

● 孙雨生^{1,2,3}, 张恒¹, 朱礼军²

(1. 湖北工业大学经济与管理学院, 湖北 武汉 430068; 2. 中国科学技术信息研究所信息技术支持中心, 北京 100038; 3. 湖北工业大学区域产业生态发展协同创新中心, 湖北 武汉 430068)

国内兴趣图谱研究进展*

摘要: 文章介绍了兴趣图谱内涵、核心内容与研究框架, 并从基础研究、构建机理、典型应用、挑战及对策、发展趋势五方面揭示了国内兴趣图谱研究进展。

关键词: 兴趣图谱; 社交图谱; 研究进展

Abstract: This article introduces the connotation, core content and research framework of interest graph and reveals its research progress in China from fundamental research, building mechanism, typical applications, challenges and countermeasures as well as development trend.

Keywords: interest graph; social graph; research progress

伴随社会化媒体发展、网络社交日常化, 用户获取信息逐步碎片化、个性化、社交化^[1]、移动化, 期望低成本高效构建轻松虚拟社交关系、寻找兴趣相同(相似)个体(群体)形成兴趣社区^[2]深度交流, 以创新社交网络构建、管理、商业模式, 基于个性化推荐精准获取、自动组合信息并使其合理流动。其中, 作为陌生用户间表达兴趣新方式的兴趣图谱扮演重要角色并随类 Pinterest 网站爆发^[3-4]备受关注, 促使传统社交网站、互联网、电子商务巨头推出兴趣图谱产品(网站)^[4-5]、工具并取得巨大商业价值且呈增长趋势^[1], 如 Facebook 的 Like 按钮和社交搜索^[6](提供按兴趣划分用户、话题等列表, 通过兴趣链接分享精准推荐信息)、Twitter 的订阅列表(基于兴趣创建关注列表, 发现兴趣相同用户并分享), 腾讯兴趣图谱 API^[7-9]、百度个人中心兴趣图谱功能^[10]、Google + 的 Circle 和 Community 等(可将社交图谱成员拖入兴趣图谱)^[6], 蘑菇街、美丽说(图片分享型社会化电子商务)^[3,5]、堆糖网(纯视觉社交, 按兴趣需求分类图片互联用户, 按兴趣、关系、幻想、交易等分类社区)^[5]。

本文介绍兴趣图谱内涵、核心内容及研究框架, 并从基础研究、构建机理、典型应用、挑战及对策、发展趋势五方面揭示国内兴趣图谱研究进展。

* 本文为国家自然科学基金资助项目“基于语义网格的数字图书馆个性化推荐模型研究”(项目编号: 71003032), 中国博士后科学基金资助项目“基于领域本体的数字图书馆用户兴趣建模研究”(项目编号: 2014M560107) 和湖北省人文社会科学重点研究基地——湖北农村社会管理创新研究中心基金项目“农村信息化知识图谱构建研究”(项目编号: HNSKY1312) 的成果之一。

1 兴趣图谱简介

1.1 定义及内涵

兴趣图谱概念于 2010 年 12 月正式提出并逐渐受互联网界关注^[1,11-12], 是 Web 3.0 时代用户网络行为(搜索、订阅、关注、购买、表达^[7,13-14])无意识形成的产物^[1], 是基于网络用户间共同兴趣、由个体兴趣关系(人际关系简单^[15])构成的表明用户兴趣集^[1]、兴趣相同或相似用户集的图谱^[3,5,13-17]。其遵循兴趣共享(用户间无需相识)理念^[1,3,5,11-14,18-20], 基于用户网络行为、用户生成内容等采集用户数据^[5,16]、提取兴趣^[1,12,21]动态反映用户兴趣关联度及喜爱度变化^[9,16]、拓展用户社交关系(基于兴趣^[14,16]互联用户^[9]), 重塑社交网络以用户为中心功能特质^[4,20]、创造商业价值、提升信息服务效率、主动性及可信赖性^[9]。

究其内涵, 兴趣图谱是由兴趣节点、用户构成^[12], 以兴趣节点为核心(周围环绕关联用户^[1,3,11-12]), 表示用户与兴趣间、用户与兴趣与用户间关系虚拟网络^[1,11,12], 是标识彰显用户特征的兴趣集及兴趣度^[7]并基于兴趣关联^[8]用户的弱关系社交模式、虚拟^[12,15]社交关系建构及管理方式(遵循用户意愿^[22]并随其兴趣动态更新^[16])。

1.2 核心内容与研究框架

主要研究兴趣图谱的表示与存储、初始化(兴趣来源及采集、预处理, 主要是数据挖掘)、进化^[8](学习与更新)、应用(用户聚类、信息推荐^[17])等^[1], 现有研究框架见图 1。

2 国内兴趣图谱研究进展

为全面掌握国内兴趣图谱研究状况, 笔者以 CNKI、

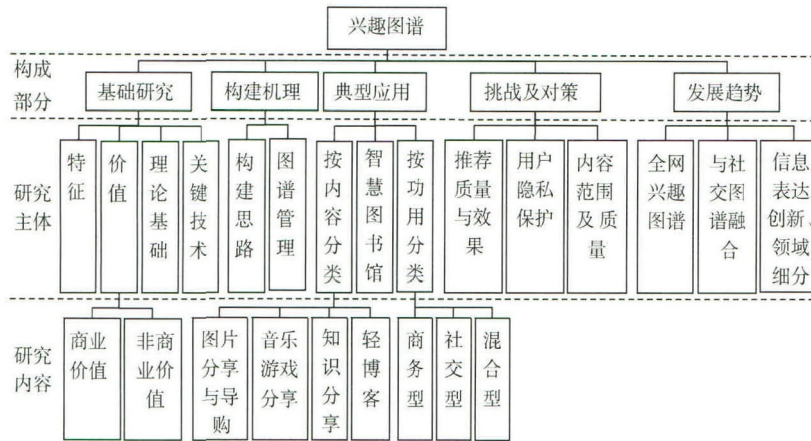


图1 兴趣图谱研究框架

万方的期刊论文库、硕博论文库、会议论文库、报纸库，维普为信息源，采用关键词在题名中检索方式进行主题信息检索。截止到2016年3月25日，以CNKI为信息源，以兴趣图谱为关键词在篇名中检索，共得期刊论文14篇，硕博论文6篇，报纸文献4篇，用相同方法在维普中检索，共得期刊论文20篇（与CNKI对比，发现5篇新文献），用相同方法在万方中检索，共得期刊论文13篇，硕博论文3篇（与CNKI、维普对比，发现1篇新文献）。其中，国内兴趣图谱最早出现在马建国2005年撰写的《广播型网格的用户兴趣图谱》中。

纵观现有成果，国内兴趣图谱研究处于起步阶段且集中在业界（多描述现象、展望前景，缺乏定量研究）^[4]；学术研究集中在基础研究、构建机理、典型应用、挑战及对策、发展趋势等方面，应用研究集中在电子商务、网络社交、智慧图书馆等领域。

2.1 基础研究

2.1.1 特征 兴趣图谱是多个社交图谱重叠部分，特征见表1^[1,11]。

2.1.2 价值

1) 商业价值。基于社会化媒体、网络社交、数据挖掘精准营销是未来商业重要形式^[20]。孙雨生^[1]、徐扬^[4]、苏慧^[5]、廖贤敏^[14]、李奕^[20]认为兴趣图谱集成社交网络中隐含但明确用户、商家、商品信息^[1]，获取并通过社交网络分享兴趣（在感兴趣商品上和已消费用户交流），激发其正面情绪、挖掘其潜在需求并提供个性化服务，在扩展社交网络形态同时改善用户体验（基于消费行为（消费种类（反映兴趣）和评价（反映品味）^[14]）计算用户兴趣相似度以分类用户、协同推荐商品，发现用户兴趣社区或朋友圈^[3]、聚类相关商家^[1]、降低搜索代价和决策成本^[3,5,14]、参与积极性（即粘度、购买欲望^[1]）、消费主动性（兴趣相同用户自发激发购物需求^[5,14]）、交易稳定性（解决信誉危机）和行动（投入时间、精力、金钱）

高效性^[16]（变现能力）；基于自发维护的社会关系、兴趣等决策，在社会化分享中口碑营销（基于意见领袖^[3,23]、消费达人^[19]），将商家对用户一对一维护变为对社区一对多维护^[9,16]、在朋友圈间快速传播以降低营销成本、满足商家针对性（抓住并基于主要兴趣精准营销）、高品质流量（实现增值）需求^[3]，通过推荐对商家商品感兴趣用户、合作商家，以便网络联盟、优势互补、资源配置优化^[1,14]；最终降低社会商业成本。

表1 兴趣图谱特征

项目	说明
本质	弱社交关系 ^[5,20] 功能图谱（基于兴趣 ^[5] 交换传播 ^[3] 引发共同活动，简化单向兴趣关注 ^[3,18] （用户间保持友好距离）、转发 ^[16] ）
核心与宗旨	核心是发现兴趣、兴趣内容，主体是兴趣，单位是兴趣点 ^[16] ；旨在通过多维兴趣定义用户 ^[6] ，基于低门槛关系纽带、交流方式（兴趣分享、标签化沟通） ^[5] 重聚、沉淀、拓展并超越社交关系（基于兴趣聚集、互联用户 ^[4,7,15-16,20,24] ），形成大用户量、强互动性 ^[17] 兴趣社区，基于用户分类、协同推荐协调用户需求 ^[5,16-17] 和网站内容关系 ^[5,16-17]
信息环境	基于兴趣、默认公开 ^[3,18] 社交信息环境，兴趣圈用户可看到公开 ^[5,22,25] 信息
用户需求层次	自我实现（基于并拓展兴趣，追求个人品质、品味 ^[16] ，重构、发现并获取个性信息） ^[1] ；社会需求（人际交往、环境适应、精神及情感满足） ^[16] ，关注将来个体及其社交圈 ^[5,17]
影响因素及稳定性	一般；群体兴趣图谱基本稳定，单用户兴趣图谱易变 ^[7,8]
直接作用对象	先事物后用户 ^[7-8] ，表明用户兴趣 ^[11]
完成者	电子商务网站：Pinterest、知美网；堆糖网 ^[5]
社交网络（站）初建及维护	通过兴趣相同用户关注同一 ^[14] 事物 ^[11] ，激发用户快速自发构建（按兴趣分类聚合信息）；通过兴趣吸引新用户发现、分享兴趣及潜在意愿，低成本维持轻松弱关系（只与兴趣相似用户交流，只关注感兴趣事物且仅分享兴趣时产生轻量社交）、用户活跃度（自选并常用所选网站自创、接收内容 ^[16] ）、内容高质性（持续拓展、个性化（单用户不同兴趣分别和相应感兴趣用户交流）、无噪音 ^[16] ）
变现能力	更强，直接将用户兴趣引入购买行动
口碑传播	市场营销易入口，用户话语权更大，传播成本更低 ^[13]
邻近匹配	有效缓解推荐中用户评分稀疏问题且精度高 ^[13]

2) 非商业价值。简化并提升^[16]兴趣发现、分类、分享、管理、传播效率,基于兴趣分享虚拟化,主动拓展互补结构网络下^[11 21]高质可信社交关系^[1],形成强人际互动兴趣社区^[17],聚类兴趣广泛性、多样性导致的碎片信息^[16]并按用户(潜在^[7])兴趣^[5 16]动态推荐,提升信息服务效率^[13](高精度、低成本^[11 25],内容生成积极及创新性^[4 20],减少用户注册、登录、关注^[15]、信息浏览量)、流动合理性、内容价值性(知识含量^[1 16]),缩短用户认知、认可事物过程^[13],提升用户体验^[3]。

2.1.3 理论基础 主要涉及六度分割、马克·格拉诺维特弱关系(除关系稳定但传播范围有限的强关系外,存在更广泛但肤浅社会认知的弱关系(一方可能掌握很多对方并不了解信息,是社交圈间信息流动主要渠道)^[1 4]、马斯洛需求层次、用户网络访问群聚性等理论^[1],应用涉及传播学、社会化营销^[11]、认知心理学(人脑对知识认知是概念语义网络,表达人类认知中知识存储结构、传递过程,节点强度代表认知深度,节点间语义相关度代表认知范围^[10])等理论^[16]。此外,陈媛媛^[10]提出兴趣图谱进化理论基础有本体、扩散激活及社会网络分析。

2.1.4 关键技术 主要涉及用户兴趣建模^[1]、社交网络^[10]、标签传播算法^[21](挖掘社交关系中圈子)、机器学习(词项、分词、分类、聚类)、数据挖掘(基于数据和关系^[14]挖掘兴趣)^[25]、数据集成^[8]、社会化标签数据库^[1]、本体^[8]、智能代理^[1 26]、数据更新^[8](兴趣图谱演化与反馈)、信息和兴趣图谱匹配^[25]等技术。

2.2 构建机理

现有研究很少,多为概念性构建^[17]。

2.2.1 构建思路 张召^[6 25]提出基于三分图模型(表示为用户(账号)、论题(语义标注文本)、词项(用自然语言工具获取,含语义,与文本多对多表示关系)、用户与论题关联边、论题与词项关联边五元组)捕获动态社交网络结构、(偏好相似)用户生成内容(含历史偏好),用带重启随机行走算法捕获三分图中节点间直接、间接关联以度量节点相关性,构建论坛活跃用户内容兴趣图谱、协同兴趣图谱;分别基于三分图模型(表示为用户、微博信息、词项、社交关系(关注者、被关注者))、用户关注列表(捕获用户生成内容及用户间关注关系,用简单启发式规则算法(双向关注,交互次数)构建微博潜水用户内容兴趣图谱并存入兴趣图谱库,用带重启随机行走算法构建社交上下文图模型,生成用户兴趣图谱;基于三分图构建星型图,在星型图上添加查询点构建并基于查询图捕获专家、专家兴趣图谱;陈媛媛^[7]、刘平峰^[8]依据基于全网数据用户兴趣图谱本体模型和 Gravity 公司新闻

用户兴趣图谱构建方法,获取全网数据并抽取用户兴趣概念、计算兴趣度生成局部兴趣图谱,基于本体映射集成兴趣概念及关系、兴趣度和规则生成并进化全局兴趣图谱;陈媛媛^[10]基于语义词典与关联开放数据集构建兴趣概念内涵与外延,借助本体思想与 dmoz ODP 构建兴趣概念关系、兴趣度及规则以形成用户兴趣领域本体,基于用户在在线社会网络中关系(表示为无向图,含用户、用户间连接及其产生时间)及其兴趣图谱本体构建用户兴趣图谱;丁绪武^[3 17]借鉴个性化服务中混合用户建模机制文本表达兴趣,结合社会网络分析构建兴趣图谱:获取用户数据、提取整合用户兴趣(提取兴趣关键词形成主题兴趣并分类、赋权重,基于兴趣分类器、用户(活跃度较高、粉丝量较多)标注兴趣类别来分类文本兴趣),构建用户兴趣模型(用 ICTCLAS 对文本兴趣分词、标注,发现并细分用户兴趣集构建用户兴趣细分树,用用户兴趣值算法得出用户兴趣模型)并引入社会网络关系值分析社会网络进行完善,基于用户及其兴趣间边构建边数据 CSV 文件导入 Gephi 聚类、调整后形成兴趣图谱;张俊林^[21 24]将用户社交圈共性信息融入微博中发布内容较少、未标注标签用户兴趣模型构建兴趣图谱,提出由实时抽取系统(基于并行语义处理单元对 Feed 流队列^[10 18]中实时发布微博进行分词(含命名实体识别)、焦点词抽取、文本分类)和离线挖掘系统(挖掘实时抽取系统中用户兴趣得出用户真正兴趣)组成的兴趣图谱构建系统;戴盈^[5]梳理读者阅读兴趣基因并构建 O2O 阅读兴趣图谱;闫霏^[9]构建动态描述用户信息行为和关注主题的图书馆用户兴趣图谱:分析较受关注主题基本特征,为其建立并基于集成动态关联图模型生成内容兴趣图谱、协同兴趣图谱;马建国^[26]将媒体网站结构转为本体结构,基于本体统计用户对信息兴趣度以构建媒体网站用户兴趣图谱,并直接基于媒体信息构建媒体栏目用户兴趣图谱。

2.2.2 图谱管理

1) 图谱表示。张召^[6 25]分别用与用户最亲近的词项集(含兴趣度且兴趣度和为1)、用户集表示内容兴趣图谱、协同兴趣图谱,用专家集及基于用户查询、按专家熟悉度排序词项集表示专家兴趣图谱;丁绪武^[3 17]提出细分树向量空间模型表示法;陈媛媛^[7]、苏慧^[5]提出用户模型(树形兴趣概念层次模型,描述用户兴趣及兴趣间关系)表示法;张俊林^[24]提出话题表示法(基于随机抽样去重后微博,用 LDA 算法挖掘话题级用户兴趣训练形成全局话题级兴趣向量,基于用户所发微博构建用户话题级兴趣向量);马建国^[26]提出矩形坐标统计图表示法(纵坐标表示兴趣度、颜色矩形表示兴趣且兴趣度从左到右依次降低)。

2) 图谱初始化。兴趣来源方面,陈媛媛^[7,10]、刘平峰^[8]提出用户全网(社交网站(关系、发布、转发、浏览、收藏、信息停留时间、评论、点赞)^[10]、博客、社会化标签、电商数据库(浏览与消费)、开放关联数据集、本体模型、语义词典等)数据;张召^[6,25]、张俊林^[21,24]提出论坛、微博(自标签^[21]、原发、转发、阅读、关注、收藏^[24]);丁绪武^[3,17]提出美丽说、用户反馈;廖贤敏^[14]、苏慧^[5]提出业务对象、消费历史(商品种类、喜好度)、商品评分;孙雨生^[1]、徐扬^[4]提出用户提交、系统学习^[1]、用户行为(含用户构成、使用情况、使用偏好^[4]);张薇^[12]提出读者交互(浏览、转发、分享、评论)、共享、采纳、争执行为;闫霏^[9]提出图书馆用户信息行为。

兴趣采集方面,张薇^[12]向社交网络平台投入话题并分析UGC采集读者兴趣,用近场无线通信、智能情境感知、社会化知识发现及体验技术发掘读者行为轨迹和兴趣矩阵;陈媛媛^[7]、刘平峰^[8]通过面向API适配器采集并抽取用户兴趣;丁绪武^[3,17]用基于HttpClient、HtmlParser组建网络爬虫采集兴趣;戴盈^[27]线上基于融合社交互动的电子商务平台、线下基于实体店体验活动形成稳定读者群并挖掘读者信息、阅读及购书兴趣;闫霏^[9]用数据挖掘技术挖掘用户兴趣;张召^[6,25]、张俊林^[21]采集活跃、潜水用户(组)间社交关系及UGC(关注列表)。

兴趣预处理方面,张俊林^[21]通过数据整合MapReduce、标签传播算法MapReduce更新用户标签集,挖掘用户社交圈、UGC、潜在兴趣(词项及标签);丁绪武^[3,17]用广度优先搜索策略获取并从个人标签、关注对象和数量、粉丝数量、分享内容数量和数量、喜欢内容和数量、关注杂志内容和数量等提取用户个体兴趣并分析分享数超50、宝贝元素相同或风格相似杂志提取用户群体兴趣,用TF-IDF算法计算用户个体、群体兴趣度(基于个体兴趣),引入社会关系值、用户重要度并加权求和;廖贤敏^[14]、苏慧^[5]进行商品分类统计、空值处理等以挖掘用户兴趣、消费习惯,计算兴趣相似度聚类、连接兴趣及消费行为相近(相同)用户;陈媛媛^[10]基于用户消费、社交网络数据(新浪、腾讯微博)计算用户对兴趣概念兴趣度及兴趣产生时间,用ICTCLAS对商品名分词获取其标签。

3) 图谱进化。陈媛媛^[10]研究基于在线社交网络用户兴趣扩散机理(含用户兴趣扩散概率计算(基于离散时间模型、用户网络空间影响力)、用户网络行为模型(基于扩散激活理论获取兴趣扩散概率,依据用户社会网络行为及行为发生时间构建)、用户兴趣扩散波纹模型(依据用户兴趣在不同社交网络间扩散概率))、基于兴趣图谱本体用户兴趣扩散模型(基于扩散激活机制、方法结合

本体模型语义结构进化兴趣图谱,包括扩散激活步骤(激活由节点向邻近节点传递,传递范围与可能性由节点权重、激活强度与节点间语义距离决定;激活扩散是迭代过程,每次迭代先更新与当前节点直接相连节点兴趣度再检查是否达到终止条件,直至所有节点被激活)、约束条件(控制扩散进程与结果)、模型(研究扩散类型、过程),先设定扩散激活函数闭值与广度扩散节点数,再依次深度扩散、广度扩散及多向链接扩散),提出融合个体(基于用户电商网站消费行为构建用户消费行为模型,借助语义词典与开放关联数据集,综合商品与商品属性或标签关联商品与用户兴趣概念)一时间(基于用户电商网站行为及时间数据构建用户兴趣指数衰减模型)一空间(综合用户社交网络中兴趣扩散与基于社会网络分析法所得用户在线社交网络拓扑结构、所花时间计算用户在网络空间影响力;用语义词典及兴趣本体关联用户发布信息与兴趣概念)的兴趣图谱进化机制;陈媛媛^[7]、刘平峰^[8]基于情境感知动态获取兴趣,基于时间序列预测、图结构链路预测、兴趣社区、复杂网络演化兴趣图谱,基于用户兴趣采纳机制、购买行为反馈进化兴趣图谱;张薇^[12]、戴盈^[27]基于用户采纳、搜索阅读行为进化兴趣图谱;孙雨生^[1]基于用户行为、评价等进化兴趣图谱;张俊林^[24]通过全局模型和个人模型整体更新方式进化话题级兴趣图谱。

2.3 典型应用

主要用兴趣图谱社交工具和网站进行个性化推荐、社交流量提升、情绪激发^[17]、需求发现^[28],典型应用见表2^[1,5,15]。

2.3.1 按内容分类

1) 图片分享与导购。多用于电子商务导购平台(图片链接电子商务平台^[3,23],单击图片或其附近提示按钮即可转至购物网站^[1],吸引商家进行社会化营销)^[4]:整合标签化图片表达内容提供灵感^[16]、表述观点,降低用户参与门槛(靠直觉与感性思考欣赏,审视个人兴趣;基于白板采集、收藏、分类、检索、优化感兴趣其他网站^[2]图片构成消费兴趣集、展现品位^[4],基于图片分享构建用户间传播、人与信息关系以发现、共享、管理兴趣,基于兴趣互联用户^[4,7,15,20,24]形成并拓展弱社交关系^[5,14,16,20];瀑布流呈现图片以改善视觉冲击^[15-16,23,29]、网购体验^[3]。

2) 音乐游戏分享^[16]。通过客户端记录并分析用户行为,向其主页添加元数据并构建兴趣图谱,推荐感兴趣对象或兴趣相似用户以构建分享和推荐小组并提供表达态度按钮方便发现、收集、共享兴趣^[1,5]。

3) 知识分享。首先基于兴趣图谱按主题分类问题、答案以便用户关注、订阅,提高知识获取与分享效率;其次设置问题搜索框并显示动态,以便用户发表、评价答案、

表2 兴趣图谱应用

类型	基于兴趣图谱优化社交网站	基于兴趣图谱新建社交网站
内容分类	图片分享	SmugMug, Flickr ^[16] , 500px, Pikeo, Friendsheet; 全景网 Pinterest ^[3-4, 16-18, 20, 24] 、Instagram ^[18, 20] ; 花瓣网 ^[3, 23, 29] 、堆糖网 ^[23] 、发现啦 ^[19] 、码图网、拼图网、topit.me、腾讯读图知天下、嘀咕网、爱采图、开心集品、我喜欢网、啪啪网、百度爱乐活、闪播网 ^[29]
	图片导购	Polyvore; 逛 ^[3, 23, 29] NatureBox; 美丽说、蘑菇街 ^[3-4, 19-20, 23, 28] 、凡客达人 ^[3] 、一淘发现 ^[4] 、hao123 逛街、哇哦网 ^[3] 、爱逛街 ^[3-4, 14] 、辣妈说、拼范网、领秀街、顽兔社区、拖拉网、新鲜网、优容网、布兜、人人逛街、知美网 ^[19-20, 23, 29] 、迷尚网、粉团网、积木网、零食控、爱物网 ^[23, 29]
	音乐游戏分享	Last.fm ^[3, 17-18, 20] , Pandora ^[6] ; 豆瓣音乐、虾米音乐 ^[18, 20] turntable.fm ^[18]
	知识分享	Google 学术 Quora ^[17-18] , Thumb; 豆瓣 ^[4, 14] 、知乎 ^[18] 、果壳
	轻博客、文字分享	Twitter, Tumblr ^[3, 17] , my6sense; 豆瓣网 ^[17] 、新浪轻博、微博 ^[4, 13-14, 28] Noovo; 人人网人人小站、网易 Lofter、凤凰快博 ^[28] 、点点网轻博客 ^[3, 17, 28] 、盛大推他网、6V 轻博客 ^[28]
	视频分享	Youtube ^[16]
社会化阅读 (定制、智能推荐)	StumbleUpon、Google Reader; 腾讯爱看 ^[22]	
移动类		Playup、谷歌新鲜汇、Socialeam、Viddy ^[18] ; 烙饼客、短趣、米吧、方维 ^[4] 兴趣图谱
其他	Facebook, Google +、Hunch; 美味书签	Genico, Traplt, Trippy, Foodspotting, Spling, Fitocracy, Gravity; 石榴网、风格网、花生共和、优推网、粉丝团、郑州逛街、爱库网、圈圈网、维锐网、合拍网、百度搜索风云榜

与感兴趣用户互动^[1, 5]。

4) 轻博客。基于兴趣图谱类社交新媒体, 按兴趣标签、话题分类帖子 (按时间排序)、用户以构建垂直频道, 显示用户关注标签动态并基于兴趣图谱优化内容质量^[24], 用户间通过关注和评论相同博主建立社交关系^[28]、维护同一主题^[3, 7]产生互动, 提升用户积极性、活跃度并拓展社交关系^[28]。

2.3.2 按功用 (盈利性) 分类

1) 商务型。主要用于电子商务导购, 涉及个性化推荐、口碑营销、广告精准定位与展示^[24]等。廖贤敏^[14]、苏慧^[5]认为兴趣图谱类社会化电子商务基于用户兴趣开展商业活动, 创新社交网络构建模式: 基于兴趣图谱构建用户自维护社区、推荐兴趣相关社区 (基于商品查找, 与个人、群体实时沟通)、可能感兴趣社区 (随意加入退出或自建并邀请隐性好友加入) 增强用户与用户、商家 O2O 互动, 以基于兴趣精准推荐、营销; 戴盈^[27]用兴趣图谱实时洞悉读者兴趣、潜在需求, 不定期精准发送其所需图书或服务促销信息以及时刺激读者读书时、产生购买需求前、对图书最感兴趣时形成购买冲动并最终购买; 刘平峰^[8]、陈媛媛^[7]设计云计算环境下基于兴趣图谱、本体电子商务智能推荐系统, 分 4 个子系统 (云计算资源调度、兴趣图谱生成、兴趣图谱动态演化与反馈、基于兴趣图谱个性化推荐) 和外部接口层 (负责推荐服务器与用户、云计算资源调度系统交互, 提供兴趣图谱本体库、数

据库管理界面), 云计算资源调度系统基于依托 Hadoop 的混合云平台 (分 IaaS、PaaS (用数据云层存储全网数据、兴趣图谱库、本体库等), 支持兴趣图谱挖掘及推荐) 进行资源管理与分配、监控、备份、安全等, 基于兴趣图谱个性化推荐系统提供兴趣图谱解析器、基于兴趣用户相似度及权值计算器、用户—兴趣—商品概念三部曲 (含兴趣度) 构建器、用户商品属性偏好挖掘器、情境感知接收器 (过滤推荐结果)、推荐引擎和推荐模式^[8], 流程为: 构建三部曲计算用户语义相似度生成候选兴趣集, 用贝叶斯分类算法基于兴趣—商品概念三部曲推荐商品概念集, 再基于用户对商品概念集属性偏好推荐商品^[7-8]。

兴趣图谱口碑营销多用半广播机制引领消费群意见 (多搜索时发现相关信息)^[13], 影响社会化电子商务产品网络口碑 (传播路径、传播量、散播率)。张飞相^[13]基于 Ucinet, 用 3 种方法测度微博用户兴趣图谱结构中心性, 通过描述中心性、聚类节点权力类型分析基于兴趣图谱口碑营销网络: 从所关注网络节点联结处采集关注关系存入邻接矩阵, 分析个体、平台口碑传播中心性; 徐扬^[4]认为蘑菇街口碑营销特色是基于微博测试用户星座、性格、爱情等并自动生成微博动态; 张佳琦^[16]、丁绪武^[3]认为 Pinterest 基于视觉化内容聚合、兴趣图谱营销品牌: 品牌生动展示 (引导用户了解分享品牌信息并购买), 品牌导流模式构建 (基于全网图片转存与品牌互动模式传递品牌信息; 先兴趣圈内分享快速建立好感再经导流通道购

买),品牌亲密感塑造(在兴趣圈内故事化、图形化多维直观传播品牌形象,提升信赖度以便消费群互动、接受),品牌深度沟通(兴趣设定满足用户线上兴趣、指导其线下生活与兴趣探索;关注并鼓励、引导意见领袖进行线下活动;使其他用户学习、接收线下建议,分享传播经验并忠诚品牌,提高平台流量;帮助用户了解感兴趣却不熟悉品牌,产生模糊需求)。

2) 社交型。强调兴趣采集分享。张佳琦^[16]认为兴趣图谱类社交网站单纯以兴趣分享为目的且用户近似匿名、自主体验^[23]轻社交乐趣,基于兴趣推荐好友;王守崑^[6]认为基于兴趣图谱个性化推荐构建融合兴趣图谱、社交图谱的用户行为模型用于统一跨域推荐;张召^[25]用内容推荐、内容兴趣图谱、推荐对象驱动实时推荐算法动态推荐论坛强时效线上对象,用协同过滤、协同兴趣图谱、推荐对象或用户驱动推荐算法推荐论坛外稳定线下对象:收集用户行为、发现用户兴趣图谱、发现推荐对象、匹配兴趣图谱和推荐对象;丁绪武^[3]用K-means算法、基于R语言发现社区及用户兴趣集内部联系。

3) 混合型。丁绪武^[3]认为兴趣图谱类第三方社会化电子商务平台融合社交和商务(用户根据兴趣参与小组或栏目,准确收集所需商品信息,结识兴趣相同或相近朋友),需定位并针对合适用户群提升其基于兴趣图谱社交、购物、分享交流等体验,提高交易转化率(以标签代表用户兴趣,通过标签算法预测用户兴趣并用于社会化营销)。

2.3.3 智慧图书馆 张薇^[12]提出基于兴趣图谱智慧图书馆 SoloMo 模式框架及构建策略:基于图书馆移动社交 SoloMo 平台汇集数字资源、图书馆知识,收集并分析读者与图书馆、读者间行为信息,基于阅读同种(相似)知识读者兴趣相关度构建用户行为模型、兴趣图谱并转为社交图谱,以改善读者知识建构,读者间知识交互、共享、进步,强化读者与读者、图书馆间关系;根据用户注册、互动、移动签到信息,聚类兴趣、生活习惯(规律)相似读者构建兴趣矩阵、高粘性且自维护兴趣社区,基于读者兴趣图谱匹配兴趣社区,用RSS协同个性化推荐知识并随兴趣图谱进化动态更新,以发现读者潜在兴趣,促进兴趣相似读者互动,提升信息获取效率,最终联手O2O模式提供碎片化、移动阅读服务并嵌入读者线下活动;闫霏^[8]提出基于兴趣图谱图书馆泛在服务模式,涉及服务方式(兴趣导向下嵌入式立体化服务:面对面咨询、分类定制、智能推送、个人帮助,嵌入参考咨询工具)、服务内容、服务途径(移动终端、通讯设备、网络、电视等),设计基于兴趣图谱图书馆泛在服务推荐算法(用Apriori算法发现关联规则算法中频繁2项集以计算用户关

注主题间关联度,构建高频信息行为关联图聚类用户信息行为,得到用户感兴趣资源集,基于本体计算用户感兴趣资源间相似度并推荐)。

此外,国琳^[30]通过社会网络用户标注数据类型分析其关注领域及兴趣度,通过幂函数拟合生成兴趣分布曲线,通过合并相似兴趣突出重点关注领域,最终发现权威专家、兴趣领域专家并提出甄别伪专家算法;陈媛媛^[10]基于兴趣图谱多粒度个性化推荐并用准确率、精确率、召回率评价效果。

2.4 挑战及对策

2.4.1 推荐质量与效果 张亚明^[2]整合云计算(基于Hadoop、MapReduce提升推荐可扩展性、并行性)、商务推荐技术(改进协同过滤算法降低计算复杂度),基于兴趣图谱、用户行为、数据挖掘优化云计算环境下网络社区海量信息推荐深度、精度、规模、实时性,用兴趣划观点理论缓解因社区用户兴趣集中导致的评分稀疏问题;王守崑^[6]提出通过增强学习或决策树法在用户展示兴趣分量较少时引导其在兴趣图谱中自行发掘,基于兴趣图谱提高跨域推荐准确率、召回率,部分缓解数据稀疏问题;陈媛媛^[7]提出改善协同推荐算法、基于全网兴趣图谱法。

2.4.2 用户隐私保护 孙雨生^[1]认为信息服务均不同程度涉及用户隐私,兴趣图谱应用需提供选择可追踪内容功能以使用户平衡信息获取成本和兴趣泄露;陈媛媛^[7]、瞿敬渤^[11]提出创建全网兴趣图谱,推荐信息时应配套用户隐私保护机制。

2.4.3 内容范围与质量 孙雨生^[1]、谢璞^[19]、孔离^[23]认为现有兴趣图谱应用内容过于女性化、娱乐化,应增加男性化、日常化内容并优化内容布局、分类、呈现;通过艺术家、专家、达人等提升内容原创性,通过进化兴趣图谱、降低内容创造与分享门槛提高内容时效性、分享率,通过语义技术改善搜索引擎智能性提升结果准确性,通过改善推荐引擎用户信息、行为分析能力满足用户潜在需求,最终提升社交氛围、用户粘度、内容质量。此外,廖贤敏^[14]、苏慧^[5]提出平台方、社区创建者、管理者应监控兴趣社区内信息,强制退出恶意广告者或封号;丁绪武^[3]认为图片质量影响社会化电子商务用户消费冲动、购买机率,应奖励乐于详细真实分享评论的用户。

2.5 发展趋势

2.5.1 开放全网兴趣图谱 王守崑^[6]认为兴趣图谱关注用户通过多维兴趣定义自我、关系且都在垂直兴趣领域有限开放;陈媛媛^[7]基于全网数据构建全网兴趣图谱并用于电子商务推荐。

2.5.2 与社交图谱融合 廖贤敏^[14]认为兴趣图谱多与社交图谱重合、包含且随用户人际关系网络发展而融合,基

于两者的社交网络产品趋于融合(基于相同兴趣的社交关系融入社交图谱^[5,16]提升UGC质量、分享率,改善用户同质性、粘性,社交图谱巩固并拓展基于兴趣交流互动^[4]),社交网站通过选择两者结合比例^[4,16]平衡内容贡献与用户基数、优化网站环境;丁绪武^[3,17]认为社会化电子商务平台融合两者扩充社交关系、将商业焦点转向兴趣社区,进行商品或服务信息精准传播、个性推荐、多平台流量共享与互动,实现平台商业价值;王守崑^[10]认为两者融合应基于用户行为建模。

2.5.3 信息表达创新、领域细分 前者体现为国内类Pinterest网站大多用白板分类兴趣,后者体现为领域性基于兴趣图谱信息服务^[20],张俊林^[24]提出构建垂直性用户兴趣模型。

3 结束语

综上,本文从基础研究、构建机理、典型应用、挑战及对策、发展趋势五方面阐述了国内兴趣图谱研究进展:基础研究集中在兴趣图谱特征、价值、理论基础、关键技术;构建机理研究集中在构建思路、图谱管理;应用集中在电子商务、网络社交、智慧图书馆等领域;面临推荐质量与效果、用户隐私保护、信息内容范围及质量挑战;发展趋势为开放全网兴趣图谱、与社交图谱融合、信息表达创新和领域细分。□

参考文献

- [1] 孙雨生,严白云,仇蓉蓉,等. 兴趣图谱及其应用研究[J]. 情报理论与实践,2014,37(3): 89-94.
- [2] 张亚明,刘海鸥. 云环境用户兴趣图谱的网络社区营销推荐机理[J]. 情报杂志,2013,32(3): 184-188,202.
- [3] 丁绪武. 基于兴趣图谱的社会化电子商务社区发现研究[D]. 上海: 上海工程技术大学,2015.
- [4] 徐扬. 基于兴趣图谱的社交网络用户行为研究[D]. 广州: 暨南大学,2013.
- [5] 苏慧. 浅析兴趣图谱网站中的社交——以堆糖网为例[J]. 新闻世界,2014(8): 165-166.
- [6] 王守崑. 兴趣图谱与社交图谱的个性化推荐[J]. 程序员,2013(9): 109-111.
- [7] 陈媛媛,聂规划,陈冬林,等. 全网兴趣图谱下电子商务推荐理论研究现状[J]. 武汉理工大学学报: 信息与管理工程版,2014,36(1): 109-113.
- [8] 刘平峰,朱孔真,杨柳,等. 基于用户兴趣图谱的个性化推荐系统设计[J]. 武汉理工大学学报: 信息与管理工程版,2014,36(3): 341-344,387.
- [9] 闫霏. 基于兴趣图谱的图书馆泛在服务模式构建研究[J]. 数字图书馆论坛,2015(12): 41-44.
- [10] 陈媛媛. 用户兴趣图谱演化机制研究[D]. 武汉: 武汉理工大学,2014.
- [11] 瞿敬勃. 全网兴趣图谱下解析电子商务推荐理论研究现状[J]. 信息通信,2014(9): 147-148.
- [12] 张薇,马卫. 基于兴趣图谱的智慧图书馆SOLOMO模式构建[J]. 商业经济研究,2015(6): 84-85.
- [13] 张飞相,宗利永. 基于社会网络分析的兴趣图谱口碑传播者角色分析[J]. 现代情报,2014,34(10): 45-49.
- [14] 廖贤敏. 基于兴趣图谱的电子商务社交化方法研究[D]. 武汉: 中南民族大学,2012.
- [15] 冰河洗剑. Pinterest——兴趣图谱引领社交网络新潮流[J]. 大众软件,2012(9): 47-53.
- [16] 张佳琦. 基于兴趣图谱的社交网站成功要素研究[D]. 上海: 上海师范大学,2014.
- [17] 丁绪武,吴忠,夏志杰. 社会化电子商务用户兴趣图谱构建的研究[J]. 情报理论与实践,2015,38(3): 90-94.
- [18] 周焯彬. 兴趣社交图谱分众社交时代来临了?[J]. 环球企业家,2012(12): 120-121.
- [19] 谢璞. 宅女的“兴趣图谱”[N]. 21世纪经济报道,2012-01-06(022).
- [20] 李奕. 兴趣图谱重塑社交网络[N]. 中国计算机报,2012-03-12(004).
- [21] 张俊林. 标签传播算法在微博用户兴趣图谱的应用[J]. 程序员,2012(7): 50-53.
- [22] 阎娜. 腾讯爱看深耕社会化阅读[N]. 中国计算机报,2011-11-21(006).
- [23] 孔离. 当兴趣图谱成为一种商务模式[J]. 创新科技,2012(1): 52-53.
- [24] 张俊林. 大规模微博用户兴趣图谱挖掘[J]. 程序员,2014(5): 102-105.
- [25] 张召. 在线论坛用户兴趣图谱发现于个性化信息推荐[D]. 上海: 华东师范大学,2012.
- [26] 马建国,邢玲,李幼平,等. 广播型网格的用户兴趣图谱[J]. 电子学报,2005,33(1): 142-146.
- [27] 戴盈. 数字化时代读者的选书智慧和阅读兴趣图谱构建[J]. 中国出版,2014(5): 40-44.
- [28] 卢松松. 轻博客: 兴趣图谱下的营销叙事[J]. 销售与市场: 管理版,2012(11): 72-74.
- [29] 齐洁. 兴趣图谱聚集用户[N]. 中国经营报,2011-12-05(C13).
- [30] 国琳,左万利. 基于兴趣图谱的用户兴趣分布分析及专家发现[J]. 电子学报,2015(8): 1561-1567.

作者简介: 孙雨生,男,1980年生,博士后,副教授,硕士生导师。研究方向: 计算机信息系统工程,知识管理技术,数字图书馆技术。
张恒,男,1994年生。研究方向: 知识管理技术。
朱礼军,男,1974年生,博士后,研究员,博士后合作导师。研究方向: 知识管理技术。

录用日期: 2016-07-18