italy e-mail: fricci@unibz.it

Lior Rokach, Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel  
e-mail: liorrk@bgu.ac.il

Bracha Shapira, Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel  
e-mail: bshapira@bgu.ac.il

翻译:燕山大学-王孝先 审核:王二朋,承皓,严强