

情报理论与实践
Information Studies: Theory & Application
ISSN 1000-7490, CN 11-1762/G3

《情报理论与实践》网络首发论文

题目：国内基于知识图谱的信息推荐研究进展
作者：孙雨生，祝博，朱礼军
网络首发日期：2019-07-22
引用格式：孙雨生，祝博，朱礼军. 国内基于知识图谱的信息推荐研究进展. 情报理论与实践. <http://kns.cnki.net/kcms/detail/11.1762.G3.20190722.1559.007.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

●孙雨生^{1,3}, 祝博², 朱礼军³

(1. 湖北工业大学经济与管理学院, 湖北 武汉 430068; 2. 湖北工业大学职业技术师范学院, 湖北 武汉 430068; 3. 中国科学技术信息研究所国家科技信息资源综合利用与公共服务中心, 北京 100038)

国内基于知识图谱的信息推荐研究进展*

摘要: [目的/意义]揭示国内基于知识图谱的信息推荐研究进展。[方法/过程]文章用内容分析法归纳了74篇文献内容,揭示了基于知识图谱的信息推荐内涵及研究框架,并从架构体系与关键技术、核心内容、领域应用三方面阐述了国内基于知识图谱的信息推荐研究进展。[结果/结论]架构体系研究集中于体系结构、功能模块和运行机理;关键技术涉及知识图谱构建、知识表示;核心内容研究集中于用户兴趣建模(涉及用户兴趣模型表示、初始化和进化)、推荐机制(按知识图谱在信息推荐中扮演角色分基于本体推荐、基于表示学习推荐及混合推荐)、信息资源管理(涉及数据挖掘、表示和存储等);应用集中于教育、工业、医疗、旅游等领域。

关键词: 知识图谱; 信息推荐; 信息资源管理; 知识表示; 用户兴趣建模

Research Development of Information Recommendation Based on Knowledge Graph in China

Abstract: [Purpose/significance] In order to reveal research development of information recommendation based on knowledge graph in China. [Method/process] Using the content analysis method, the authors summed up the content of the 74 articles, revealed the connotation and research framework of information recommendation based on knowledge graph, and introduced its research development in China from three aspects of architecture and key technologies, core content and domain application. [Result/conclusion] Architecture system research focus on architecture, function modules and operating mechanisms; Key technology involves construction of knowledge graph, knowledge representation; Core content research focus on user interest modeling (it involves representation, initialization and evolution of user interest model), recommendation mechanism (According to the role of knowledge graph in information recommendation, there are three types of recommendation: ontology-based recommendation, representation learning-based recommendation and hybrid recommendation), information resource management (it involves data mining, data representation and data storage, etc); Applications focus on education, industry, medical, tourism and other fields.

Keywords: knowledge graph; information recommendation; information resource management; knowledge representation; user interest modeling

近年移动互联网、物联网、社会化媒体迅猛发展,在方便用户获取信息的同时也加剧了信息过载、数据多态异构分布等问题,致使缺乏语义技术支持的传统信息推荐面临更严峻的数据稀疏、冷启动、时效性、智能性等挑战而无法高效挖掘用户与项目信息以发现用户兴趣并有效融合、智能推荐其所需信息,严重影响用户体验及忠诚度、信息资源传播及利用率。凭借知识结构化、语义关联化、推理智能化等优势,新近出现的知识图谱为该系列问题解决提供了新途径:通过提供表示客观世界概念(实体)及其关系(属性)的知识体系(支持机

*本文为教育部人文社会科学研究规划基金项目“基于本体的数字图书馆语义用户兴趣模型构建机理及应用模式研究”(项目编号:17YJA870016),中国博士后科学基金项目“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),全国教育科学规划基金项目“基于多数据源、多方法融合的学科知识图谱构建方法研究”(项目编号:DIA160326)和全国大学生创新创业训练计划项目“夕龄乐连养老服务平台”(项目编号:201810500027)的成果之一。

读化信息处理、智能化知识推理)支持语义用户兴趣建模(基于三元组形式本体动态、智能采集并结构化、概念化异构用户信息、兴趣、行为等数据以挖掘其语义信息、潜在兴趣,基于属性构建概念语义网络以揭示其语义关系并支持智能推理,进而自动提取、精准量化(基于一义多词降维兴趣项、融合兴趣度)用户兴趣以构建、进化语义用户兴趣模型)、语义项目建模及相似度计算(动态采集、智能处理、结构化表示异构项目数据语义及关系进而挖掘推理潜在信息并形式化存储,以全面揭示项目属性(项目间、项目与用户间)关系进而管理项目信息)、语义推荐算法及其智能混合,最终实现信息推荐智能化、动态化、精准化、多样化(基于实体关系^[1]丰富推荐结果并提升其可信度),因此,研究基于知识图谱的信息推荐有重要意义。

本文以知网、万方的学位论文库、期刊论文库及维普的期刊论文库为信息源,以“知识图谱”与“推荐”为关键词组合在主题中检索文献(截止2019年5月22日,从CNKI获硕博论文30篇、期刊论文36篇;从万方获硕博论文43篇(新发现16篇)、期刊论文52篇(新发现23篇);从维普获期刊论文21篇(新发现0篇),阅读并剔除主题无关文献后共72篇相关),为进一步获取相关文献,笔者以“知识图谱”与“推送”为关键词组合在主题中检索(从CNKI获硕博论文3篇、期刊论文6篇;从万方获硕博论文7篇(新发现6篇)、期刊论文6篇(新发现4篇),从维普获期刊论文0篇,阅读并剔除主题无关文献后共两篇相关),合计有效文献74篇,然后详读74篇文献归纳基于知识图谱的信息推荐内涵、研究框架及其研究进展并按提及频次、内容质量详细标注,本着最大限度反映国内基于知识图谱的信息推荐研究进展重要文献、优中选优(剔除标注次数少、与其他标注文献内容重复文献)原则选出25篇参考文献(内容覆盖74篇文献),最后从架构体系与关键技术、核心内容、领域应用三方面阐述国内基于知识图谱的信息推荐研究进展。

1 基于知识图谱的信息推荐简介

1.1 定义及内涵

基于知识图谱的信息推荐依托知识图谱结构化实体、丰富语义关系、智能推理优势,基于实体表示与特征关联(实体化信息以构建知识图谱,基于关联属性多路径遍历实体(涉及下位实体、二元(互补或相似)实体)以挖掘实体间关系)动态采集、深度挖掘用户信息及行为并智能推理以提取用户(潜在)兴趣(特征)或关系并结合场景(上下文)信息构建并动态进化语义用户兴趣模型以全面精准揭示用户个性需求;动态采集并形式化多源异构项目信息、深度挖掘项目间潜在语义关系并精准表示、规范存储、智能更新以实现语义层项目信息管理,基于本体形式化描述、实体向量化集成表示用户、项目信息并支持语义层知识推理、相似度计算及推荐算法混合,实现形式化、智能化、个性化推荐及结果图谱化,与传统信息推荐对比见表1。

表1 基于知识图谱的信息推荐和传统信息推荐对比

项目		基于知识图谱的信息推荐	传统信息推荐
概述	目的	结构化、语义化海量分布异构信息,提升推荐精确性、时效性、智能性	基于用户兴趣动态发现、预测并推荐信息以降低用户认知负荷、合理化资源配置及流动
	优势	支持海量异构分布数据语义信息形式化表示、智能化关联及推理以实现语义用户兴趣、项目建模、推荐算法混合、推荐结果图谱化	简单、成熟且应用广
	劣势	应用垂直化,知识图谱构建及在线进化较难,处于探索阶段	冷启动、语义缺失(难处理多源异构数据),时效性、精准性及智能性弱
用户	模型表示	基于知识图谱精准表示用户兴趣、进行逻辑推理及	案例表示、神经网络、向量空间模型、

兴趣		动态挖掘以便机器理解处理	用户—项目评价矩阵、关键词
建模	模型初始化	基于聚类、关联规则、神经网络、贝叶斯网络等采集、预处理用户数据并挖掘潜在兴趣关系以确定用户兴趣项及兴趣度、图谱化兴趣项间关系	基于数据挖掘、机器学习等采集并预处理用户数据，确定用户兴趣项及兴趣度
	模型进化	基于语义关系推理用户潜在兴趣进而智能进化	定期进化、动态跟踪但智能性较弱（难挖掘、预测用户兴趣）
推荐机制	基于本体推荐	基于本体结构化、语义化表示用户兴趣、项目信息及各自关系并基于语义关系智能推理以提高推荐精确性	不详
	基于表示学习推荐	通过向量化三元组形式存储的实体信息以用较少语义维表示实体语义及关系，进而计算语义层实体相似度进行推荐	基于特征向量表示标签、项目并加权融合进行推荐但空间维度高、计算复杂且语义缺失
	混合推荐	基于知识图谱揭示项目属性并通过推理补全语义，进而基于混合推荐机制（多协同过滤推荐）语义推荐	综合各算法优点，分为 ^[2] 加权、条件变换、特征（属性）扩充、特征组合、层叠和元层次
信息资源管理	数据采集	基于机器学习、网络爬虫采集领域知识库及相关数据支持机读化处理、形式化存储	基于网络爬虫、人工录入、自然语言理解采集用户、项目信息
	数据挖掘	基于分类、关联规则、潜在特征向量、统计学习结合知识图谱实体关系实现语义挖掘、推理	基于分类、关联规则挖掘、推理非语义层异构数据关系
	数据表示	基于知识图谱节点及边表示实体及关系以语义化、形式化资源并支持机读化逻辑处理、降维式相似度计算，实现语义层关联化、智能化表示	逻辑、产生式、框架、特征向量
	数据存储	RDF、图数据库	（文件、关系）数据库、本体库
	数据更新	基于语义层动态采集、智能进化构建新图谱、扩充原图谱但图谱规模较大时较难	基于用户反馈及兴趣变化定期自适应更新

1.2 研究框架

纵观现有文献，国内相关研究成果最早是赵佳男 2014 年发表的《数字学习资源推荐技术研究现状及趋势分析》，目前整体处于研究初期但呈上升趋势，分析现有文献，笔者提出基于知识图谱的信息推荐研究框架，见图 1。

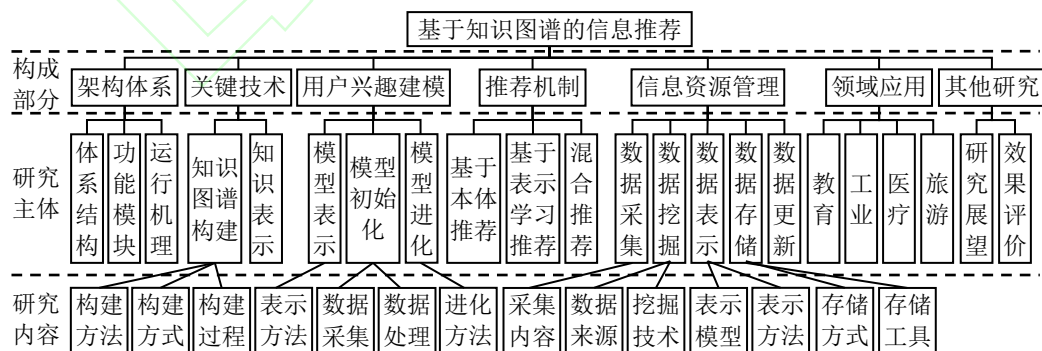


图 1 基于知识图谱的信息推荐研究框架

2 架构体系与关键技术

2.1 架构体系

2.1.1 体系结构

1) 基于知识图谱的教育推荐系统。赵佳男^[3]、黄华升^[4]提出 B/S 结构基于知识图谱课程^[3]、学习资源^[4]推荐体系分为表示层（以知识图谱可视化知识^[4]）、数据（媒体资源^[4]）处理层（链接、预处理学习（学科^[4]）资源）、数据层（转换、存取和处理数据^[2]并以图谱存储，构建学科资源本体模型（涉及定义逻辑、描述关系、标注标签）^[4]）；朱臻^[5]提出基于知识图谱初中英语题目推荐体系分为应用层（基于知识图谱管理师生、题目和错题）、功能支撑层（知识水平定级、题目推荐、难度管理）、系统集成层（基于 Neo4j、MySQL、MongoDB 存储数据）。

2) 基于知识图谱的工业推荐系统。邵领^[1]、张亿^[6]提出 B/S 结构基于知识图谱的信息检索^[1]、保险产品推荐^[6]体系分表示层^[6]、（知识检索^[1]）逻辑层、数据层（采集结构化实体并基于本体库^[1]处理、存储用户、产品信息^[6]）、知识图谱构建层^[1]。

3) 基于知识图谱的医疗推荐系统。屈峰林^[7]提出 B/S 结构基于知识图谱的医疗知识推送体系分为前端 UI、表示层、业务层（管理、推荐用户等）、数据层（基于 Neo4j、MySQL 管理数据）。

2.1.2 功能模块

1) 基于知识图谱的教育推荐系统。赵佳男^[3]、朱臻^[5]提出分为知识图谱表示^[3]、学习推荐^[5]（基于打分排序^[3]、难度值相似度^[5]推荐）、资源管理（教育知识图谱^[3]（基于 SCORM、果壳网及本地资源）、学习资源^[5]（基于图谱进化机制调整题目及难度））模块；黄华升^[4]提出分为学习推荐（基于协同过滤结合知识图谱知识点关系推荐资源列表）、资源管理（学习者、专业知识图谱）模块。

2) 基于知识图谱的工业推荐系统。邵领^[1]提出分为搜索（基于本体库解析用户请求以抽取关键词进而检索）、爬虫及索引（爬取信息并抽取实体进而索引）、知识图谱构建模块；张亿^[6]提出分为推荐（分析用户需求并协同推荐）、用户信息挖掘及存储、产品信息管理模块；邵元新^[8]提出分为前台表示（展示采集信息、三元组抽取结果及图谱）、后台处理（资源采集及预处理（实体识别及关系抽取）、知识图谱构建）模块。

3) 基于知识图谱的医疗推荐系统。范灿华^[9]提出分为解析表示（转化医疗数学模型及知识以便理解）、知识引擎（依次机器预诊、与医生交互诊断、按患者需求推荐医生）、医生实例维护及存储、知识库维护模块。

4) 基于知识图谱的旅游推荐系统。唐慧琳^[10]提出 SSH 架构基于知识图谱旅游问答推荐系统分为前端表示、语义分析（提取关键词）、系统底层（存储知识图谱及 Aspect 分数）模块；孙文平等^[11]提出基于知识图谱旅游路线推荐系统分为路线推荐、路线序列挖掘、知识抽取与处理模块。

2.1.3 运行机理

1) 基于知识图谱的教育推荐系统。赵佳男^[3]、黄华升^[4]提出基于网络教育平台^[3]课程、基于 MOOC 学习线路^[4]推荐流程：数据（学习资源^[3]）采集^[4]、信息资源管理^[4]（基于 XML 描述资源及关系^[3]）、服务推荐（基于知识图谱结合知识点权值及关系^[3]、学习者—资源矩阵^[4]计算相似度进而推荐）；王家辉等^[12]提出微信刷题系统推荐流程：基于错题关联知识点与学习知识图谱进而检索知识点（题目）、推荐题目。此外，邹弘智等^[13]提出基于知识图谱文献推荐流程：离线构建维护通用知识图谱、建模文献语义树，在线建模用户知识需求并依其与候选文献知识距离推荐。

2) 基于知识图谱的工业推荐系统。张亿^[6]提出保险产品推荐流程：用户选择知识图谱或推荐列表中险种节点、产品，基于用户购买意向推荐保险服务；邵元新^[8]提出基于知识图谱产品推荐流程：基于 Web 采集并结构化资源（识别实体、抽取关系）进而构建知识图谱，基于用户需求、产品候选项推荐；此外，高广尚等^[14]提出知识面板中知识推荐流程：基于

知识图谱存储知识、挖掘并结合实体类别及关系深度推理用户需求，以面板形式推荐知识。

3) 基于知识图谱的医疗推荐系统。屈峰林^[7]提出医疗知识推送流程：基于用户症状提取实体并在知识图谱中检索后推送；范灿华^[9]提出医生推荐流程：基于患者病症、知识图谱进行预诊，基于医生信息相似度、预约挂号推荐。

4) 基于知识图谱的旅游推荐系统。汪沛^[15]提出旅游推荐流程：采集、提取用户兴趣实体并基于图聚类模型计算实体相似度以挖掘其潜在兴趣、判断实体与旅游知识图谱景点实体相似度进而推荐；唐慧琳^[10]提出旅游问答推荐流程：离线采集景点及评分数据以构建并补全景点知识图谱，在线解析用户提问并检索、排序后推荐；孙文平等^[11]提出旅游路线推荐流程：基于结构化知识库、网站构建知识图谱并结合用户游记生成路线数据库，基于用户兴趣、约束挖掘频繁线路序列并推荐。

2.2 关键技术

2.2.1 知识图谱构建 知识图谱知识准确性、覆盖面直接影响推荐效果，国内研究较多且相对成熟，多为领域知识图谱。

1) 构建方法。按构建主体分为专家构建、社群协同构建；按数据源分为基于半结构化数据（人工定义规则、基于（监督、半监督）学习构建）、基于非结构化数据^[10]（基于机器学习、自然语言处理技术）自动构建。

2) 构建方式。分为自顶向下（顶层学习、设计数据源，识别、填充实体）、自底向上（归纳底层实体以抽象出概念体系）^[16]和混合式（使用最多且实体抽取更准，常按领域本体特点设计知识图谱组织方法）。

3) 构建过程。分为信息抽取、知识推理、图谱更新，其中信息抽取^[5]包括实体识别（用归纳学习、DOM 树^[1]、自然语言处理^[5]、基于规则^[16]、机器学习^[7-8]、本体、正则表达式^[5]、模式匹配^[8]等从异构数据抽取实体及其属性、关系^[5,17]）和知识融合（用基于规则、统计机器学习、聚类、深度学习^[8]将实体消歧、对齐（将描述相同实体标识唯一化）、链接）；知识推理用基于关联挖掘、基于规则（用语言专家按任务要求设计的模板从文本中匹配实体并推理实体间关系）、基于潜在特征向量模型、聚类等，结合图谱关系推理实体间潜在关联以补全图谱，此外蔡鸿博^[18]基于协同过滤推荐算法补充实体上位词：计算已挖掘上位词与实体相似度并基于频繁项集、中文构词法构建并扩展上位词间关系以补全知识图谱。

2.2.2 知识表示 知识图谱构建和应用关键，旨在结构化表示实体及关系，分为^[15]基于符号逻辑（规则）（分谓阶逻辑、产生式、框架表示）、Web 描述语言（核心是 XML、RDF、本体^[1]、OWL^[9]等）和表示学习（基于机器学习、神经网络、深度学习^[15]向量化富语义信息）的知识图谱并嵌入稠密低维空间以高效计算相似度、全面表示项目属性）^[19]。其中表示学习算法^[15,19]分为基于潜在特征向量模型（基于张量分解（涉及 SME、MLP、NAM、NTN、RESCAL 算法，旨在重构知识库信息、预测关系但常张量维度高、计算量大）、距离翻译（势能函数法，涉及 TransE、TransH、TransR/CTransR 等算法，核心是头实体到尾实体翻译操作）、神经网络、矩阵分解模型^[16]）、图嵌入（基于随机游走算法（涉及 Node2Vec^[20]、DeepWalk^[17]）生成节点序列及其关系并嵌入低维向量空间^[17]，基于特征（邻节点或节点间路径）计算节点相似度^[10]（指标分为局部、全局和半局部））^[15,17,20]，此外，唐慧琳^[10]基于表示学习计算实体间语义关联以推理知识、补全旅游知识图谱：基于融合三角形子图的嵌入表示模型（基于不同模式三角形结构、实体间关系训练关系向量）和融合拓扑信息的语义平嵌入表示模型（计算实体语义相似度以挖掘关联实体）提升表示学习准确性进而构建补全知识图谱。

3 核心内容

3.1 用户兴趣建模

基于知识图谱用户兴趣建模旨在实现语义用户兴趣建模，常基于知识图谱层次化、形

式化、语义化采集并挖掘用户与项目、用户间信息以推理潜在兴趣、全面揭示其需求，分为模型表示、初始化和进化。

3.1.1 模型表示 基于知识图谱知识形式化、实体关联化、推理智能化优势分析用户需求以全面精准表示其兴趣模型（用节点及边分别表示兴趣项及兴趣度），现有研究集中于表示方法，主要有关键词、基于向量空间模型^[2]、类型层次结构模型、用户—项目矩阵^[19]、语义网络、本体^[9]、决策树、简单贝叶斯概率模型等。此外，朱臻^[5]、邵礼旭^[21]基于知识图谱建模学习者模型表示其知识点掌握度。

3.1.2 模型初始化 基于知识图谱采集、挖掘并量化兴趣度以提升用户兴趣模型准确性，现有研究集中于数据采集及处理。

1) 数据采集。主要采集显式信息（用户属性（基本信息^[6]，涉及姓名、性别、年龄、语言^[13]、职业、国籍^[14]等）、兴趣^[17]（爱好、标签、属性、情感、商品、购买意向^[6]、时间戳^[2]等）和隐式信息（用户历史行为和潜在关系，如访问日志^[6,14]、记录（浏览^[15]、检索^[5]）、评论^[2]（购物、社区论坛、游记^[11]）、查询语言^[14]、客户端数据等），此外，黄华升^[4]、邵礼旭等^[21]采集学习者特征数据（认知水平、记录、策略、学习路径^[4]、学习目标^[3]等），孙文平等^[11]采集用户游记信息（评论、签到、照片等）。

2) 数据处理。主要研究用户兴趣数据预处理和兴趣度量化：前者分为基于图聚类^[15]、关联规则^[1]、机器学习^[17]（神经网络、贝叶斯网络）等预处理兴趣数据进而以节点距离、边权值量化兴趣度，基于分词、词性标注^[10]等处理 Web 日志及评分等两类；汪沛^[15]基于用户行为数据（浏览记录）聚类实体节点以挖掘用户兴趣。对于后者，赵佳男^[2]基于资源访问量量化兴趣度；孙文平^[11]综合景点价值度、用户兴趣量化兴趣度；邹弘智^[13]基于伪相关反馈法匹配用户初始查询及其知识背景确定主题关注度以量化兴趣度并挖掘潜在兴趣；杨晋吉^[2]用用户行为时间及次数动态调整图谱节点权值以量化兴趣度；黄华升^[4]、王一鸣^[22]用用户—项目矩阵评分量化兴趣度。

3.1.3 模型进化 用户兴趣模型随用户兴趣变化动态进化，分为模型学习与更新：基于自适应^[13]、机器学习、动态兴趣建模^[17]等机读化采集、处理用户兴趣（反馈）以智能学习、动态进化模型。邹弘智^[13]基于激活扩散法拓展用户需求、挖掘潜在兴趣、进化用户兴趣模型；杨晋吉等^[20]构建知识图谱模型（混合反馈模型、用户迁移模型）调整权值进化用户兴趣模型；邵礼旭等^[21]按三层资源处理架构（数据图谱、信息图谱和知识图谱）结合用户评分反馈进化用户兴趣模型。

3.2 推荐机制

分析现有文献，笔者认为基于知识图谱信息推荐机制按推荐过程中知识图谱扮演角色分为基于本体推荐、基于表示学习推荐及混合推荐。

3.2.1 基于本体推荐 基于本体构建知识图谱可更结构化、层次化、精准化表示实体属性及关系并基于三元组关系及实体相似度推理新实体及潜在语义关系，在扩展知识图谱的同时支持高效语义计算以精准推荐。

范灿华^[9]构建医疗本体知识图谱（基于患者需求、半结构医疗文档，用爬虫粗粒度、人工细粒度抽取并优化），基于本体、实例语义计算相似度进行推荐；邹弘智^[13]基于领域本体库、语义词典构建知识图谱，基于领域本体库、主题语义属性构建文献语义树，计算用户兴趣模型和文献知识模型余弦相似度进行文献推荐。

3.2.2 基于表示学习推荐 基于表示学习向量化用户、项目结构及其语义并嵌入低维语义空间，用语义距离（基于链接数^[17]、欧氏距离、余弦距离^[9]）、相关系数^[22]、隐式语义反馈路径、熵权系数^[16]、关联规则^[18]等相似度算法计算实体、关系间相似度^[19]进行推荐以缓解数据稀疏和冷启动问题、提升推荐效果、存储效率。

现有研究分为基于潜在特征向量、图嵌入两类：针对前者，唐慧琳^[10]基于距离监督、

嵌入表示学习更精确向量化实体结构以补全旅游知识图谱,按用户需求检索知识图谱并结合评分排序推荐;陈平华等^[23]基于 TransR-MF 算法将项目实体嵌入低维向量空间结合矩阵分解模型推荐项目。针对后者,汪沛^[15]用基于图的随机游走算法结合人工标注的实体关系图谱降维计算实体相似度进而提高其消歧准确性,构建实体关系网络并基于图的层叠分析分类关系类型以丰富实体属性,基于实体关系图模型、用户兴趣模型(基于随机游走计算图核相似度挖掘用户潜在兴趣)聚类并结合领域实体分析模型推荐;贾中浩等^[24]结合随机游走、基于景点的属性子图生成节点序列并嵌入低维空间,用特征学习挖掘更富语义的节点特征,计算游客和景点向量余弦相似度推荐列表。

3.2.3 混合推荐 综合基于知识图谱推荐和其他推荐提升推荐精确度,多混合基于表示学习推荐与基于项目协同推荐。吴玺煜^[19]、王一鸣^[22]提出基于表示学习(降维用户显式反馈^[19](用户-项目评分矩阵^[22]),基于翻译算法 TransE 计算项目语义相似度)、协同过滤混合推荐算法;杨晋吉等^[20]提出融合知识图谱的排序学习混合推荐算法:用 Node2Vec(基于深度学习的网络表示法,将实体嵌入低维空间,通过余弦相似度计算实体关系)算法抽取知识图谱特征,混合评分推荐模型、基于排序学习的反馈模型、用户兴趣迁移模型进行 Top-N 推荐。此外,黄华升^[4]、王家辉等^[12]提出基于教育知识图谱推荐、面向用户协同过滤混合推荐算法(基于用户学习水平与邻居用户群相似度结合知识图谱推荐学习资源^[4]、基于错题反馈推荐和用户知识点相似题目^[12])推荐学习资源(微信题目^[12]);门瑞^[25]混合基于 MOOC 知识图谱推荐(针对用过 MOOC 资源用户基于资源间关系推荐)、基于案例推理推荐进行课程推荐(针对未用过 MOOC 资源用户基于问卷构建用户案例模型并基于案例相似度推荐)。

3.3 信息资源管理

基于知识图谱信息资源管理研究用户及项目数据库中数据采集、挖掘、表示、存储及更新以管理用户、项目、推荐过程生成信息,现有研究集中于数据挖掘、表示和存储。

3.3.1 数据采集 基于知识图谱结构化采集多源异构数据(多为领域知识及其关系),内容多为用户^[14,17]、项目信息(源自垂直、百科网站、结构化数据^[3])、交互信息(日志、社区评论、评分^[20]、上下文信息(用户自主录入^[17]、传感器采集))等;此外,目前知识图谱构建多基于国外开源数据库致使国内知识图谱发展依赖国外数据源^[9]。

3.3.2 数据挖掘 基于知识图谱实体及关系推理挖掘潜在信息,多研究数据来源、挖掘技术:前者包括用户搜索日志^[3]及隐式信息、百科类网站(如维基百科、Freebase、DBpedia^[13])、结构化数据^[3](如开放图谱知识库、领域分类体系或本体库、结构化知识库^[11])、半结构化数据^[9]、非结构化数据^[8];后者主要涉及网络爬虫^[1,10]、自然语言处理^[5]、Web 数据挖掘^[1-2](涉及关联规则、分类离线挖掘(聚类、Apriori^[18]、贝叶斯网络)、统计学习)。

3.3.3 数据表示 基于知识图谱(节点、边及其权值)以较少语义维结构化精准表示用户、项目实体语义及关系以高效推荐,多研究表示模型及方法:表示学习常用模型^[16]分为代表模型(涉及距离(向量空间模型)、神经网络、双线性^[16](隐变量模型)、神经张量、矩阵分解、翻译模型)、复杂关系模型(TransH、TransR、TransD 等),领域知识库表示法^[9]有语义网络法、框架法、产生式法等。此外,常用^[6]D3.js^[3]、Echarts、Processing、Gephi 等工具列表、网状或树状可视化知识图谱结构。

3.3.4 数据存储 用数据库存储用户、项目知识图谱(数据)及推荐过程信息,多研究存储方式、工具。知识图谱存储方式分为 RDF(数据序列化^[7])、图结构;常用工具分为 RDF 式(Allegro Graph、Jena^[6])、图数据库(Neo4j(支持 Cypher 查询知识属性图数据^[1,8]、实体间关系可视化)、OrientDB^[7])、关系数据库(MySQL、PostgreSQL^[9])。此外,朱臻^[5]分别用 MySQL、Neo4j、MongoDB(文档型)混合存储用户及其学习信息、知识图谱、题目;屈峰林^[7]分别用 MySQL、Neo4j 存储用户及药品信息、疾病实体及其关系。

3.3.5 数据更新 目前较难^[17],主要基于知识图谱更新用户(基于用户兴趣模型进化机制^[2])、

项目（基于知识图谱更新机制^[17]（更新节点属性及关系^[1]）和数据库自带程序^[2]）及推荐过程生成信息。邵礼旭等^[21]基于学习者能力、外部环境分别更新用户兴趣模型、资源框架（由5W引入的数据、信息和知识图谱）；门瑞^[25]（不）定期构建新专业图谱并融入原知识图谱。

4 领域应用

集中于教育、工业、医疗和旅游等领域且前景良好，其中教育、工业领域已有实现方式及框架。

4.1 教育

基于知识图谱表示课程（知识点）及关系建模知识结构并结合学习者兴趣图谱化推荐以提升学习全局性、连续性^[3]、系统性：黄华升^[4]提出基于知识图谱学习资源推荐系统：基于知识点间关系构建资源本体模型并映射到知识图谱并结合基于用户协同过滤推荐学习资源；王家辉等^[12]提出智能微信刷题系统：基于知识图谱揭示用户、题目及知识点关系并基于用户、题目及知识点间相似度协同过滤推荐题目。此外，部分学者基于知识度量计算推荐列表，郅弘智等^[13]基于知识图谱推荐文献以缓解主题多义性问题：基于兴趣挖掘在线建模并拓展用户知识需求，离线构建基于“背景知识库—词簇—主题网格”通用知识图谱按知识距离推荐文献；门瑞^[25]提出新 MOOC 推荐算法（经平台验证可显著提升国内 MOOC 平台推荐效果）：基于课程次序构建知识图谱并按知识点相似度推荐。部分学者融合教育理论和知识图谱进行推荐：朱臻^[5]提出基于知识图谱选题系统：将教育理论（过度学习、艾宾浩斯遗忘曲线）融入知识图谱构建（涉及知识抽取、融合），结合学生学习特征构建知识图谱选题模型并基于练习反馈动态推荐题目。

4.2 工业

张亿^[6]提出基于知识图谱保险产品推荐系统：基于知识图谱结构化保险实体及其关系、关联化挖掘用户兴趣并进行协同过滤推荐；邵元新^[8]提出基于 Web 工业知识图谱构建系统：Web 采集并结构化工业产品信息，基于混合模型法识别实体构建知识图谱以可视化并基于用户所选属性推荐产品；邵领^[1]提出基于知识图谱的搜索系统：构建“机器学习”领域本体库结构化实体及其关系，基于用户检索语句遍历知识图谱并计算实体相似度以推荐实体；高广尚^[14]以知识图谱为知识面板核心，基于知识图谱挖掘用户需求及场景信息（上下文、地理位置）并返回结构化实体进而图谱化推荐知识体系（实体及其关系）以提升用户体验。

4.3 医疗

屈峰林^[7]提出基于知识图谱医疗知识推荐系统：用知识图谱结构化医疗实体及关系，基于医疗结构特征、CRF 模型实体化电子病历，基于医疗知识图谱、患者检索语句、病症二次选配并推荐医疗知识；范灿华^[9]将知识图谱用于导医系统：基于本体医疗知识图谱、患者疾病模型匹配疾病实例并计算与医生语义相似度以推荐医生。

4.4 旅游

汪沛^[15]提出旅游推荐系统：基于知识图谱构建实体属性关系网络、基于属性图聚类（层叠分析）分类实体，将用户兴趣输入属性图聚类模型推荐旅游线路聚类结果；唐慧琳^[10]提出旅游问答推荐系统：机器处理用户自然语言形式提问，基于景点知识图谱按属性筛选景点并结合其评分（从评论中提取）排序智能推荐；孙文平等^[11]提出基于频繁序列旅游路线算法并结合知识图谱结构化多源数据以提升用户体验：基于旅游知识图谱、用户游记结合 I2-PrefixSpan（基于人物类型的频繁序列算法）生成、挖掘旅游路线序列并结合用户兴趣、评分推荐。

5 其他研究

5.1 研究展望

5.1.1 知识图谱构建智能化 当前知识图谱构建流程、应用（存储、检索^[9]）相对成熟稳定但知识推理能力有限、难更新问题严重影响推荐效果，基于机器学习、数据挖掘、深度学习^[17]提升知识图谱构建效率（主要是数据及其关系抽取、实体识别及表示，以更好融合多源异构数据提升知识图谱覆盖率）、拓展知识图谱应用、提升知识推理能力（挖掘用户与项目间深层关系及相关知识^[17]以丰富知识图谱语义）是研究趋势。

5.1.2 用户兴趣建模精准化 用户兴趣模型精准性直接影响推荐质量，采集并挖掘跨领域、应用、平台、载体的用户（潜在^[15]）兴趣及用户兴趣模型精准表示是研究趋势。

5.1.3 推荐精准化、动态化、多元化 推荐精度、实时性随用户、项目信息增加面临挑战，需加强融入上下文^[17]、地理位置^[11]、社交信息^[17]提升面向群体和个体（行为^[15]）的用户兴趣建模及推荐精度、实时性研究；同时，当前基于知识图谱的混合推荐多混合知识图谱与协同过滤推荐，与其他推荐机制混合亟待研究^[15]。

此外，急需强化用户隐私保护机制、基于通用知识图谱的推荐模式^[3]研究。

5.2 效果评价

目前多基于传统指标评价推荐效果，分为用户满意度（准确率（推荐项占用户喜欢项比重^[17]以评价实体识别准确度^[8]）、召回率（评价用户满意项被推荐比重）、*F-measure* 值（准确率和召回率的加权平均^[19]）、推荐项覆盖率、推荐项评分预测准确度（平均绝对值误差（项目评分和预测评分误差绝对值和的均值，评价真实和预测评分均差^[23]）、均方根误差（项目评分和预测评分误差的平方和的均值的根值^[23]）、推荐项排序预测准确度（平均排序分（评价用户对推荐列表中推荐项位置满意度）、*P@N*（评价特定查询 TOP-N 排序准确率^[20]）、平均查准率（评价推荐结果排序精度^[20]）4 类。此外，邵弘智等^[13]提出平均排序公式 TXARS 评价文献推荐效果。

6 结束语

综上，本文从架构体系与关键技术、核心内容、领域应用三方面阐述了国内基于知识图谱的信息推荐研究进展：架构体系研究集中于体系结构、功能模块和运行机理；关键技术集中于知识图谱构建、知识表示。核心内容集中于用户兴趣建模、推荐机制和信息资源管理，其中，用户兴趣建模核心研究用户兴趣模型表示（集中于表示方法）、初始化（集中于数据采集及处理）、进化但亟待加强；推荐机制集中于基于本体推荐、基于表示学习推荐和混合推荐；信息资源管理集中于数据采集（集中于采集内容）、数据挖掘（集中于数据来源及挖掘技术）、数据表示（集中于表示模型及方法）、数据存储（集中于存储方式及工具）及数据更新。应用集中于教育、工业、医疗、旅游等领域。此外，该主题研究呈知识图谱构建智能化、用户兴趣建模精准化及推荐精准化、动态化与多元化等趋势。

下一步，笔者将研究基于知识图谱的信息推荐机理，以供相关研究与实践参考。□

参考文献

- [1]邵领.基于知识图谱的搜索引擎技术研究与应用[D].成都：电子科技大学，2016.
- [2]孙雨生，张晨，任洁，等.国内电子商务个性化推荐研究进展：核心技术[J].现代情报，2017，37（4）：151-157.
- [3]赵佳男.基于知识图谱的课程数字化资源推荐方法研究[D].北京：北京邮电大学，2015.
- [4]黄华升.基于知识图谱的个性化学习资源推荐研究[J].软件工程，2018，21（10）：24-26，23.
- [5]朱臻.基于知识图谱的初中英语选题系统及应用[D].北京：中央民族大学，2016.
- [6]张亿.基于知识图谱的保险产品个性化推荐系统构建[D].武汉：武汉科技大学，2016.
- [7]屈峰林.基于知识图谱的健康医疗知识推送系统研究[D].海口：海南大学，2018.
- [8]邵元新.基于 web 的工业产品知识图谱构建及应用[D].沈阳：沈阳航空航天大学，2017.

- [9] 范灿华.基于知识图谱的精准导医系统[D].广州: 华南理工大学, 2017.
- [10] 唐慧琳.融合结构和语义信息知识图谱补全算法研究[D].北京: 北京邮电大学, 2017.
- [11] 孙文平, 常亮, 宾辰忠, 等.基于知识图谱和频繁序列挖掘的旅游路线推荐[J].计算机科学, 2019, 46 (2): 56-61.
- [12] 王家辉, 文欣秀, 赵敏, 等.基于知识图谱的微信智能刷题系统[J].实验室研究与探索, 2018, 37(11): 125-128, 140.
- [13] 郇弘智, 闫健卓, 陈建辉.一种知识驱动的个性化文献推荐方法[J].计算机应用研究, 2018, 35 (12): 3603-3608.
- [14] 高广尚, 张智雄.面向实体的知识面板推荐模式分析[J].图书馆学研究, 2015 (1): 56-64.
- [15] 汪沛.基于领域知识图谱的个性化推荐方法研究[D].昆明: 昆明理工大学, 2017.
- [16] 徐增林, 盛泳潘, 贺丽荣, 等.知识图谱技术综述[J].电子科技大学学报, 2016, 45 (4): 589-606.
- [17] 常亮, 张伟涛, 古天龙, 等.知识图谱的推荐系统综述[J].智能系统学报, 2019, 14 (2): 207-216.
- [18] 蔡鸿博.基于上位词的中文实体关系图谱构建[D].哈尔滨: 哈尔滨工业大学, 2014.
- [19] 吴玺煜, 陈启买, 刘海, 等.基于知识图谱表示学习的协同过滤推荐算法[J].计算机工程, 2018, 44 (2): 226-232, 263.
- [20] 杨晋吉, 胡波, 王欣明, 等.一种知识图谱的排序学习个性化推荐算法[J].小型微型计算机系统, 2018, 39 (11): 2419-2423.
- [21] 邵礼旭, 段玉聪, 周长兵, 等.数据、信息和知识三层图谱架构的推荐服务设计[J].计算机科学与探索, 2019, 13 (2): 214-225.
- [22] 王一鸣.基于知识图谱的推荐技术研究及应用[D].成都: 电子科技大学, 2018.
- [23] 陈平华, 朱禹.融合知识图谱表示学习和矩阵分解的推荐算法[J].计算机工程与设计, 2018, 39(10): 3137-3142.
- [24] 贾中浩, 古天龙, 宾辰忠, 等.旅游知识图谱特征学习的景点推荐[J].智能系统学报, 2019, 14 (3): 430-437.
- [25] 门瑞, 杨芳萍.基于知识图谱及案例推理的 MOOC 资源推荐算法研究[J].陇东学院学报, 2018, 29 (1): 10-13.

作者简介: 孙雨生 (ORCID: 0000-0002-1302-981X), 男, 1980 年生, 副教授, 博士后, 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, 发表论文 65 篇, 主编教材 1 部。**祝博** (ORCID: 0000-0001-5413-6518), 男, 1998 年生。研究方向: 电子商务技术, 知识管理技术。**朱礼军** (ORCID: 0000-0002-7741-9286), 男, 1973 年生, 研究员, 博士后, 博士后合作导师。研究方向: 知识管理技术。

作者贡献声明: 孙雨生, 设计论文选题、研究框架, 指导论文并修改。**祝博**, 文献整理、阅读、论文撰写。**朱礼军**, 论文后期修改指导。

录用日期: 2019-06-17