

社会化标注系统中个性化 信息推荐模型研究

武慧娟 著



科学出版社

社会化标注系统中个性化 信息推荐模型研究

武慧娟 著

科学出版社

北 京

内 容 简 介

个性化信息推荐是充分利用互联网上用户产生的数据来解决目前的个性化信息服务问题的有效方法,本书针对社会化标注系统中的个性化信息推荐模型问题展开研究,在综述“社会化标注系统”“个性化信息推荐”的基础上,对国内外“社会化标注系统中的个性化信息推荐模型”的研究现状进行分析,通过自组织理论、社会网络分析理论、系统动力学等理论与方法,围绕社会化标注系统的演化过程、用户关系网络、个性化信息推荐模型构建等核心问题进行详细分析,最后从静态、动态两个角度分别构建社会化标注系统中的个性化信息推荐模型。

本书可供从事信息管理和应用、信息系统设计与开发、企业信息系统等相关领域的人员参考,也可作为高等院校信息管理系统、计算机应用、电子商务等专业的高年级本科生和研究生的教学参考书。

图书在版编目(CIP)数据

社会化标注系统中个性化信息推荐模型研究/武慧娟著. —北京:科学出版社, 2016.3

ISBN 978-7-03-047417-9

I. ①社… II. ①武… III. ①社会化—信息管理—研究
IV. ①G203

中国版本图书馆 CIP 数据核字(2016)第 038542 号

责任编辑:王 哲 董素芹 / 责任校对:包志虹

责任印制:张 倩 / 封面设计:迷底书装

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>

新科印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2016 年 3 月第 一 版 开本:720×1 000 1/16

2016 年 3 月第一次印刷 印张:8 1/2

字数:158 000

定价:45.00 元

(如有印装质量问题,我社负责调换)

前 言

在 Web 3.0 时代,人们不再满足由机器挖掘给用户的各种信息,而是要结合自身的偏好享用个性化的信息服务,无论是产品还是服务,都将为每个用户量身打造,信息世界将变得越来越聪明和智能,似乎比用户还了解他想要的是什么,这就是个性化的信息服务,即 Web 3.0 的内涵。因此,如何在浩如烟海的信息中寻求个性化的信息,是目前学术界需要解决的热点研究问题。同时,随着互联网的产生和使用,用户在社会化标注系统(Web 页面)上产生了大量的代表自己个性化信息的数据,那么如何充分利用用户产生的数据来解决目前的个性化信息服务问题,就成了情报学术界研究的一个热点学术问题。

社会化标注系统主要是以“用户-资源-标签”三元关系为研究对象的典型复杂动态网络,用户可以根据个人的需要自由选择词汇对所喜爱的资源进行标注,每添加一个词汇称为对资源添加一个标签,用户、资源和标签组成了社会化标注系统的三个基本元素。从这三个最基本的组成元素开始,从中提炼用户的个性化信息,进而形成个性化信息推荐,成为本书的研究起点。

本书是教育部人文社会科学研究青年基金项目(项目编号:15YJC870024)的成果之一,针对社会化标注系统中的个性化信息推荐模型问题展开研究,在综述“社会化标注系统”“个性化信息推荐”的基础上,对国内外“社会化标注系统中的个性化信息推荐模型”研究现状进行分析,通过自组织理论、社会网络分析理论、系统动力学等理论与方法围绕社会化标注系统的演化过程、用户关系网络、个性化信息推荐模型构建等核心问题进行详细分析,本书内容比较新颖,反映了作者多年来的研究成果。

本书研究的内容主要分为五大部分。

(1)研究目前国内外社会化标注系统、个性化信息推荐两方面的研究进展,分析研究的热点与前沿问题,掌握在具体的研究中其所面临的不足,然后从不足处展开,形成本书所要研究的起点,即利用社会网络分析研究用户关系网络,作为分析个性化信息推荐的逻辑起点。

(2)研究社会化标注系统的演化形式和耗散结构。对社会化标注系统中的用户、资源、标签以及三者之间的关系进行分析,利用自组织理论研究社会化标注系统的演化机理,具体使用超循环和耗散结构分别分析系统中的用户、标签、资源各自的演化形式和系统的多层级耗散结构。

(3)分析社会化标注系统中的用户关系网络。主要分析用户关系网络的结构特征,如网络密度、核心-边缘结构、中心性等,同时,从用户关系网络中挖掘凝聚子群的各种信息偏好,从两个角度(群内和群际)对信息进行分析,发现个性化的信息。

(4) 构建社会化标注系统中的个性化信息推荐模型。从静态、动态两个角度分别构建社会化标注系统中的个性化信息推荐模型；然后利用系统动力学绘制个性化信息推荐的系统模型，主要由系统的因果关系和系统流图两个步骤构成。

(5) 对个性化信息推荐进行实证研究，首先从豆瓣网中抓取数据，然后对其进行自组织演化、用户关系网络结构分析、个性化信息推荐模型构建等。

本书的创新点体现在如下两个方面。

(1) 研究视角的创新。

从社会网络的角度研究在用户标注行为中产生的用户关系网络，从中提取个性化信息，并从群内和群际两个角度进行个性化信息推荐分析。当前的相关研究主要局限于社会化标注系统的网络结构分析，少部分研究涉及了凝聚子群，群内和群际信息推荐的研究成果几乎没有。因此，本书为个性化信息推荐问题的研究引入了全新的分析视角。

(2) 研究内容的创新。

采用“移植”借鉴的方法，通过科学“移植”信息科学领域的经典理论，对社会化标注系统中的个性化信息推荐模型进行深入研究，创新认识，具体体现在：根据自组织理论提出社会化标注系统的超循环演化形式和多层级耗散结构，深化系统演化的过程，为模型建立打下深厚的理论基础；根据社会网络分析理论提出基于群内的个性化信息推荐和基于群外的个性化信息推荐两种推荐方法，深化了方法论思想；构建个性化信息推荐模型的静态模型和动态模型，并运用系统动力学对个性化信息推荐模型进行动力学影响因素分析，提高了模型化的程度。

本书的研究结论具体如下。

(1) 基于自组织理论的社会化标注系统的演化形式遵循超循环理论，同时具有层级耗散结构特征。社会化标注系统具有自组织系统的开放性、远离平衡态、非线性相关性、随机涨落等特征；然后运用超循环理论分别对用户集、资源集、标签集和社会化标注系统的自组织演化机理进行了探讨，并构建了社会化标注系统的自组织演化模型；最后根据耗散结构理论，从多层级的角度，对社会化标注系统的序化形成过程进行研究，提出了系统序化的平衡极点。

(2) 从社会网络分析的角度，可以发现社会化标注系统的用户关系网络结构和凝聚子群。主要分析了用于关系网络的网络密度、核心-边缘结构、中心性等，还分析了用户关系网络的凝聚子群，利用块模型的“结构对等性”对行动者进行聚类，利用 K-核分析对块模型进行补充，利用结构洞方法发现存在结构洞的用户，即基于群际的用户。

(3) 社会化标注系统中的个性化信息推荐模型可以分为静态模型和动态模型来构建。首先从系统学的角度研究构成个性化信息推荐系统的六大要素；然后在遵循构建概念的原则下，构建了个性化信息推荐的静态模型和动态模型；最后利用系统动力学对个性化信息推荐的模型进行影响因素分析，即当网络的密度较小时，基于群内的个性化信息推荐与基于群际的个性化信息推荐受网络密度的影响较大；当网络的密度较

大时，基于群内的个性化信息推荐受核心人物的影响较大，基于群际的个性化信息推荐受图的中心势值的影响较大。

本书获得了教育部人文社会科学研究青年基金项目（项目编号：15YJC870024）、东北电力大学博士科研启动基金项目（项目编号：BSJXM-201420）、东北电力大学“十二五”科研提升工程资助计划项目——软科学及人文社会科学专项资助计划项目（项目编号：TSSK-201405）的支持。

在本书的写作过程中参考了许多国内外的研究成果，并尽可能详细地在书中列出，在此对参考文献的作者表示真诚的感谢，也可能因为多方面的原因而有所疏漏，恳请读者批评、指正。

武慧娟

2016年1月

目 录

前言

第 1 章	绪论	1
1.1	研究背景	1
1.2	研究目的与意义	5
1.2.1	研究目的	5
1.2.2	研究意义	5
1.3	研究内容	6
1.4	研究方法与技术路线	7
1.4.1	研究方法	7
1.4.2	技术路线	7
1.5	本章小结	9
第 2 章	社会化标注系统与个性化信息推荐	10
2.1	社会化标注系统研究	10
2.2	个性化信息推荐研究	14
2.3	社会化标注系统中个性化信息推荐模型研究综述	17
2.3.1	国外社会化标注系统中个性化信息推荐模型研究综述	17
2.3.2	国内社会化标注系统中个性化信息推荐模型研究综述	21
2.4	社会网络分析在社会化标注系统中的应用研究综述	22
2.5	本章小结	23
第 3 章	信息推荐的相关理论	24
3.1	熵与自组织理论	24
3.1.1	熵的提出	24
3.1.2	自组织理论	24
3.2	社会网络分析理论	26
3.2.1	社会网络分析的内涵	26
3.2.2	社会网络分析的研究方法	28
3.3	系统动力学理论	30
3.3.1	系统动力学的内涵	30
3.3.2	系统动力学的形成与发展	31

3.3.3	系统动力学的研究方法	32
3.4	本章小结	34
第4章	社会化标注系统的内涵	35
4.1	社会化标注系统的类型	35
4.2	社会化标注系统的自组织特征	36
4.3	社会化标注系统的自组织演化形式	38
4.3.1	二元超循环	39
4.3.2	三元超循环	41
4.3.3	社会化标注系统的超循环	42
4.4	社会化标注系统的耗散结构	44
4.4.1	社会化标注系统的层级关系	45
4.4.2	社会化标注系统的层级耗散结构	45
4.4.3	社会化标注系统的熵	46
4.5	本章小结	48
第5章	社会化标注系统中的用户关系分析	49
5.1	社会化标注系统中的用户关系	49
5.1.1	用户需求分析	49
5.1.2	用户关系的形成	52
5.1.3	用户关系的特征	53
5.1.4	用户关系的类型	55
5.1.5	用户关系的分析流程	58
5.2	用户关系的维度分析	58
5.2.1	知识维	59
5.2.2	信任维	60
5.2.3	结构维	62
5.3	用户关系的社会网络结构	63
5.3.1	幂律分布	63
5.3.2	网络密度	65
5.3.3	核心-边缘结构	65
5.3.4	中心性	66
5.4	用户关系的凝聚子群	69
5.4.1	块模型	69
5.4.2	K-核	73
5.4.3	结构洞	76
5.5	本章小结	79

第 6 章 社会化标注系统中个性化信息推荐模型构建	80
6.1 社会化标注系统中个性化信息推荐模型框架	80
6.1.1 个性化信息推荐模型构成要素	80
6.1.2 个性化信息推荐模型构建原则	81
6.1.3 个性化信息推荐静态模型	82
6.1.4 个性化信息推荐动态模型	83
6.2 社会化标注系统个性化信息推荐模型系统动力学分析	85
6.2.1 个性化信息推荐因果关系分析	85
6.2.2 个性化信息推荐的存量流量图	86
6.3 模型评价	88
6.4 本章小结	89
第 7 章 个性化信息推荐模型实证分析	90
7.1 豆瓣网概况	90
7.1.1 豆瓣的网站架构	90
7.1.2 豆瓣网的自组织演化	90
7.2 豆瓣网的样本数据	91
7.3 豆瓣网的用户关系网络	96
7.3.1 幂律分布	96
7.3.2 网络密度	96
7.3.3 核心-边缘结构	97
7.3.4 中心性	98
7.4 豆瓣网的用户凝聚子群	100
7.4.1 块模型	100
7.4.2 K-核	105
7.4.3 结构洞	107
7.5 豆瓣网的个性化信息推荐模型测试	109
7.6 结果评价	114
7.7 本章小结	115
第 8 章 结论与展望	116
8.1 研究结论	116
8.2 研究展望	117
参考文献	118

第 1 章 绪 论

1.1 研究背景

1. Web 3.0 内涵

2005 年, Bill Gates 提出了互联网新的概念模式——Web 3.0, 这一概念的提出引起了学术界和企业界的巨大反响, 接下来, 专家学者开始对 Web 3.0 的概念和特征进行描述。2007 年, 瑞达网络 (Radar Networks) 公司的创始人 Spivak 认为: Web 3.0 可以描绘互联网发展的第 3 个十年, 即 2010~2020 年, 如图 1.1^[1]所示。

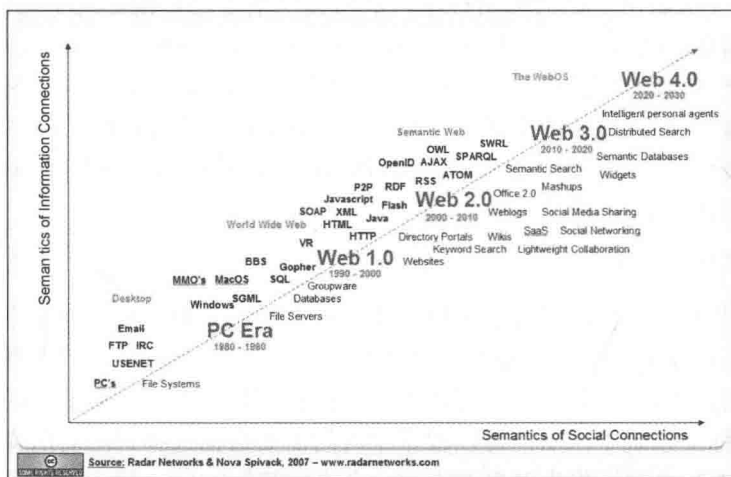


图 1.1 Web 发展史

从互联网发明开始, 即 1990~2000 年, 人类进入了 Web 1.0 的时代, 用户通过各种门户网站的搜索引擎, 快速地搜索任何想要的资源, 这些资源以文本、视频、声音等多种类型的文件出现, 所以 Web 1.0 时代解决了“搜索、下载”的问题。2000~2010 年, 人类进入了 Web 2.0 的时代, 用户可以在网上共同参与一个游戏, 共同解决一个问题, 一夜之间全民都学会了偷菜, 所以 Web 2.0 的时代解决了“参与、互动”的问题。2010~2020 年, 用户急需在杂乱无章的信息中寻找适合自己的、精准的、高效的信息, 即个性化的信息服务面世, 其中主要使用语义网、资源描述框架 (Resource Description Framework, RDF)、OpenID、应用程序编程接口 (Application Programming Interface, API)、Mash-up 等技术进行个性化的信息服务。

Web 3.0 用于信息服务的关键技术主要有 OpenAPI、OpenID、Mash-up 等。其中 OpenAPI 主要实现了不同程序应用平台之间数据的共享,可以容纳更多的第三方应用^[2],实现信息的聚合;使用 OpenID 技术可以使用户通过一个统一资源定位符(Uniform Resource Locator, URL)用户名进行身份的识别,同时可以进入相互兼容的不同平台,既方便了后台对用户的管理,又增强了用户的体验感;Mash-up 是一个 Web 应用程序,它集成了来自多个信息源的内容并将其交付到一个页面中进行显示,服务器向每个内容源发出请求,解析收到的信息,并将结果综合到一个页面中发给浏览器^[3]。

文献[4]指出 Web 3.0 的核心思想就是:Web 3.0 不仅是使用 RDF、OWL(Web Ontology Language)、SPARQL(Simple Protocol and RDF Query Language)标准技术产生的、可以作为各种应用程序数据库的智能化网络,而且对于 Web 2.0 深入发展的结果,是对当前开源软件、资源共享、广泛参与等互联网观念的升华,是本体技术和知识组织观念在网络空间中的延伸和深入发展。文献[5]指出 Web 3.0 的特征有四个,分别是“个性化的信息聚合”,即使用 Mash-up 对轻量级的内容进行集成,用户可以根据自己的需要组装出自己的应用程序,从而提供基于用户偏好的个性化聚合服务;“信息检索的高精度度”,使用用户生成内容(User-Generated Content, UGC)的筛选性分离出可信度更高的信息提供给用户;“搜索引擎的智能化”,使用语义网让人与机器之间的沟通像人与人之间的沟通一样顺畅,自动聚合各种内容 and 应用;“多种终端平台、跨网站和跨语言信息交互”,使用 OpenID 身份识别框架让用户在不同的平台上进行信息交互,实现无国界的沟通。

2. 社会化标注系统

目前,有关社会化标注系统的严格定义还没有统一,但从相关文献中分析,社会化标注系统(social tagging system)主要是以“用户-资源-标签”三元关系为研究对象的典型复杂动态网络,具体是从内部结构、演化机理和应用三方面开展的一个研究方向。在社会化标注系统中,用户可以根据个人的需要自由选择词汇对所喜爱的资源进行标注,每添加一个词汇称为对资源添加一个“标签”(tag),用户、资源和标签组成了社会化标注系统的三个基本元素。

社会化标注自从产生后就得到了大众的广泛关注,情报学、图书馆、博物馆等信息组织领域和计算机、信息构建等信息技术领域的学者都对社会化标注产生了浓厚的兴趣。目前,国内外对社会化标注的概念仍然没有一个统一的定义,国外社会化标注(social tagging)等同于 Folksonomy、Collaborative Tagging、Social Classification、Social Indexing,在我国称为大众分类法、公众分类法、民俗分类法和合作标注等,这主要是从标签的分类角度来研究的。但是标签是由用户对某个资源标注的,体现了用户、资源、标签是一个完整的系统,所以对 Folksonomy 的研究是建立在社会化标注系统的内容结构基础之上的。

3. 信息推荐

信息推荐指的是通过某种过滤技术将满足用户信息需求的信息通过某种方式推荐给相关的用户,更常见的是将最新的满足需求的信息推荐给用户^[6]。这里强调的是“最新的满足需求的信息”,而目前的状况却是用户面对的信息很多,但却不是他想要的。那么如何将用户想要的信息推荐给他,并且使他满意,这就是个性化的信息推荐服务的实质。

例如,在一个阳光明媚的周日上午,当用户端起一杯咖啡,打开笔记本电脑浏览当日的新闻时,网站展示给用户的已经不是那些与他毫不相关、让他毫无兴趣的“垃圾”新闻了,取而代之的是他一直在关注的房价拐点问题和他的股票池中重仓持有的中信证券的最新财报。相应的广告也不再是千篇一律的“减肥、美白”等让他看都不想看就想关掉的内容,取而代之的是他刚刚开始练习钢琴的一些培训班、销售钢琴的店铺打折消息等。所以在 Web 3.0 时代,如何根据用户的信息需求、兴趣模式、情境模式、用户之间的关系来推荐用户感兴趣的信息、服务、产品等,即如何满足用户个性化的需求,帮助用户在过载信息中快速发现真正所需的商品,提高用户黏性,促进信息点击和商品销售,是 Web 3.0 时代急需解决的问题^[7]。

在国外,SCI (Science Citation Index) 的学术论文检索结果显示信息推荐自 2000 年以来逐渐成为研究的热点,2004 年达到一个小高峰,以此为界分为两个阶段;2005 年有明显的回落,然后又逐年上升,直到 2009 年再次达到顶峰。根据研究内容得知,2004 年以前的研究主要是单纯的用户需求分析和信息内容组织基础上的推荐,即用户需求与信息内容的匹配;从 2005 年起,研究逐渐转向通过挖掘信息之间的关系、用户之间的关系甚至用户和信息的关系,进一步提高推荐的效率和效果^[8]。在国内,从中国知网学术期刊网络出版和万方数字知识服务平台的知识脉络分析可知,个性化信息推荐自从 2001 年开始被国内学者研究以来,逐渐引起了学者的注意,总体上,国内比国外的研究要稍晚一些,具体如图 1.2 所示。

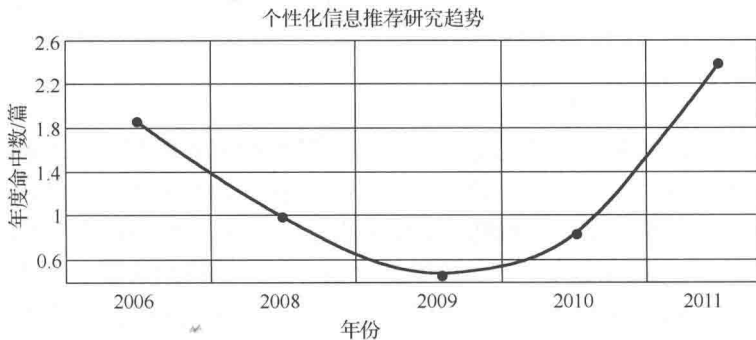


图 1.2 万方个性化信息推荐研究趋势

目前研究信息推荐的理论主要有复杂网络理论、用户行为学、社会化标注理论等。文献^[9]将复杂网络理论引入图书馆个性化信息服务的实际问题中,探讨在复杂网络理

论下个性化信息推荐服务的模型。文献[10]利用行为学理论中的效用、声誉和风险态度提出个性化信息推荐模型。文献[11]将众多用户与信息资源联系起来,形成用户与资源的关系网络,更好地进行信息推荐。还有利用标签可以进行用户聚类^[12]、资源聚类^[13],或者通过网络的共现建立用户聚类^[14]等。

目前研究信息推荐的理论主要是协同过滤、内容分析、关联规则。

(1) 协同过滤是最早提出的,研究最深入、商业应用最广泛。它服务的对象是个体,却利用了所有用户或商品的信息,并从中发现品味相近的用户或属性相近的商品,据此构造经过排序的推荐列表。

(2) 内容分析在最初只是协同推荐技术的延续与发展,它不需要用户对项目的任何评价意见,而是依据用户已经选择的商品内容信息计算用户之间的相似性,进而进行相应的推荐。目前很多网站的信息推荐都是通过分析商品的文本信息进行推荐的。

(3) 关联规则挖掘的根本目的是寻找商品销售记录的相关性,从而更好地知道销售策略的制定。挖掘物品在空间上的相关性,即发现同时被访问或购买的物品,据此为用户推荐。商品之间的关联规则可以分为空间关联和时间关联两种,时间关联又可以分为周期关联和顺序关联两种。

有学者提出一些相关方法,如矩阵分解、潜层语义分析、能力扩散与热传导等来提高算法的精度。还有学者提出混合推荐,就是综合运用以上多种推荐算法进行推荐,不同算法之间可以取长补短,但增加了一定的复杂性,而且推荐结果也不容易解释。

总体上,目前的信息推荐主要从基于商品和基于用户两个角度来进行个性化的信息推荐,基于商品主要是通过商品之间的关联度,来判断被用户购买过的某一商品,可能也会购买与这个商品相关联的其他商品,使用到的技术有协同过滤、内容分析、关联规则等。基于用户主要是通过判断用户之间的相似性,即判断单用户属于的群体,于是把群体的信息资源推荐给群内的单用户,使用到的技术主要是协同过滤,它主要是通过分析用户的购买记录历史,判断用户之间的相似性,如使用协同过滤技术的 Jason Olim 领导的音乐 CD 零售网站,对用户进行个性化的音乐专辑推荐。

4. 社会网络分析

面对 Web 3.0 的特征和核心思想,个性化信息推荐需要能够适应 Web 3.0 的特征而进行个性化的推荐。具体可以从以下三个方面加强:加强对用户隐性需求的理解,要根据用户的知识结构、心里倾向等方面进行挖掘,如基于情景的个性化推荐;加强信息的交流与共享,不同的门户平台可以整合在一起,让用户实现一站式服务;建立多用户兴趣模型,通过用户之间的信任关系,形成用户兴趣群。文献[15]提出了基于信任的社会化推荐系统的模型,文献[16]提出了基于信任、社会关系和语义分析的综合博客推荐方法,从以上三方面可以看出,Web 3.0 需要对用户进行各方面的挖掘,如用户的隐性知识、用户的兴趣、用户的关系等,而基于社会化网络分析的个性化信息推荐是目前信息推荐研究的热点问题。

科学网的研究结果显示,“与家庭成员相比,朋友会对人的行为和发展产生更深远的影响^[17]”,这说明社会关系对人的决定的影响是非常大的。它的核心是以用户关系为中心,建立用户关系网络模型,对关系结构进行分析,挖掘对信息推荐有益的知识,增加用户的滞留时间,提高用户黏性。

综上,在 Web 3.0 时代,人们不再满足由机器挖掘给用户的各种信息,而是要结合自身的偏好享用个性化的信息服务,无论是产品还是服务,都将为每个用户量身打造,信息世界将变得越来越聪明和智能。而目前的个性化信息推荐不能满足从用户关系的角度进行协同推荐的需求,所以本书是在 Web 3.0 的客观需求和目前的信息推荐理论无法满足需求的背景下,挖掘社会化标注系统中隐含的用户关系网络信息,利用社会网络分析理论对社会化标注系统中的个性化信息推荐进行研究分析。

1.2 研究目的与意义

1.2.1 研究目的

本书的研究目的主要就是构建基于社会化标注系统的个性化信息推荐模型,解决当前 Web 环境下个性化信息推荐所面临的信息推荐准确率低、数据稀疏、冷启动等诸多问题。

目前,基于社会化标注系统的个性化信息推荐的相关研究多是分散的、应用性的,而系统的、理论的方法较少。本书通过自组织理论详细分析了社会化标注系统的演化形式和结构,在此基础上,利用社会网络分析方法对社会化标注系统中的用户关系网络结构和凝聚子群进行分析,并构建个性化信息推荐的静态模型和动态模型,最后利用系统动力学对信息推荐模型进行动力学影响因素分析。

1.2.2 研究意义

本书的研究意义主要体现在理论与实践两个方面。

(1) 理论意义。从理论角度看,目前在社会化标注系统中研究个性化信息推荐处于刚刚起步阶段,至今仍没有形成一个统一的理论框架。本书在对社会化标注系统的形成、演进等进行系统科学的分析基础上,通过对社会化标注系统内的用户关系网络信息进行挖掘,发现用户关系网络的结构特征和凝聚子群;从系统学的角度构建个性化信息推荐的静态和动态模型。因此,本书的研究成果一方面将极大地丰富当前个性化信息推荐的理论体系,为个性化信息推荐研究的纵深拓展提供新的思路;另一方面,将在一定程度上完善信息服务的过程,进一步丰富信息服务的基本理论与方法,为信息服务创新提供借鉴,对推进情报学的学科发展具有重要的学科价值。

(2) 实践意义。从实践角度看,随着 Web 3.0 的发展,社会化标注系统中的信息急剧增长,急需个性化信息推荐的相关理论进行指导。利用社会化网络分析技术实现

个性化信息推荐能够有效缓解互联网带来的“信息泛滥”“信息迷航”等问题。对社会化标注系统用户来说,特别有利于他们获取相关的信息资源,以提高他们的信息获取能力;对各类社会化标注站点来说,可以有效提高自适应能力,能够动态满足系统用户的信息需求,从而提高站点的竞争力。总体上,本书的研究结果将对各类社会化标注系统的应用和发展起到良好的启示作用,并对各种电子商务运营商和用户起到重要的理论指导意义。

1.3 研究内容

在全面研究国内外相关学术成果的基础上,按照标准的“理论研究—实证研究”研究范式,应用了情报学、文献计量学、系统学、复杂网络等理论知识作为研究的理论基础,首先分析社会化标注系统的形成、演化过程;然后深入研究社会化标注系统中形成的用户关系的网络结构,进一步挖掘关系网络中蕴涵的个性化用户信息,先从静态和动态的角度构建个性化信息推荐模型,后从系统动力学的角度构建个性化信息推荐系统模型;最后利用豆瓣网的统计数据进行全面的实证研究。

主要分为五大部分,如下。

(1) 研究目前国内外社会化标注系统、个性化信息推荐两方面的研究进展,分析研究的热点与前沿问题,掌握在具体的研究中其所面临的不足,然后从不足处展开,形成本书所要研究的起点,即利用社会网络分析研究用户关系网络,作为分析个性化信息推荐的逻辑起点。

(2) 研究社会化标注系统的演化形式和耗散结构。对社会化标注系统中的用户、资源、标签和三者之间的关系进行分析,利用自组织理论研究社会化标注系统的演化机理,具体使用超循环和耗散结构分别分析系统中的用户、标签、资源的各自演化形式和多层级耗散结构。

(3) 分析社会化标注系统中的用户关系网络。首先,基于马斯洛需求理论分析了用户信息需求的形成,在信息需求的推动下,用户之间形成了错综复杂的关系,即社会化标注系统中用户关系网络的形成。然后,分析用户关系网络的结构特征,如网络密度、核心-边缘、中心性等;进一步,从用户关系网络中挖掘凝聚子群的各种信息偏好,从两个角度(群内和群际)对信息进行分析,发现个性化的信息。

(4) 构建社会化标注系统中的个性化信息推荐模型。从静态、动态两个角度分别构建社会化标注系统中的个性化信息推荐模型。其中,静态模型主要从系统学的角度展开,动态模型主要从信息推荐的过程角度展开;然后利用系统动力学绘制个性化信息推荐模型,具体由系统的因果关系和系统流程图两个内容构成。

(5) 对个性化信息推荐进行实证研究,首先从豆瓣网中抓取数据,然后对前面的系统自组织演化、用户关系网络结构分析、个性化信息推荐模型共三个方面进行整体实证研究。

1.4 研究方法与技术路线

1.4.1 研究方法

本书采用的研究方法主要如下。

(1) 文献分析与统计分析相结合。

本书采用了科学的文献计量方法——知识图谱,对大量的国外文献进行梳理与归纳,掌握了国外社会化标注系统研究和个性化信息推荐研究的知识基础、研究热点和研究前沿,了解了目前存在的问题;同时分析国内两个领域的研究现状和存在的问题,以此为依据,展开理论研究。在最后的实证阶段,收集豆瓣网的实际数据,用统计学的方法对调查结果进行整理与分析,获取研究所需的数据,进行实证分析。

(2) 整体分析与局部分析相结合。

本书首先从系统学的角度整体上分析了社会化标注系统的形成和演化,它是由用户、资源、标签和他们之间的关系构成的一个有机整体,遵循超循环和耗散的演化过程;然后从整体到局部,具体分析社会化标注系统中的用户关系网络,研究网络的结构特性和凝聚子群等,发现个性化的用户信息;最后由局部构成整体,构建社会化标注系统中的个性化信息推荐模型,全书逻辑关系较强,思维缜密,体现了由整体到局部再到整体的演绎与归纳思想。

(3) 定性分析与定量分析相结合。

在实际的研究工作中,定性分析与定量分析并不能完全划分开来,定性分析是定量分析的前提,没有定性的定量是一种盲目的、毫无价值的定量;定量分析是定性分析的具体化,它使定性更加科学、准确,促使定性分析得出广泛而深入的结论。本书首先在文献分析中,采用了定量分析(知识图谱)和定性分析相结合的方法对文献进行了梳理;然后在分析社会化标注系统的形成和演化机理时采用了定性分析方法(超循环和耗散结构),最后采用了定性定量相结合(系统学和系统动力学)的方法对个性化信息推荐动态模型进行构建。

(4) 理论分析与实证分析相结合。

理论分析法和实证分析法是两种相互联系又相互区别的研究方法。理论分析法是通过已有的相关理论对提出的问题进行逻辑上的分析,提出拟解决问题的办法;实证分析法是通过对客观存在的事物进行验证来概括和说明理论分析中提出的假设是否正确,主要说明研究对象“是什么”。本书在相关理论的支撑下,对社会化标注系统进行分析研究,提出个性化信息推荐模型,最后通过实证的方法进行验证。

1.4.2 技术路线

按照本书的研究内容,绘制相应的技术路线图,具体如图 1.3 所示。

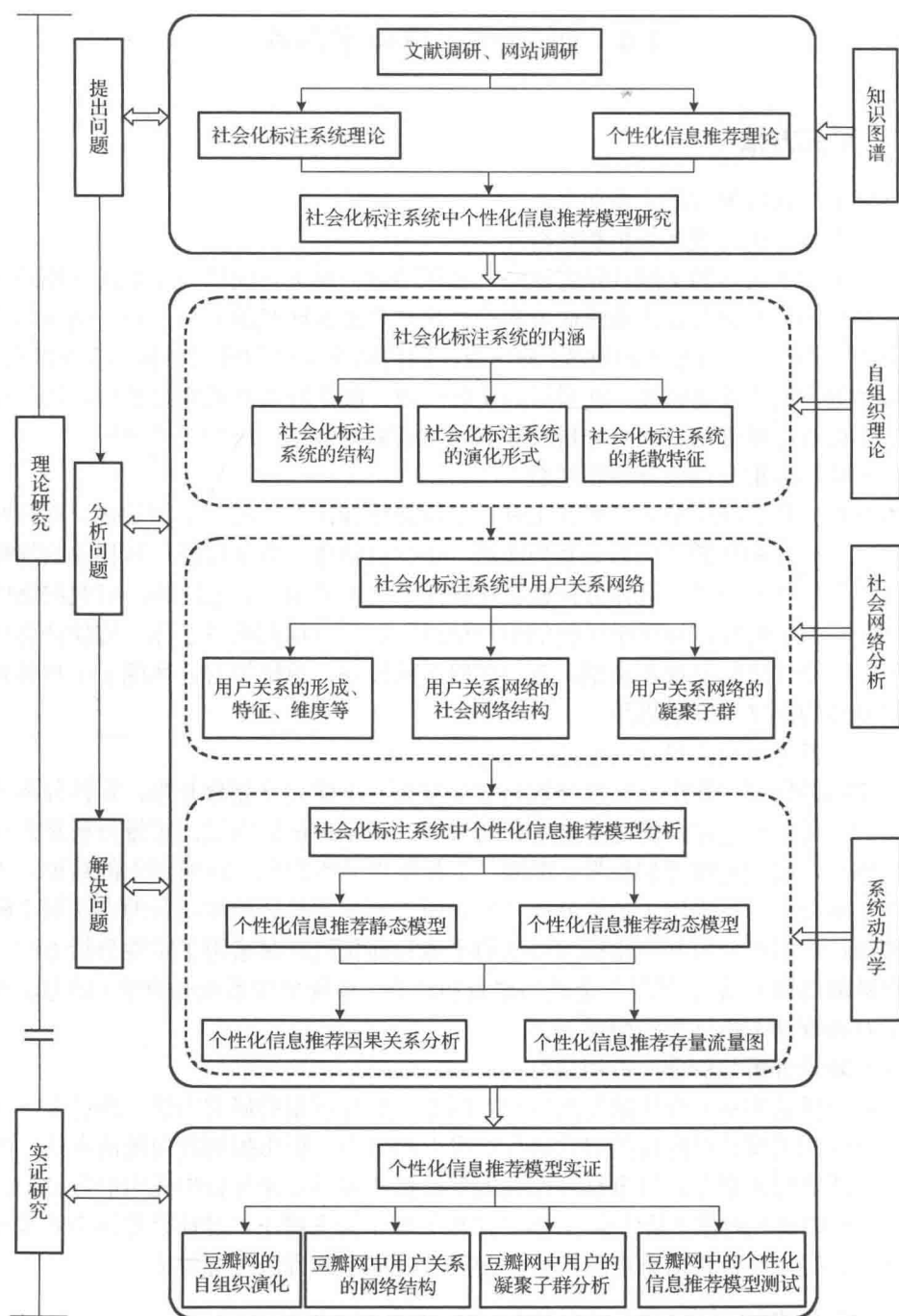


图 1.3 技术路线图