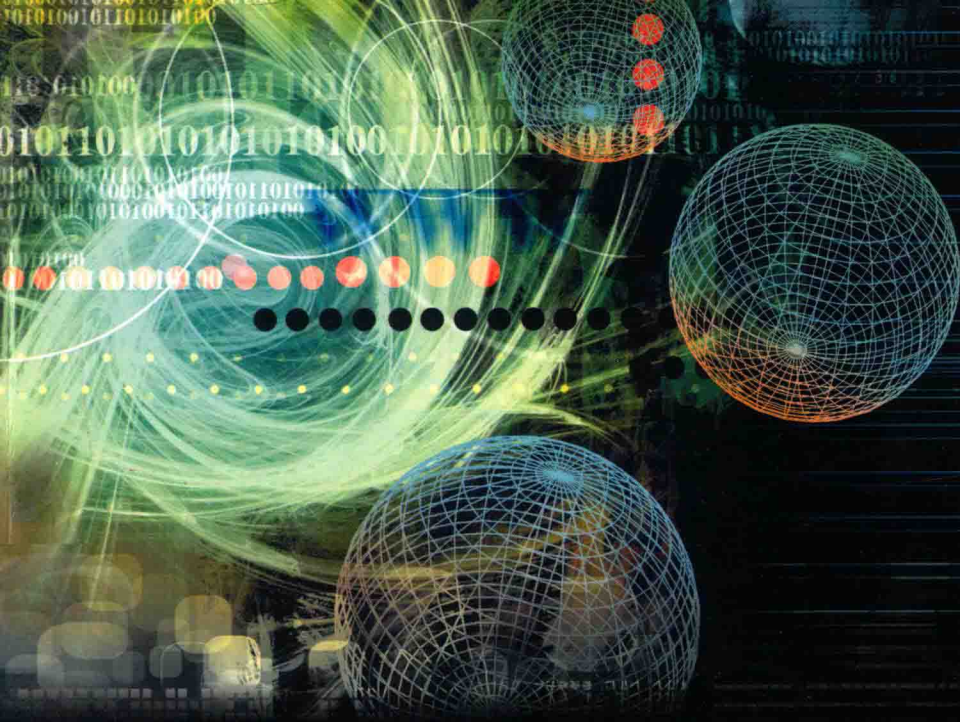




多Agent推荐系统及应用

◎ 吴 兵 编著



同济大学出版社
TONGJI UNIVERSITY PRESS

上海远程教育集团学术专著出版基金资助

多 Agent 推荐系统及应用

吴 兵 编著



同济大学出版社
TONGJI UNIVERSITY PRESS

内 容 提 要

本书主要将 Agent 技术应用于推荐系统,以改善现有的推荐方法,包括以下五个方面的内容。1. 推荐系统及应用中相关理论分析;2. 基于“效用”(Utility)理论的推荐方法研究;3. 基于 BDI Agent 的协作推荐模型研究;4. 基于遗传算法的粗糙集属性约简研究;5. 基于合同网的协作粒子群优化方法研究。

本书同时也探讨了粗糙集理论及智能优化方法在其中综合应用问题。

本书适用于高校计算机专业的本科生及硕士生参考阅读。

图书在版编目(CIP)数据

多 Agent 推荐系统及应用/吴兵编著. —上海:同济大学出版社,2014. 7

ISBN 978-7-5608-5545-5

I. ①多… II. ①吴… III. ①计算机网络—研究
IV. ①TP393

中国版本图书馆 CIP 数据核字(2014)第 125038 号

多 Agent 推荐系统及应用

编著 吴 兵

责任编辑 卞玉清 责任校对 张德胜 封面设计 潘向葵

出版发行 同济大学出版社 www.tongjipress.com.cn

(地址:上海市四平路 1239 号 邮编:200092 电话:021-65985622)

经 销 全国各地新华书店

印 刷 大丰市教育印刷有限公司

开 本 787mm×1092mm 1/16

印 张 8

字 数 199 000

版 次 2014 年 7 月第 1 版 2014 年 7 月第 1 次印刷

书 号 ISBN 978-7-5608-5545-5

定 价 30.00 元

本书若有印装质量问题,请向本社发行部调换 版权所有 侵权必究

前 言

面对日益增长的信息,人们在寻觅自己感兴趣的、有用的、有价值的信息时,需要花费大量的时间和精力。“信息过载”、“信息迷航”、“缺乏个性化服务”等日益成为阻碍人们享受数字化生活的严重问题。推荐系统的出现成为解决上述问题的重要方法。推荐系统在电子商务、电子政务、网络学习、电子图书馆等多个领域中有重要的应用价值及经济价值。然而,随着信息数量的迅猛增长、对服务的个性化需求以及系统环境的不断变化,现有的推荐系统面临很多新的挑战。因此,推荐系统需要借鉴新的技术及理论方法实现提升及改善。与此同时,一些新理论、新技术的出现为改善推荐系统提供了基础,包括:Agent 技术,粗糙集理论、智能优化方法等。

在上述背景下,本书将 Agent 技术应用于推荐系统,以改善现有的推荐方法。本书同时探讨了粗糙集理论及智能优化方法在其中综合应用问题。本书研究的内容主要包括如下几个方面:

1. 推荐系统及应用中相关理论分析。对推荐系统的概念、分类及推荐方法方法进行了综述。对 Agent 的概念、结构及工作原理进行了综述。对粗糙集理论的概念、定义及主要方法进行了分析。对智能优化的概念及方法进行了综述。这些理论为本书提供了理论和技术基础。

2. 基于“效用”(Utility)理论的推荐方法研究。针对当前推荐方法缺乏效用评价能力,本书提出基于用户偏好的多属性效用推荐方法。其中,对用户偏好的识别是实现效用推荐的关键问题。对此,本书利用可变精度粗糙集挖掘用户对属性的偏好,并采用 Agent 对用户偏好动态地识别及学习。然后,本书采用 TOPSIS 方法,结合用户偏好对推荐对象进行效用评价,实现最佳的个性化效用推荐。通过相关实验分析,表明新方法有更好的个性化效用推荐能力及更高的推荐准确性。

3. 基于 BDI Agent 的协作推荐模型研究。针对现有推荐系统缺乏自适应能力、自组织能力,本书提出基于多 Agent 的协作推荐模型。模型将 Agent 认知理论方法应用于协作推荐,同时借鉴合同网实现推荐任务拍卖,让多种推荐方法参与推荐,并依据用户的反馈评价各种推荐方法,从而形成协作机制。模型利用 BDI 理论对系统中的角色进行心智状态建模,设计了 Agent 心智状态与行为的驱动逻辑,给出形式化的描述方法。基于 BDI 的 Agent,实现了意图与行为的模式匹配,使得系统工作有序。通过相关实验,表明基于新模型的推荐系统有较好的适应能力及自组织能力。

4. 基于遗传算法的粗糙集属性约简研究。针对推荐方法中存在的数据挖掘问题,本书将仿生学的遗传优化算法引入到属性约简求解中,使得在属性约简解空间中,每个可能的解都有机会参与运算。新方法消除了启发式算法约简属性过早收敛问题,为获得更好的属性约简提供了新的求解途径。针对现实数据中的噪声数据问题,通过引入误差

因子,使得算法的适应性更高、容错性更强。在算法实现上,新方法用关系演算描述了算法,并在关系型数据库基础上实现了算法。通过含有噪声数据的实验,验证了方法的有效性。

5. 基于合同网的协作粒子群优化方法研究。针对协作推荐系统中存在的优化问题,本书将 Agent 技术与粒子群算法相结合,进行了优化方法研究。新方法利用 Agent 来代理粒子群,采用合同网实现协作机制。新方法不仅实现了粒子群间进化知识的共享,而且有并行进化能力,克服了独立粒子群进化能力的不足。通过基于 Agent 的背包问题求解实验,验证了新方法有较好的优化效率和效果。

由于推荐技术研究涉及的领域非常广阔,又限于作者水平,本书难免会有不完善之处,恳请广大专家、学者批评指正。本书参考的相关资料尽可能逐一列出,向所有文献的作者表示感谢,但由于文献较多,如有疏漏,在此表示歉意。

本书内容是上海市教育委员会创新课题(11YZ256)的部分研究成果,其出版获得了上海远程教育集团学术专著出版基金资助。在此,向资助单位表示衷心感谢。

编者

2014 年 3 月

目 录

1 绪 论	1
§ 1.1 研究背景及意义	1
§ 1.1.1 研究背景	1
§ 1.1.2 研究意义	2
§ 1.2 推荐系统概论	3
§ 1.2.1 推荐系统的概念	3
§ 1.2.2 推荐系统的研究内容	4
§ 1.2.3 推荐系统研究与应用现状	5
§ 1.2.4 推荐系统面临的问题	8
§ 1.2.5 智能 Agent 与推荐系统	9
§ 1.3 本书的主要工作	11
§ 1.3.1 研究的内容	11
§ 1.3.2 本书的结构	12
2 相关技术研究	14
§ 2.1 推荐系统	14
§ 2.1.1 推荐系统的分类	14
§ 2.1.2 推荐系统的主要算法	16
§ 2.1.3 推荐方法的评价	20
§ 2.2 Agent 技术	21
§ 2.2.1 Agent 的概念及应用	21
§ 2.2.2 Agent 的体系结构	23
§ 2.2.3 Agent 通讯语言	24
§ 2.2.4 多 Agent 系统的概念	26
§ 2.2.5 多 Agent 系统的交互与协作	27
§ 2.3 粗糙集理论	30
§ 2.3.1 粗糙集理论概述	30
§ 2.3.2 粗糙集理论基本概念及定义	31
§ 2.3.3 粗糙集理论的主要工作	33
§ 2.4 智能优化概念及方法	34
§ 2.4.1 进化计算概述	34
§ 2.4.2 智能优化的基本方法	35

§ 2.4.3	优化算法的基本工作步骤	36
§ 2.4.4	基于 Agent 的智能优化技术	37
§ 2.5	本章小结	38
3	基于粗糙集及多属性效用的推荐研究	39
§ 3.1	引言	39
§ 3.2	现有推荐方法分析及问题提出	39
§ 3.2.1	现有推荐方法分析	39
§ 3.2.2	用户偏好研究	40
§ 3.2.3	个性化效用推荐问题的提出	40
§ 3.3	多目标决策概述	41
§ 3.3.1	多目标决策概念	41
§ 3.3.2	多属性效用理论	42
§ 3.4	TOPSIS 方法	43
§ 3.4.1	TOPSIS 方法概述	43
§ 3.4.2	TOPSIS 数学模型	43
§ 3.5	基于粗糙集的属性重要度识别	45
§ 3.5.1	粗糙集知识表达及基本概念	45
§ 3.5.2	属性依赖度	46
§ 3.5.3	属性重要度	46
§ 3.5.4	基于粗糙集的属性权重识别算法	46
§ 3.5.5	属性重要度权重算例分析	47
§ 3.6	基于粗糙集及多属性效用推荐研究	48
§ 3.6.1	基于粗糙集及多属性效用推荐工作方法	48
§ 3.6.2	基于粗糙集及多属性效用的推荐流程	49
§ 3.6.3	基于粗糙集及多属性效用的推荐系统框架	49
§ 3.6.4	基于可变精度粗糙集的用户偏好识别	49
§ 3.6.5	基于用户偏好的多属性效用推荐算法	52
§ 3.6.6	基于用户偏好的协同过滤推荐方法	53
§ 3.6.7	实验及分析	54
§ 3.7	本章小结	56
4	基于 BDI Agent 的协作推荐研究	58
§ 4.1	引言	58
§ 4.2	现有推荐方法分析	58
§ 4.2.1	现有推荐方法的不足与挑战	58
§ 4.2.2	基于多 Agent 的推荐系统的优势	60

§ 4.3 BDI 模型及合同网概述	61
§ 4.3.1 BDI 模型	61
§ 4.3.2 合同网	63
§ 4.4 基于 BDI Agent 协作推荐模型	63
§ 4.4.1 Agent 的分工与职能	64
§ 4.4.2 基于协作推荐的改进合同网	65
§ 4.4.3 基于 ACL 的协作推荐合同网会话协议	66
§ 4.4.4 BDI Agent 工作逻辑	68
§ 4.4.5 协作推荐结果的生成方法	68
§ 4.4.6 协作权重设置及更新	68
§ 4.4.7 协作推荐系统框架模型	69
§ 4.4.8 实验及分析	70
§ 4.5 本章小结	72
5 基于遗传算法的粗糙集属性约简研究	73
§ 5.1 引言	73
§ 5.2 信息系统属性约简问题概述	73
§ 5.3 粗糙集理论属性约简问题分析	74
§ 5.3.1 粗糙集理论	74
§ 5.3.2 粗糙集属性约简方法分析	74
§ 5.3.3 粗糙集属性约简相关概念及定义	75
§ 5.3.4 粗糙集求属性约简方法	77
§ 5.4 遗传算法分析	79
§ 5.4.1 遗传算法概述	79
§ 5.4.2 遗传算法相关概念及定义	80
§ 5.4.3 遗传算法步骤	81
§ 5.5 基于遗传算法的变精度粗糙集属性约简	81
§ 5.5.1 关系模型及关系数据库	81
§ 5.5.2 基于关系运算的可变精度粗糙集	83
§ 5.5.3 基于遗传算法的可变精度粗糙集属性约简	85
§ 5.5.4 实验及分析	88
§ 5.6 本章小结	89
6 基于合同网模型的协作粒子群优化研究	91
§ 6.1 引言	91
§ 6.2 协作推荐的优化问题分析	91
§ 6.3 粒子群算法概述与分析	92

§ 6.3.1	粒子群算法概述	92
§ 6.3.2	粒子群算法相关定义及算法过程	92
§ 6.3.3	粒子群算法研究的现状与不足	94
§ 6.4	多 Agent 系统的协作	95
§ 6.4.1	多 Agent 协作的概念与优点	95
§ 6.4.2	多 Agent 协作相关模型及相关研究	96
§ 6.5	基于合同网的协作粒子群优化模型	98
§ 6.5.1	基于合同网的协作粒子群优化模型概述	98
§ 6.5.2	基于合同网的协作粒子群优化模型相关定义	100
§ 6.5.3	基于合同网的协作粒子群背包求解算法	102
§ 6.5.4	实验及分析	104
§ 6.6	本章小结	105
7	结论与展望	106
§ 7.1	本书总结	106
§ 7.2	进一步研究展望	107
参考文献		109

1 绪 论

§ 1.1 研究背景及意义

§ 1.1.1 研究背景

随着信息时代的到来以及 Internet 的广泛应用,信息在人们工作、学习、生活中的运用达到了前所未有的广度及深度。依据中国互联网络信息中心(CNNIC)发布的第 31 期《中国互联网络发展状况统计报告》,截至 2012 年 12 月底,我国网民规模达 5.64 亿,2012 年共计新增网民 5090 万人。互联网普及率为 42.1%,较 2011 年底提升 3.8 个百分点。我国域名总数为 1341 万个,其中“.CN”域名总数为 751 万,占比为 56.0%;“.中国”域名总数为 28 万个。中国网站总数(即网站的域名注册者在中国境内的网站数,包括在境内接入和境外接入)继续回升至 268 万个。网民的年龄结构中,20—29 岁占据 30.4%,30—39 岁占据 25.3%。从网民职业结构中,学生占据比例最高,达到 25.1%。从学历结构分析,高中/中专/技校的占据 32.3%,大专占据 9.8%,大学本科以上占据 11.3%。2012 年,我国搜索引擎用户规模为 4.51 亿,较 2011 年底增长了 4370 万人,年增长率 10.7%。搜索引擎作为互联网的基础应用,是网民获取信息的重要工具,其使用率自 2010 年后保持在 80%左右水平,稳居互联网第二应用之位。

上述这些统计资料反映出,互联网应用日益融入人们生活的方方面面。根据 CNNIC 发布的《2012 年中国网络购物市场研究报告》显示,2012 年我国网络购物市场交易金额达到 12594 亿元,较 2011 年增长 66.5%。2012 年网络零售市场交易总额占社会消费品零售总额的 6.1%。截至 2012 年 12 月底,我国网络购物用户规模达到 2.42 亿,网民使用网络购物的比例提升至 42.9%。从网络购物满意度分析,2012 年,对网络购物整体表示满意的用户为 89.3%,与 2011 年基本持平。

从用户的网络购物行为分析,根据 CNNIC 的报告,购物者经常使用社会化分享网站如:微博、蘑菇街等,获得信息进行购物。据分析有 28.4%的网购用户表示自己最近半年使用过社会化分享网站。使用社会化分享网站的网购用户中,有 52.8%的人表示自己在社会化分享网站上浏览、关注过商品购物方面信息。从不同年龄代的购物行为分析,在购买不熟悉的产品时,60 后看重购物网站的品牌;70 后追求品牌的性价比,看重购物网站的品牌和产品价格的高低;80 后更看重用户评价与网站的知名度/口碑;90 后则更看重于用户评价。购物行为与不同年龄代际的生活态度和购物习惯相关。上述情况表现出,在电子商务领域,社会化分享影响用户购物的作用不容忽视,有共同偏好的群体之间分享信息的作用十分明显。同时,不同用户群体使用信息、评价信息价值的方式各不相同,商家需要提供不同的信息服务,使用不同的销售策略。

与此同时,网络教育在互联网的支持下蓬勃发展。网络教育是指在网络环境下,以

现代教育思想和学习理论为指导,充分发挥网络的各种教育功能和丰富的网络教育资源优势,向受教育者和学习者提供一种网络教和学的环境,传递数字化内容,开展以学习者为中心的非面授教育活动。^[1]网络教育的形式很多,包括学历教育、职业教育、继续培训等。随着高等学历教育在网络教育中的实施,网络教育发展十分迅猛。根据《中国教育统计年鉴 2012》显示,2012 年网络教育本科生在校生总数为 2 002 698 人,专科生在校生总数为 3 701 414 人,分别比 2011 年增长 14.13%、16.76%。随着知识经济的到来,学习型社会的建设,人们普遍发现依靠传统的学校教育不能满足人们长期发展的需要,只有通过终身学习才能适应社会发展的需要。此外,求职的困难也使得一些学习者不愿意放弃现有的工作机会。因此,原来越来越多的学习者参与到网络学习中。《国家中长期教育改革和发展规划纲要(2010—2020 年)》提出:“把教育信息化纳入国家信息化发展整体战略,超前部署教育信息网络,建设数字化教育服务体系,充分利用优质资源和先进技术,创新机制和管理模式,整合现有的资源,构建先进、高效、实用的数字化教育基础设施。”网络教育与传统的教育有很多的区别,学习者的知识背景、学习能力、学习进度等差异较大。教学形式也有别与传统的课堂教学,往往依托学习管理系统(Learning Management System)。学习者学习形式包括:个体学习、协作学习、教师指导学习等。因此,网络教育中学习者需要更有质量的信息辅助其学习,个性化信息服务的要求也更显著。

在互联网应用发展到今天,虽然获得信息仍然是人们使用互联网服务的主要目的,但是信息使用的方式、方法正在发生深刻的变化。以电子商务、网络教育为代表的各类互联网应用中,人们对信息的使用表现出社会化共享、个性化服务、更好的用户体验需求等特征。

§ 1.1.2 研究意义

随着网络应用的蓬勃发展,人们对信息服务的需求进入到了一个新阶段,表现出许多新特点。人们对信息的使用从简单的信息查询、检索需求,转变为:信息的度身定制需求、信息的个性化服务需求、信息服务高质量高效率需求、信息深度再利用需求等。然而,在享受信息时代便捷的同时,人们却陷入另外一种困境——信息过载、信息迷航等问题日益凸显。人们为了从海量的信息中辨识出自己实际需要的、感兴趣的、有价值的信息,不得不花费大量的时间、精力,乃至财力。对此,一些新的技术不断地涌现出来,包括搜索引擎技术、垂直搜索引擎技、RSS 信息聚合技术等等。然而,现有的方法和技术与用户的需求和用户体验之间还存在很大的差距,具体表现在如下几个方面:

1. 检索能力较差。传统的搜索引擎的工作是依据用户检索关键词触发的。检索的质量,往往与用户本身对概念的概括能力、分词能力等信息使用能力相关。用户输入不恰当的检索词,检索工具不能为用户提供好的检索结果。同时,检索工具也不具备辅助用户检索的功能。

2. 信息过载问题严重。计算机网络是一个开放式的环境,信息资源不仅迅猛增长,而且资源中的概念重叠,词序、词义错位现象普遍存在。应用系统往往将大量的信息一次性地提供给用户,对此用户需要花费很多的精力和时间进行选择、过滤,表现出严重的信息过载情况。

3. 个性化服务能力不足。现有系统提供统一的、无差别的服务,缺乏个性化服务能力。部分系统开始采用个性化服务功能,但是对用户的个性化分析能力不足,服务的质量和针对性有待改善。

4. 用户之间的联系能力不足。传统的信息服务,将用户与资源分隔开来,没有记录用户对资源的评价及使用行为。因此,无法寻找到用户与资源之间、用于与用户之间以及资源与资源之间的隐形联系,使得信息服务的质量不高。

5. 系统的交互能力不足。现有系统与用户的交互需要用户输入较多的信息,如:检索词、查询对象的具体属性等,依赖性高,交互过程不友好。同时,系统仅对具体的检索、查询请求给出反馈,没有针对用户的活动动态地给出建议和参考。虽然部分系统开始使用联想功能,但是针对性较差。此外,系统的交互能力处于单向的被动状态,主动交互的能力不足。

针对上述问题,从用户需求角度来看,用户需要现有的信息服务提供更好的服务质量,包括:更准确的服务内容、更好的服务效率、更贴切的服务效果,因此对推荐系统的研究具有很好的应用价值。从理论方法上来说,现有推荐系统需要借助新的技术、方法、途径,来解决日益凸显的服务能力不足与用户需求之间的矛盾。推荐系统与新技术的结合研究具有很好的理论价值。同时,在国家对国民经济和社会发展的规划中,以及国家对教育改革和发展规划内容中,都将信息服务作为发展的重点,对推荐系统的研究有较好的社会效益及经济效益。因此,本书基于 Agent 的个性化推荐系统的研究具有理论意义和实践意义。

§ 1.2 推荐系统概论

§ 1.2.1 推荐系统的概念

在生活中,当人们由于缺乏足够的经验无法对事物做出准确的选择或判断时,往往需要借鉴他人的意见和建议信息。这些信息包括:简述、建议书、评论等等。推荐系统正是据此为原型,实现推荐信息的采集、处理及传递。推荐系统涉及多个学科,包含多种方法,包括:管理信息系统、管理学、检索技术、过滤技术等等。然而,它作为一个独立的发展方向,受到研究者及商业应用的广泛关注。Resinck 和 Varian 在参考文献[2]中,较早提出推荐系统的概念。他们认为推荐系统的主要工作是采集用户给出推荐信息,将推荐信息整理后推送给适合的人。其中,“整理信息”的工作是推荐系统的主要工作。此外,将推荐信息与需求者之间有效的匹配问题,也是推荐系统的研究的重要内容。在参考文献[2]中,Resinck 和 Varian 对推荐系统面临的困难进行了讨论,这些困难包括:用户评价的多样化;进入系统的不明确性;匿名使用的问题;评价信息的整合问题;评价结果用途的多样性。电子商务作为一种新型的商务运行模式,较早地应用了推荐系统。其中,参考文献[3]、[4]调查了多个商业网站,包括:Amazo. com、CDNOW、Ebay、Drugstore. com、MovieFinder. com、Reel. com,评价推荐系统对电子商务的影响及应用模式。上述文献认为推荐系统适应了电子商务发展的需求,它为商家实现基于用户个性化的商业活动提供了有效的支持。在电子商务运营中,传统的商务模式的优点已不复存在。商家只

有通过为用户提供更好的个性化服务,才能使得用户拥有更好的服务体验,从而产生“忠实感”,最终让商家形成长效的商业价值。因此,推荐系统对电子商务影响力巨大。

推荐系统在电子商务中的作用,可以归结为如下几个方面^[3]:

1. 利用推荐列表,帮助新用户和老用户找到他们需要的商品,留住消费者。电子商务营销的主要内容之一是吸引浏览者的眼球,将浏览者留在自己的网站上。在将浏览者吸引到网站之后,又要防止浏览者由于信息过多而不知所措或产生反感的情绪,而再次选择离开。因此,通过对商品进行分类、整理,利用推荐列表的形式展现给浏览者,引导浏览者进一步关注自己感兴趣的内容,从而将浏览者转变为消费者。推荐列表有很好的效果,可以在不需要用户个性化信息的情况下,取得广而告之、显而易见的引导作用。

2. 通过参与评价与打分来提高用户对商家的忠实度。电子商务的卖家给消费者的感受往往是一个电子网站,不是一个实实在在的商户,因此普遍存在重视度低的问题。而品牌的忠实度往往是企业、商家得以长期生存发展的基础。让消费者有宾至如归、一对一的个性化服务的感受,是提高用户忠实度的方法之一。这要求商业网站能依据用户的特点提供个性化服务,为此商务网站需要采集大量的用户浏览和评价信息。通过上述工作,让用户体验到了区别于其他网站的服务,从而提高了忠实度。

3. 通过相关性商品推荐,提高总体销售量。当商家可以依据消费者的兴趣实现商品推荐时,就可以实现将相似种类、相关联的种类商品进行交叉推荐。这样可以提高整体销售的额度,从而实现多元化的赢利。

4. 建立长期的联系以实现深化的个性化服务。零售商业的终极目标是与消费者建立长期的联系,从而长期的赢利并形成竞争壁垒。深化的个性化服务是根据用户长期的浏览行为、购买记录、网站导航记录形成。建立深化的个性化服务是形成长期联系的重要手段,同时也非常难以实现。

目前,许多知名的网站都试用了推荐服务。国外知名的网站包括:Amazo. com、CD-NOW、Ebay、Drugstore. com、MovieFinder. com、Reel. com。国内知名的网站包括:Taobao、EachNet 等,也开始采用推荐的方法。这些网站提供的各种推荐服务能较好地帮助用户快速地定位自己的所需的商品等。

§ 1.2.2 推荐系统的研究内容

推荐系统是一个人机交互系统,综合了信息检索、信息过滤、数据挖掘、人工智能等多个学科的内容。虽然它有一定的综合性特征,但是很多组织、研究单位都将它作为一个独立的领域开展研究,例如:人工智能会议(AAAI)、协同工作会议(CSCW)、美国计算机协会(ACM)都在年度的会议及活动中,将推荐系统作为一个研究主题。推荐系统的研究内容主要有:

1. 推荐算法的研究。推荐算法是实现推荐服务的核心内容。当前,推荐算法主要包括:基于内容的推荐、协同过滤技术、混合推荐方法等。由于协同过滤技有较好的适应性,因此成为当前的研究热点。同时,由于协同过滤技术依赖用户对推荐对象的评价信息,因此也存在很多不足之处。

2. 数据挖掘在推荐系统中的应用。随着数据挖掘技术应用领域的拓展,各种数据挖

掘技术在推荐中的应用逐渐成熟起来,包括:关联规则挖掘、粗糙集方法、聚类分析、序列挖掘等。数据挖掘研究为推荐系统的发展融入了很多新的元素。随着系统各种数据规模的增长,基于内存的推荐方法不仅难以满足推荐实时性的要求,并且由于空间开销问题,难以完成计算任务。因此,借助数据挖掘技术,对用户模式进行离线的挖掘,是解决上述问题的途径之一。利用挖掘得到的知识、规则等结果,指导推荐工作。因此,数据挖掘技术与推荐系统的结合研究是推荐系统研究的主要内容之一。

3. 推荐信息的采集。对用户信息的采集是实现推荐的基础。因此,如何更有效、准确、友好的采集到用户的兴趣偏好信息,是推荐系统研究的重点问题。当前推荐系统多依赖于显示的兴趣采集方法,如何利用隐式的采集方法更合理有效地实现信息采集是推荐系统研究的主要方面。

4. 用户隐私保护研究。由于推荐系统需要搜集、分析用户的兴趣偏好等信息,这涉及用户隐私信息的保护问题。因此,如何在推荐系统既为用户提供良好的推荐服务,又能有效地保护用户的个人信息安全是当前需要解决的问题。

5. 推荐系统的可视化研究。如何为用户提供一个友好的可视化的交互界面成为推荐系统重要的研究主题。这部分研究的重点是推荐结果的可视化研究和推荐结果的有效解释等方面内容。

6. 多种技术的集成应用。随着研究的深入、应用的深化以及应用领域拓展,新型的推荐系统需要有更强的信息采集能力、自组织能力、环境适应能力以及协同工作能力。因此,需要通过集成新的技术,通过吸取其他技术的优点,来不断改进和提升现有推荐系统的服务质量及能力。

§ 1.2.3 推荐系统研究与应用现状

按照实现方法,目前有代表性的推荐系统的研究与应用如下:

1. 基于内容的推荐系统(Content-based Recommendation System)^[5, 6]

该类型推荐系统方法的核心是将现有的资源与用户过去访问的资源进行内容比较,找出最相似的资源作为候选集合。再依据相关的选择策略,从候选集合中选取部分进行推荐。该类型的推荐系统可以采集用户对文本类资源的使用情况,然后建立用户兴趣及偏好模型,从而能形成长效的推荐。对用户信息的采集方式,可以是显示的问卷形式,也可以是隐式的动态采集。该类型的推荐算法适用于图书、电视节目、Web 站点信息、BBS 信息等文本类信息资源的。对于非文本类资源如音频、视屏类资源的推荐有比较大的困难,主要原因是对上述资源的内容分析技术尚无突破性的进展。

参考文献[7]、[8]中,Thorsten Joachims 等人设计了一个名为 WebWatcher 的推荐系统。在 WebWatcher 系统中,基于 HTML 网页的结构化特征,通过网页之间的链接关系,通过建立“最小描述长度规则”方法寻找网页资源之间的相关性。当用户访问了某网页后,系统会推送“最小描述长度规则”计算出的相关网页给他们。这种方法利用了 HTML 网页的特殊结构,在一定程度上消除了文本分析的困难。

参考文献[9]中,提出了一个名为 Siteseer 的推荐系统。这个推荐系统将用户对网页的标注看作是兴趣点的描述信息。同时将用户对信息的分组,看作是基于语义的分类或

者相关分组。系统的核心思想是根据不同用户兴趣之间的重叠度来实现推荐。推荐方法屏蔽了基于语法、语义的推荐的复杂性。

参考文献[10]中,Bamshad Mobasher 设计了一个名为 WebPersonalizer 推荐系统。系统设计者认为,目前的个性化网站过多地依靠用户的兴趣输入。因此,在他们设计的系统中,可以在用户匿名登录及不需要输入喜好信息下实现个性化网站服务。这个系统利用离线的数据准备及数据挖掘,并与在线的推荐相结合为用户实现推荐。系统的核心是运用基于聚类分析的数据挖掘方法对用户的网上浏览信息实现进行挖掘。

基于内容的推荐系统目前仍然应用广泛,并且很多研究者将它与协同过滤方法联系起来,改善后者的很多不足之处,形成混合型推荐系统。

2. 基于协同过滤的推荐系统(Collaborative Filtering Recommendation System)

协同过滤技术的基本思想是:依据用户对项目/商品的评价,找出与当前用户相似的用户集合,将他们评价过的项目经过选择、排序等处理后推荐给当前用户。典型的协同过滤的推荐系统,有如下一些:

Ringo^[11]是推荐音乐的个性化推荐系统。用户听过系统提供的音乐后,对其打分。系统记录下用户的评分,形成用户的描述信息,并通过其推荐音乐。

FireFly^[12](<http://www.firefly.com>)早期的使用协同过滤技术与 Agent 技术的商业网站。它提供的推荐商品包括:音乐、电子书籍、服装、生活用品等。Firefly 提供搜索功能,帮助用户快速找到所需的商品,并能提供价格比较服务,缩短用户查找商品的时间。

GroupLens^[13]是一个新闻推荐系统。它让用户看完新闻后,使用一个叫作 Better Bit Bureaus 的评分系统对此新闻进行评分。然后,系统依据用户评分等信息实现新闻推荐。

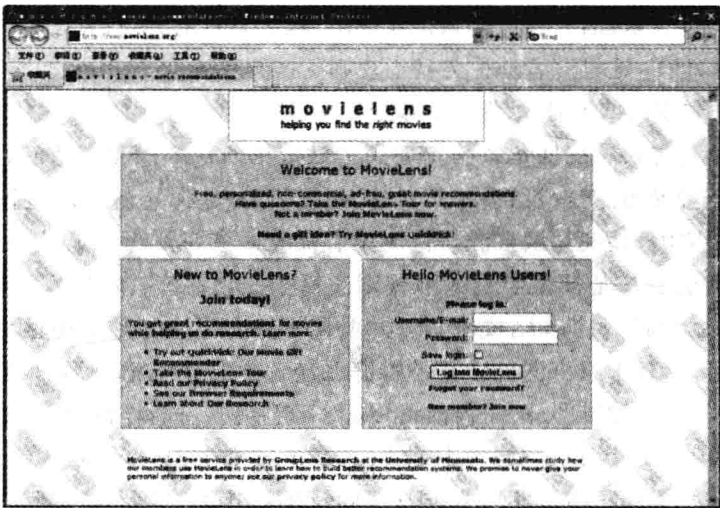


图 1-1 MovieLens 系统界面

Video Recommender^[14]与 MovieLens^[15]是两个视频推荐系统。其中,前者是 Will Hill 等在参考文献[14]中提出的依据虚拟社区(Virtual Community)用户的对视频的评价,实现的视频推荐的系统。这里的虚拟社区不是指虚拟现实,也不是智能 Agent 技术,

而是一个大家共同参与的视频评价环境。MovieLens (<http://www.movie.org>) 是明尼苏达大学研究者们设计的一个视频推荐系统,系统界面如图 1-1 所示。这个推荐系统保存了大量的用户对电影的评价信息,并且将这些数据公布出来,让研究推荐系统的工作者分享。

通过研究者们改进与发展,协同过滤推荐系统又可以分为:基于用户的协同过滤(User-based),基于项目的协同过滤(Item-based),基于模型的协同过滤(Model-based)。协同过滤推荐系统的最大优点在于:它是数据驱动的推荐系统;它不需要对推荐项目做预处理,也不需要为推荐项目做内容分析工作。因此,可以适应各种类型的推荐内容,包括文本、视屏、音频等。

协同过滤技术也存在不足之处,包括:①评价数据稀疏问题。当系统中保存的用户评价数据量较少时,会出现稀疏评价矩阵问题,导致较差的推荐质量。同时,对于一个新添加的项目,由于没有用户参与评价,因此无法对其实现推荐。②新用户零启动问题。由于新用户没有对项目做过评价,因此无法为新用户提供推荐服务。③系统的实时推荐能力问题。随着系统规模、数据规模、用户规模的扩大,传统的协同过滤推荐系统的时间复杂度、空间复杂度问题凸显,使得系统无法实现实时推荐任务。很多学者对协同过滤推荐系统提出了改进方法,如参考文献[16]中提出的基于降维的方法;参考文献[17], [18]中提出的基于聚类的改进方法;参考文献[19]中讨论了基于 Memory-based 与 Model-based 混合协同过滤的方法。此外,一些学者从用户兴趣变化的追踪角度,探讨了协同过滤的改进,如参考文献[20]中提出了基于时间的数据权重和基于资源相似度权重的改进方法,将两者进行有机结合,取得了较好的改善效果。

3. 基于规则的推荐系统

这一类型的推荐系统将关联规则、神经网络、基于案例的分析、基于知识的分析等技术融入到推荐系统中,这里笔者通称为基于规则的推荐系统。

(1) 基于关联规则的推荐系统

参考文献[21]讨论了在推荐系统中使用关联规则对用户及项目进行规则挖掘。在挖掘扩充中采用不限定支持度的方法,保证获取的规则数量在一定范围内。通过在 EachMovie 数据集测试表明,该方法提高了推荐质量。参考文献[22]认为,由于关联规则的发现会产生很多条规则,直接将关联规则应用于推荐系统是困难的。对此提出基于模板的规则,通过设计一个规则模板,提取对用户有效的规则再实现推荐。参考文献[23]中,提出了一个基于用户浏览信息的推荐系统。通过系统记录用户的浏览记录,再采用基于关联规则数据挖掘方法改善推荐质量,实现将网页推荐给用户。

(2) 基于案例的推荐系统 (Case-based Recommendation System)

类型的推荐系统从案例出发形成推荐规则及推荐知识。如参考文献[24]分析了基于案例的推荐系统的优点及与其他推荐方法的比较,认为在商品推荐方面基于案例的推荐系统有很多优势。参考文献[25]提出了基于案例的目的地(如:餐馆)的推荐系统。参考文献[26]提出了基于案例的旅游推荐系统。基于案例的推荐可以较少受到由于用户评分数据少的影响,可以弥补协同过滤方法的不足。

(3) 基于知识的推荐系统(Knowledge-based Recommendation System)

基于知识的推荐系统是基于用户需求的推理,知识可以在一定程度上表达用户的具体需求。参考文献[27]提出协同过滤与知识的推荐系统相结合的推荐系统。在数据比较稀疏的情况下,这个系统使用知识推荐,利用协同过滤方法对知识推荐结果再过滤。参考文献[28]认为基于知识的推荐系统是独立运行的,不是基于用户对商品的打分,所以协同过滤方法中的不足点对它的影响不大。

(4) 基于数据挖掘的推荐系统

这一类型的推荐系统包括使用各种具体数据挖掘方法的推荐系统。参考文献[29]探讨了将 Web 挖掘、决策树、关联规则挖掘、分类方法应用于推荐系统,提高推荐质量的方法。参考文献[30]建立了一个拥有两种类型 Agent 的推荐系统,一种使用协同过滤技术,另外一种是使用 k-mean 的发掘方法,并且比较了两种方法的优越性。通过比较认为,在用户评分数据量少的情况下 k-mean 的 Agent 能帮助获得较准确推荐。参考文献[31]提出一个电视节目推荐系统,这个系统内使用神经网络(Neural Network)方法综合了基于个人浏览历史的贝叶斯推荐/决策树方法,基于家庭浏览历史的贝叶斯推荐/决策方法,以及直接的推荐五种方法。

基于数据挖掘的推荐方法有如下特点:对用户评分的依赖程度较小,对用户的干扰较弱;往往不需要领域知识;自动化程度较高、可以实现持续的推荐能力;有发现隐藏的关系能力,包括项目之间的和用户之间的;能较好地追踪用户的兴趣变化,通过较少的训练数据能快速地获得用户的兴趣模型并能更新。

(5) 混合型推荐系统(Hybrid Recommendation System)

混合型的推荐系统是通过综合多种推荐方法改善推荐质量。参考文献[32]提出一个基于图(Graph-based)推荐系统,将基于内容的推荐与协同过滤的推荐综合起来,形成一个混合推荐系统。前面讨论的参考文献[29]、[30]、[31]也都属于混合推荐系统,它们从不同角度探讨了各种方法的综合应用问题。参考文献[33]比较了四种推荐系统,七种组合策略,认为层次性的、增广的组合方式工作得更好,特别是在组合两种不同方法的推荐系统中。

§ 1.2.4 推荐系统面临的问题

随着信息技术的发展,网络中信息量的海量增长,信息服务的负载能力不断地增加,现有的推荐系统在发展过程中面临着新的挑战和问题,总结如下:

(1) 系统缺乏柔性。目前,推荐系统大多是针对特定的应用定制开发的,并且系统模型一旦确定,就难于更改与扩充。现有的推荐系统缺乏灵活性、组装性和重用性,并且系统集成代价高、移植性差。

(2) 系统缺乏智能性。对所有的用户采用相同的推荐策略是目前推荐系统存在的问题之一。系统不能根据当前用户的个性化信息丰富程度,推荐方法的适应性,动态地选择不同的推荐策略及方法,从而为用户提供更好的个性化服务。

(3) 缺乏主动性。目前,大部分的系统需要用户给出比较明确的检索信息,如:名称、类别、产地等,才能实时检索或推荐。系统对用户的依赖程度较高,干预用户的浏览行