

第 1 章 引 言

随着智能互联网时代的到来和发展,个性化推荐作为理解用户的核心技术之一,成为智能网络的重要组成部分,并在各种实际系统中得到广泛应用。长期以来,个性化推荐技术的研究集中于如何为用户提供更为准确的被推荐物品,而在很大程度上忽视了推荐系统的可解释性,不利于推荐系统对用户的透明度和可信度。本书从数据、模型和结果三个方面对推荐系统的可解释性进行研究,力图做到不仅知其然更知其所以然,并在互联网真实用户场景下对理论模型进行实验验证。本章旨在阐述研究背景,简要回顾个性化推荐系统的主要技术与历史现状,给出本书的研究问题、面临的主要挑战及其实际意义和科学价值,并描述本研究的主要贡献和章节安排。

1.1 研究背景

互联网的快速发展开启了人类活动线上化的进程,越来越多传统上只能在线下完成的任务变得可以方便快捷地在互联网上完成。已经深入人们日常生活的电子商务就是这一进程的典型代表,例如阿里巴巴^①、京东商城^②、亚马逊网络商城^③等电子商务网站的普及,使得人们不必走出家门即可购买自己所需要的商品,并且可以在更多的备选商品中进行挑选。不仅限于电子商务应用,社交网络平台(如新浪微博^④和 Facebook^⑤)的兴起使得人

① 阿里巴巴,中国最大的电子商务企业平台, <http://www.alibaba.com>。

② 京东商城,中国最大的自主经营式电子商务平台, <http://www.jd.com>。

③ 亚马逊网络商城,美国著名电子商务网站,在世界主要国家提供服务, <http://www.amazon.com>。

④ 新浪微博,中国主要的微博客社交网络平台, <http://www.weibo.com>。

⑤ Facebook,全球最大的社交网络平台, <http://www.facebook.com>。

们可以在互联网上交友、沟通、获取实时资讯；在线叫车服务（如滴滴^①和 Uber^②）的发展使得用户不再需要线下街头打车；在线 P2P 借贷服务（如宜信^③和 Prosper^④）使得用户线上借贷和理财成为可能；在线房地产业务（如 Zillow^⑤和 Airbnb^⑥）的发展使传统的房地产业务逐步线上化；在线自由职业平台（如猪八戒网^⑦和亚马逊 MTurk^⑧）的迅速发展甚至使得自由职业者在线工作和任务分配成为可能。

伴随着各种互联网应用的迅速发展，个性化推荐系统成为网络应用中不可缺少的重要组成部分，并在各种场景下以不同的方式影响着人们网络生活的方方面面，其研究也对国民生产、生活的多方面具有重大意义。

第一，个性化推荐技术的研究对互联网在线服务和信息系统具有重要的经济和市场价值。随着人类线下活动的逐步线上化，互联网不再仅仅是一个信息流通和传播的平台，而是逐渐成为了一个完整的在线经济和社会系统，大量的社会生产、生活和消费商贸活动以在线交易的方式在互联网上完成，而个性化推荐系统及其相关技术在这一过程中发挥着资源配置的核心作用。例如在电子商务网站中，个性化推荐将商家待销售的商品与具有相应需求的用户进行匹配，从而提高整个在线经济系统资源配置的效率，进而促进消费，推动国民经济的发展。据京东商城推荐搜索部透露^⑨，京东商城基于大数据的个性化推荐算法在 PC 端和移动端都已经为京东贡献了 10% 以上的订单；而据著名的科技咨询公司 VentureBeat 统计^⑩，亚马逊的个性化推荐系统更是为其贡献了 35% 以上的销售额，推荐系统对在线经济的重要作用可见一斑。

不仅是在电子商务系统中，在线租房和房地产业务中的房屋推荐通过综合考虑地理位置和价格等信息，对用户需求进行精确定位，从而提高住房利用率和降低房产空置率；在线自由职业平台中的工作任务推荐系统更是综合考虑自由职业者的技术能力和预期报酬以及雇主的任务需求，通过精确的职业

① 滴滴，中国最大的互联网在线打车服务平台，<http://www.xiaojukeji.com>。

② Uber，美国在线打车服务公司，在世界许多国家提供服务，<http://www.uber.com>。

③ 宜信，中国最早开始网络贷款服务的公司，<http://www.creditease.cn>。

④ Prosper，美国在线 P2P 借贷服务平台，<http://www.prosper.com>。

⑤ Zillow，美国主要的在线房地产买卖业务平台，<http://www.zillow.com>。

⑥ Airbnb，全球主要的 P2P 租房服务平台，<http://www.airbnb.com>。

⑦ 猪八戒网，中国最大的在线自由职业平台，<http://www.zbj.com>。

⑧ MTurk，亚马逊旗下自由职业平台，<http://www.mturk.com>。

⑨ 订单贡献率 10%，京东个性化推荐系统持续优化的奥秘，CSDN 资讯。

⑩ Aggregate knowledge raises \$5M from Kleiner, on a roll, VentureBeat.

匹配令雇佣双方各得其所,使高效的在线劳务市场成为可能。除了显式的用户可见的个性化推荐系统之外,隐式的推荐系统也大量存在于网络之中,例如在线叫车服务系统对用户需求和付费意愿进行预测并对道路交通状况实施建模,从而为用户自动匹配最合适的司机,在满足双方出行需求的同时进行路线优化、缓解交通负担。

第二,个性化推荐系统的研究对国民生产和国家安全具有重大意义。在个性化推荐系统的信息匹配和资源配置过程中,恰当地融合经济效益及风险控制等因素的考量对互联网经济安全和网络环境的稳定可控具有重要作用。例如在网络借贷服务的理财产品推荐中,对理财产品的风险评估和用户风险承受能力的评估是产品匹配以及面向用户的推荐过程中所要考虑的重要因素,向不同用户推荐和展示合理的产品是网络金融服务实现风险可控的重要手段。再如社交媒体的兴起和快速发展使得人们可以更加迅速快捷地发布、分享和传播信息,打破了长期以来新闻讯息由国家和新闻媒体机构垄断的局面。自媒体的兴起使信息传播更为高效、信息获取成本大大降低、社会生产和生活更富有活力,但同时也为网络谣言、信息诈骗、极端思想、恐怖主义的传播带来了便利,而社交网络中的个性化信息排序和推荐技术则在满足用户个性化信息需求的同时,起到积极引导社会舆论的作用。

个性化推荐技术的研究不仅具有重要的实际应用意义,更具有重要的科学研究价值,如图 1.1 所示。

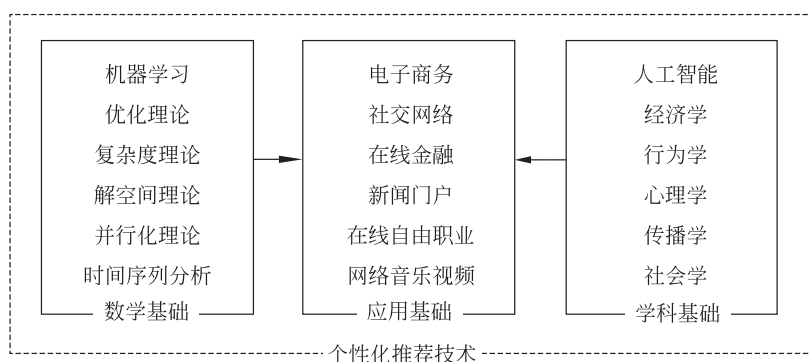


图 1.1 个性化推荐研究涉及的学科领域

具体而言,其科学研究价值和意义主要包括如下三个方面。

第一,个性化推荐技术的研究涉及多个重要的数学分支,有利于促进和

拓展相关理论研究的深度和广度。充分理解用户行为模式和个性化的信息需求需要对用户行为和偏好进行深入的数学建模,而互联网用户行为信息规模庞大且多种多样,例如电子商务网站中的用户浏览、购买、数值评分、文本评论等历史信息,在线视频音乐网站中的用户点击、观看、收听、时长等行为信息,以及社交网站中的用户好友关系、关注关系、地理位置、登录时间等社交信息等。数据的多样性和异质性为用户行为分析和偏好建模带来了新的难度,而庞大的数据规模也为网络大数据的处理带来了极大的挑战,这些都对机器学习方法和相关数学模型理论的发展与应用提出了新的要求,为矩阵运算、并行化理论、解空间理论、时空信息处理、系统复杂度控制等相关理论的研究发展提供了重要的问题背景。在本书中,将在矩阵分解、局部优化理论、解空间分析、并行化算法、时间序列分析等多个方面对理论前沿做出进一步的拓展。

第二,个性化推荐技术的研究涉及诸多学科的交叉综合,有利于促进跨学科学术研究的进一步发展。个性化推荐技术的核心在于用户需求理解,只有对用户兴趣和需求进行精确的建模,才能给出具有针对性的个性化推荐,因此,个性化推荐技术的研究需要互联网用户行为学和心理学的支持;在社交网站、新闻门户等应用场景中,对好友、新闻、信息的推荐则依赖于对信息传播学和社会学的深入理解和应用;同时,电子商务、在线金融、在线职业网站、在线打车等网络业务的发展正不断将人类的线下经济学活动线上化,对互联网经济现象和用户在线经济行为的深入理解和正确建模对提供合时合地合情合理的推荐具有重要作用,而这依赖于经济学相关理论的应用与发展。本研究将借助经济学、心理学、行为学等学科的基本概念和主要结论,对互联网用户行为进行分析建模,并进一步给出个性化的推荐。

第三,个性化推荐技术的研究涉及众多互联网应用场景,有利于促进互联网整体的进一步个性化和智能化。个性化推荐以其技术的基础性和方法的通用性已经成为诸多网络应用中不可或缺的组成部分,以显性或隐性的方式渗入到人们网络生活的方方面面。个性化推荐技术以其“理解用户”的核心思想,成为向用户提供智能服务的基础和关键的第一步,因此在未来以智能化为核心特征的下一代互联网的发展中具有重要的基础性意义,同时也是未来个性化生活和办公助理等平台化系统,以及智能家居等线下智能系统的核心技术之一。本书不仅限于理论研究和技術拓展,更进一步将相关理论应用到

电子商务、在线娱乐、网络金融、在线自由职业等多个不同的各有特点的网络应用场景中，从而验证相关理论的实用性和有效性。

1.2 问题的提出

个性化推荐的研究及其实际应用包含三个依次递进的核心关切，分别是输入数据、算法模型和线上应用的经济效益，如图 1.2 所示。一个典型的个性化推荐系统将用户在物品上的浏览、点击、购买、评论等行为信息作为输入数据，通过构建个性化推荐模型对用户进行偏好建模并给出个性化推荐列表，进而通过用户在推荐列表上的购买和消费行为产生经济效益。

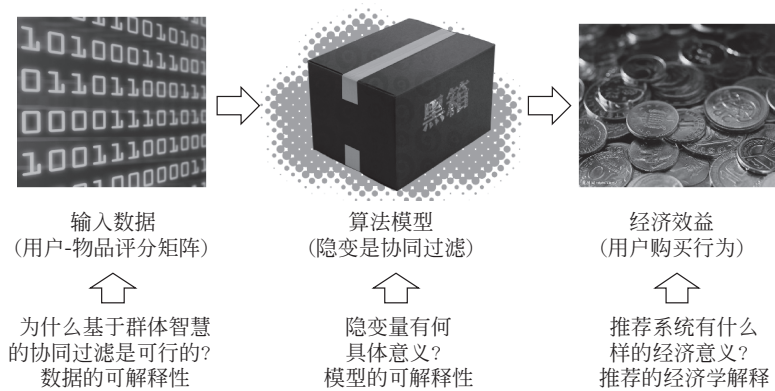


图 1.2 典型个性化推荐系统的主要构件及本研究需要解决的核心问题

输入数据是个性化推荐系统的基础，用户-物品评分稀疏矩阵是个性化推荐实际系统中最基本和最常用的输入数据形式。在用户-物品评分矩阵中，每一行代表一个用户、每一列代表系统中的一个商品，而矩阵中的值表示相应用户在相应物品上的打分，例如电子商务网站中典型的 1~5 星数值评分等。由于系统中的商品往往有数百万甚至数亿个，而每一个用户只购买过少数的商品，因此矩阵中有大量的未知打分，因而打分预测成为个性化推荐的核心问题之一。近年来，矩阵分解等协同过滤技术在打分预测问题上取得了较好的效果并得到广泛应用，然而为何在稀疏矩阵上基于群体智慧的协同过滤得以可行这一根本性的问题并未得到很好的回答，这涉及数据层面上的可解释性问题。本书提出稀疏矩阵的迭代双边块对角结构，指出和证明了稀疏矩阵

内在的用户-物品群组关系，并证明了矩阵分解协同过滤对群组的可拆分性这一重要的数学性质，从而为协同过滤的可行性找到了理论依据。同时，将杂乱无章的原始矩阵转化为内在的用户-物品社区结构，一方面降低了数据稀疏性，另一方面将具有相似兴趣的用户及其历史评分聚在一起，从而提高协同过滤算法的预测效果。

在数据的基础上，个性化推荐模型试图对用户的个性化偏好进行建模，从而给出个性化推荐结果。基于隐变量的协同过滤（例如矩阵分解）技术是目前应用最为广泛的推荐模型之一，然而隐变量模型将数据投影到一组未知的空间变量上，因此难以解释原有数据中用户的具体偏好，也难以据此为推荐结果给出直观的推荐理由，例如大多数电子商务推荐系统只是简单地给出“其他用户也购买了”等与真正模型无关的推荐理由。在本书中，提出与传统隐变量分解模型相对应的显式变量分解模型来解决模型的可解释性问题，通过引入属性词显式变量构建用户在属性词上的偏好矩阵，从而使得协同过滤算法具有明确的内在意义，并且可以为最终的推荐结果给出模型内生的推荐理由。

在数据和模型的基础上，个性化推荐算法通过用户在推荐列表上的点击、购买等行为为系统创造经济效益，因此个性化推荐实际上以推荐列表的方式隐形地控制着系统中物品与用户之间的匹配，从而影响推荐算法所在经济系统（如电子商务网站）中的资源分配。虽然电子商务、在线金融等网络应用已经是一个完整的线上经济系统，然而传统的个性化推荐研究主要从计算机科学家的视角出发，关心点击率（click through rate, CTR）、购买率等性能指标，而对个性化推荐在整个系统中的经济学意义较少研究。本书将个性化推荐作为经济系统中资源分配的手段，提出了网络经济系统的生产者-消费者建模框架，并给出了系统最终所实现社会效益（系统总福利）的计算方法，从而对个性化推荐的经济学意义进行解释。在此基础上，进一步提出基于总福利最大化的推荐算法，从而在提高推荐系统用户体验的同时，也提高整个经济系统的社会效益。

1.3 面临的主要挑战

个性化推荐技术的主要研究对象丰富多变，概括而言包括两大部分：一是广泛存在于各种互联网应用中的被推荐物品，包括商品、视频、音乐、电

影、新闻、金融产品、工作任务等；二是购买、消费和操作这些物品的网络用户。用户与物品之间交互方式的多样性、行为记录的丰富性、兴趣偏好的动态性为个性化推荐技术的研究及其解释带来了诸多挑战，如图 1.3 所示，这主要包括如下几个方面。

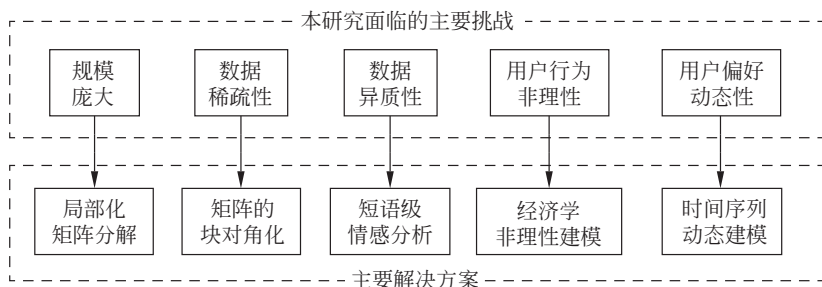


图 1.3 本研究面临的主要挑战及其对应的解决方案

第一，庞大的数据规模。在典型的互联网应用中往往存在着数量极为庞大的用户和物品，例如据全球最大的用户评论网站 Yelp^① 统计报告指出，截至 2016 年初，其日活跃用户达 1.35 亿，并且拥有 9500 万用户评论历史记录；中国主要的电子商务网站淘宝网和京东商城的活跃用户同样数以亿计；而百度、谷歌、Facebook 等搜索和社交网站的活跃用户更是达到数以十亿计。庞大的数据规模对用户偏好建模和个性化推荐算法的可行性和实时性提出了较高的要求。本书提出了矩阵的迭代双边块对角矩阵 (bordered block diagonal form, BBDF)，并证明了该结构与用户-物品二部图社区发现的等价性，从而为用户-物品行为数据的内在结构提供了解释框架。我们进一步证明了双边块对角矩阵在分解问题上的数学性质，并基于其性质提出了局部化矩阵分解 (localized matrix factorization, LMF) 框架，使得矩阵分解在数据层面上的并行化和局部化成为可能；不仅如此，我们还证明了该框架与许多常见矩阵分解算法的兼容性，从而为矩阵分解问题提供了一个统一的并行化解决方案。

第二，用户行为数据极其稀疏。虽然网络应用中往往存在数以千万乃至亿计的用户和物品，然而单个用户却往往只与其中极少的一部分物品存在历史行为记录，这导致多数互联网用户行为日志存在严重的数据稀疏性。同样

^① Yelp, 全球最大的用户评论网站, <http://www.yelp.com>。

在 Yelp 中,有 49% 以上的用户仅仅只有一条历史评论,用户-物品行为矩阵的平均密度仅有 0.043%,稀疏度达到了 99.957%。用户历史记录的稀疏性为用户行为的建模带来了挑战,使得我们不得不从有限的历史行为中估计用户的兴趣和偏好,并给出符合其个性化信息需求的推荐。在本书中,利用矩阵的双边块对角结构来剔除其中不包含信息量的部分,在增强矩阵密度的同时提高用户偏好的预测精度。

第三,数据异质性。由于用户行为方式和种类的多样性,网络应用所积累的用户行为历史数据往往具有明显的异质性,从而为异质信息的融合与协同处理带来了挑战。例如在电子商务等诸多网站的评论系统中,用户一方面可以对购买物品给出数值化的评分,另一方面可以同时以文本的方式给出评论,从而更为具体地表达自己的态度和偏好。长期以来,基于协同过滤的个性化推荐算法,尤其是基于矩阵分解的隐变量方法的相关研究只关注数值化评分的使用,而忽视了文本评论中所包含的丰富的用户个性化信息。这一方面是由于个性化的文本评论的异质性带来处理难度,另一方面也由于文本处理技术的瓶颈而难以发挥其应有的作用。近年来,随着短语级情感分析技术的不断发展和成熟,从用户评论文本中抽取结构化信息成为可能,从而为数值和文本异质信息的处理带来新的思路。本书利用短语级情感分析技术从用户评论中进行产品属性和用户情感的抽取,并基于此提出显式变量分解模型(explicit factor model, EFM),一方面同时处理用户数值评分和文本评论的异质信息,另一方面对推荐模型和结果给出属性级的个性化推荐理由与解释。

第四,用户行为的非理性。用户历史行为数据虽然均为用户在实际环境下的真实行为记录,但并非全部都是用户在理性状态下的最优决策。经济学研究指出,用户在实际决策中具有一定的非理性,这主要是由于人并非完全能够做到对目标对象的价值和效用进行精确估计并做出严格最优的决策。例如在购物网站中,用户常常由于时空和搜索量的限制不能找到最优的商品而选择了次优的替代品;用户往往也无法精确计算出最优的购买数量而选择了一个估计的数量。非理性行为的存在为用户行为建模和模型估计带来了一定的偏差和挑战。本书引入经济学相关理论,对用户行为的非理性因素进行考虑和建模,从而消除由于用户非理性行为而带来的估计偏差。

第五,用户偏好的动态性。用户的偏好并非一成不变,而是随着时间的推

移发生兴趣的增强、减弱,或者转变。例如在化妆品购物领域,用户在夏季更加关注防晒相关的产品,而在秋、冬季节则更为关注保湿相关的产品,不断变化的行为偏好为用户兴趣的建模提出了动态化和实时性的要求。本书采用经济学的时间序列分析理论对用户动态变化的兴趣进行建模,同时为了解决大数据环境下时间序列模型参数过多带来的计算可行性问题,提出了基于傅里叶级数辅助的时间序列分析模型,为大数据环境下用户偏好的动态跟踪和预测提供了解决方案。

1.4 主要贡献

本书从数据、模型和经济学意义三个方面对个性化推荐的可解释性进行系统性的研究,主要贡献分为三点,总结如下。

第一,数据的可解释性,提出基于双边块对角矩阵的局部化矩阵分解框架,并将其应用于矩阵分解的并行化。用户-物品评分矩阵是个性化推荐算法,尤其是基于矩阵分解的个性化推荐算法的数据基础和主要输入。传统的矩阵分解算法将原始矩阵看作一个整体进行分解和预测,缺乏对矩阵内在结构的理解。本书提出矩阵的双边块对角结构,并在理论上证明该结构与二部图上社区发现算法的数学等价性,从而解释矩阵内在的社区结构和社区关系。在社区结构的基础上,进一步提出了局部化的矩阵分解框架,并理论证明了它与传统矩阵分解算法的兼容性,从而为常用的矩阵分解算法提供了一个统一的并行化框架,在提高预测精度的同时大幅提高计算效率。该部分内容将在第3章进行介绍。

第二,模型的可解释性,提出基于短语级情感分析的显式变量分解模型,及其基于时间序列分析的动态化建模。基于矩阵分解的隐变量模型由于其较好的评分预测效果和可扩展性,逐渐成为了个性化推荐的基础算法,并在实际系统中得到广泛的应用。然而由于变量在本质上的未知性,隐变量模型难以对推荐算法和推荐结果给出直观可理解的解释,进而降低了推荐系统对用户的可信度。本书利用短语级情感分析技术从大规模的用户评论中抽取产品属性词及用户在不同属性上表达的情感,进而引入显式变量,并提出基于显式变量分解模型的个性化推荐算法,一方面使得模型的优化过程具备了直观意义,另一方面给出在模型层面可解释的推荐结果和个性化的推荐理由。由

于用户在不同属性上的偏好具有时间周期性，利用时间序列分析对用户偏好进行动态建模和预测，从而实现动态时间意义上的可解释性推荐。基于浏览器的大规模真实用户实验显示，本书的方法在点击率、购买率、一致性等多个线上指标上效果显著。该部分内容将在第 4 章进行介绍。

第三，推荐的经济学解释，提出基于互联网系统总福利最大化的个性化推荐框架，并给出典型应用场景中的具体实现。随着人类传统线下活动的不断线上化，常见的互联网应用均可以形式化为“生产者-服务-消费者”模型，例如在电子商务网站中，网络商家（生产者）提供在线商品（服务），而网络用户（消费者）则在众多的商品中进行选择和购买。基于传统经济学的基本定义，本书首先给出了互联网环境下效用、成本和福利的基本概念与统一形式，并进一步给出了互联网应用中总社会福利的通用计算方法。在此基础上，以互联网服务分配为基本问题，提出基于网络福利最大化的个性化推荐框架。随后本书在典型的网络应用（例如电子商务、P2P 借贷、在线众包平台）中对该框架进行具体化，并进行个性化的网络服务推荐与评测。实验结果表明，该方法可以在为用户提供高质量服务推荐的同时提升社会总福利，即在提升用户体验的同时又增强了社会效益。该部分内容将在第 5 章进行介绍。