DOI:10.12154/j.qbzlgz.2020.01.009

在线学术资源中知识图谱的应用研究综述*

熊回香 景紫薇 杨梦婷 （华中师范大学信息管理学院 武汉 430079）

摘 要：[目的/意义]文章针对国内外近年来在线学术资源中知识图谱的应用研究成果进行了梳理，以期为将来知识图谱应用于在线学术资源更多方面的深入研究提供参考和借鉴。**[方法/过程]**通过文献调研，对知识图谱近年来应用于在线学术资源研究中的情况进行了分析，并整理了学术知识图谱构建过程中的研究情况。**[结果/结论]**目前，知识图谱在在线学术资源的应用研究集中在学术知识图谱、科学文献推荐、学术知识发现、语义出版、学术社交网络、学者画像和学术资源共享等方面，但研究深度层面仍处于初期阶段，在这些基础之上对当前相关研究中存在的问题进行阐述，并指出了未来的研究方向。

关键词：知识图谱 在线学术资源 学术知识图谱

A Summary of Research on the Application of Knowledge Graph in Online Academic Resources

Xiong Huixiang Jing Ziwei Yang Mengting
(School of Information Management, Central China Normal University, Wuhan, 430079)

Abstract: [Purpose/significance] This paper sorts out the application research results of knowledge graph in online academic resources in recent years, in order to provide reference and reference for the further study of the application of knowledge graph in online academic resources. **[Method/process]** Through literature research, this paper analyzes the application of knowledge graph in online academic resources research in recent years, and organizes the research situation in the process of constructing academic knowledge graph. **[Result/conclusion]** At present, the application of knowledge graph in online academic resources focuses on academic knowledge graph, scientific literature recommendation, academic knowledge discovery, semantic publishing, academic social networks, scholar portraits and academic resource sharing, but the depth of research is still in the early stage. Based on these foundations, the problems existing in the current research are expounded, and the future research directions are pointed out.

Keywords: knowledge graph online academic resources academic knowledge graph

1 引言

随着大数据、海量在线学术资源在满足科研人员信息需求的同时，也带来了严重的信息过载压力，使得科研人员难以在短时间内充分挖掘隐含在其中的知识，无形中增加了选择合适学术资源的时间和精力。而知识图谱在组织、管理和理解互联网海量信息上具有的超强能力，为解决这一问题提供了新的方法和途径。知识图谱的前身是语义网，它继承了语义网和本体在知识组织和表达方面的理念，通过实体-关系的描述方式来展示客观世界各个概念实体之间的关系网络，不仅使计算机在文本语义理解上变得更加智能，而且为计算机

* 本文系国家社会科学基金年度项目“融合知识图谱和深度学习的在线学术资源挖掘与推荐研究”（项目编号：19BTQ005）、中央高校基本科研业务费重大培育项目“基于语义网的在线健康信息的挖掘与推荐研究”（项目编号：CCNU19Z02004）的研究成果之一。

用户提供了一种从关系视角来看待世界的方式。

当前工业界和学术界都开始加入面向在线学术资源的知识图谱应用研究中。在工业界，国外的Microsoft Academic Graph（MAG）和国内的Aminer科学知识图谱都走在了学术知识图谱产品的前列，而且为了进一步拓展跨国际的学术交流，两大公司还合作开发了开放学术图谱（Open Academic Graph，OAG）；在学术界，也有很多学者关注到了知识图谱在在线学术资源组织和管理中的价值，并做了一系列探索，本文对国内外近年来在线学术资源中知识图谱的应用研究成果进行了梳理，主要从在线学术资源中知识图谱的应用、学术知识图谱构建过程以及知识图谱在在线学术资源应用中面临的问题与研究趋势这三个部分进行了论述，在总结分析了当前该领域研究现状的同时，为将来知识图谱应用于在线学术资源更多方面的深入研究提供参考和借鉴。

2　在线学术资源中知识图谱的应用

知识图谱这一术语在2012年5月被谷歌正式提出[1]，它本质上是一种通过结构化的形式描述客观世界中存在的各种实体、概念及其关系的方式，其中实体指的是具体事物，概念是具有相同属性的实体的概括抽象，而关系则是实体、概念之间存在的联系[2]。如今知识图谱在智能搜索、知识问答和个性化推荐等方面展现出了很好的性能，而且应用领域涉及各行各业，包括企业、金融、医疗、教育等。其中，在线学术领域中知识图谱的应用近年来也逐渐进入人们的视野，由于学术资源主要服务于科研的属性，决定了其面向的用户多为科研院校的研究人员和学生，一定程度上限制了该领域的服务人群，但由于学术资源涵盖了各个学科的专业知识，不同学科的研究方向和研究方式各有不同，这为知识图谱在该领域的应用带来了很多机遇与挑战。

目前，在线学术资源中知识图谱的应用主要分布在以下几个方面：学术知识图谱构建、科学文献推荐、学术知识发现与语义出版、学术社交网络结构挖掘、学者画像构建和学术资源共享。

2.1　学术知识图谱构建

学术知识图谱是面向学术资源构建起的知识图谱应用，与医学知识图谱、企业知识图谱等垂直知识图谱一样，都是针对特定领域的特定需求，对海量数据进行管理、共享及应用的工具，不同的是，相比其他垂直知识图谱而言，学术知识图谱的学术性更强，对实体、关系和属性的严谨性要求更高；除此之外，和同领域中基于文献外部特征（标题、关键词、作者等）进行可视化分析的学科知识图谱相比，学术知识图谱的应用范围更广，而且不局限于文献类资源，其构建的资源集面向科学文献、专业词典、社交网络上的学术信息等更多形式的在线学术资源，为学术领域在智能检索、智能问答和个性化推荐等方面的发展奠定了基础。

相关研究中，Huang等[3]提出了一个名为AKMiner（Academic Knowledge Miner）的系统，用于从特定领域的文章中自动挖掘有用的知识，然后在视觉上向用户呈现知识图谱；秦玥[4]构建了一个面向科技论文的知识图谱框架，主要通过实体识别、摘要的语义模块划分等技术手段，丰富了科技论文知识图谱相关实体以及实体间的属性和关系；Sadeghi等[5]为了更好地整合分布在不同数据源上的学术信息，提出了一种创建高质量知识图谱的数据管道，以DBLP和MAG的异构来源数据为基础，构建了一个集成的学术交流元数据知识图谱（SCM-KG），展示了基于规则的数据映射中的并行化能力。

在基于文献数据库资源进行学术知识图谱的构建上，还有很多其他的实践案例。汤庸等[6]以学者网SCHOLAT为背景，给出了学术知识图谱的模式设计，具体包括实体关系模型构建、学术信息获取和学术实体匹配三大环节，其中基于学者网的学术信息获取来源包含两部分，一部分是由注册用户产生的个人学术信息，另一部分是收录站外的学术文献资料；张晔等[7]针对原AceMap学术大数据分析系统由于依靠关系型数据库存储数据而限制了知识地图种类的问题，开发了AceKG知识图谱，将关系型数据存储方式改为三元组形式的图存取方式，保存了更丰富的语义信息，也拓展了知识地图的可显示种类。

除了可以在基于文献数据库资源的基础上实现学术知识图谱的构建以外，还可以基于专业词表来构建，Qiao B等[8]以农业领域学术知识为例，设计了一个基于农业叙词表（AT）构建起的农业知识图谱模型，该模型实现了从农业叙词表到农业知识图谱的自动转换，为将来基于语义的农业信息检索和问答系统的构建奠定了坚实的基础。

2.2　科学文献推荐

知识图谱的出现，为个性化推荐可以从语义关联角度切入提供了一系列解决方法，而且由于科学文献知识的专业性，可以大大降低知识抽取阶段的难度，所以近两年来，通过知识图谱的相关技术来提升科学文

献资源推荐的研究也开始涌现出来。

刘康[9]为了提高现有论文知识推荐的多样性与惊喜度，提出了一个基于不确定图的知识差异论文推荐算法，该算法是在离线数据的基础上进行验证的，首先根据论文语料库的“文档-主题”矩阵生成知识图谱，并在加入概率模型之后形成不确定知识图谱，在这个基础上来针对用户背景知识和目标知识之间存在的差异进行学术论文的个性化推荐，虽然最终的实验结果证明了该算法的可行性，但还缺乏真实用户的体验数据。邵弘智等[10]则实现了在线状态与离线状态的融合，在离线阶段将本体、语义词典和知识库融合在一起，构建了基于“主题网格-词簇-背景知识库”的跨领域知识图谱，用于挖掘文献的主题分布及背景知识特征；在在线阶段，通过建立模型实现用户与目标文献知识之间的关联，最终基于知识距离的度量方法来推荐文献。

在引文推荐领域，Ayala-Gómez等[11]提出了一种使用知识图谱来建立全球引文推荐的方法，通过使用知识图谱扩展来挖掘给定摘要中的语义特征，并将它们与其他特征组合以适应学习排名模型，最后通过这一模型来生成引文推荐；Huang等[12]提出了一个新的引用级联的概念，并定义为一个引文关系网络，该网络包括一篇论文与其所引用论文之间的引用关系，还包括所引用论文中的引用关系。

2.3 学术知识发现与语义出版

以往的学术搜索引擎都是基于关键词、主题、篇名、作者等独立的关键词进行检索，而忽视了文章内部知识单元间内在和隐含的关联，这使得在科学文献信息骤然增长的环境下，很难保证检全率与检准率，同时，随着科学出版物的指数增长，在线出版的文献资源越来越多，也给发现和获取科学文献中有用的学术知识带来了一些实际挑战。

Dong等[13]将学术知识发现和获取的问题视为众包数据库问题，并提出了一个用于学术知识发现和获取的混合框架，该框架通过从PDF文档中识别和提取知识单元及其关系，集成了人工的准确性和自动算法的速度来解决问题；Vahdati等[14]提出了一个知识驱动的框架——Korona，该框架与之前Traverso-Ribón等[15]提出的KOI知识发现方法不同的是，它不再局限于自我网络的实现，而是通过依靠语义相似度来发现知识图谱中隐藏的语义关系，进而揭示未知关系，在这一框架基础上构建起的学术知识图谱，成功揭示了合作者网络，而且通过扩展本体，还可以用于预测其他学术模式，例如共同引用或学术合作的关系等。

目前流行的在线出版方式像PDF格式在格式转换过程中会有乱码等不规范文本的出现，这对在线学术资源的智能检索带来了很大的影响，所以借助到本体、关联数据等语义相关的信息技术发展起来的语义出版相继产生，随着语义出版技术的不断成熟，基于学术文献的知识抽取和知识图谱构建也会跟着有更大的突破。语义出版与学术知识图谱的发展是双向互利的，语义出版可以帮助知识图谱实现语义丰富化[16]，同时知识图谱也可以促进语义出版的发展。

2.4 学术社交网络结构挖掘

学术社交网络是科研人员聚集的专业性网络社区，具有一般社交网络的基本属性，同时还兼具学术社交在基本功能、用户行为、服务模式等方面的特殊属性。身处社交网络中，每个用户都会不自觉地受到社交行为的影响，而挖掘用户的社交行为信息以及预测用户间社交行为的影响对社交网络平台的发展都是至关重要的。

一般社交网络中社交影响力的预测方法都是通过基于手动建立的模板来提取的，这些方法过于依赖专家的领域知识，Qiu等[17]在深度神经网络的启发下，设计了一个端到端的框架——DeepInf，来学习用户的潜在特征表示，将网络结构和用户特定功能纳入卷积神经网络和关注网络中，最后该框架被应用在了四个不同领域的社交网站上，包括开放学术图谱（OAG）、Digg、Twitter和微博，证明了DeepInf模型可以显著提高社交影响的预测性能，展示了社交网络信息挖掘任务的表示学习前景。

2.5 学者画像构建

通过构建学者画像，能够为预测用户的学术需求和行为提供可参考的直观描述，还可以提升智能检索和个性化推荐的效果。知识图谱的思想可以增强传统学者画像构建方法中关键词之间的相关性，还可以借此构建学者画像图谱，所以姚远等人[18]通过基于本体的方法为图书馆读者的学术行为构建了用户画像，并通过知识图谱的视角考察了用户画像的构建方法，具体将用户画像向量空间模型中的向量词与领域知识图谱中的概念相对应，将知识图谱中概念间的关系映射到用户画像中的向量词之间，来最终获得用户的学术画像本体。针对开放互联网中学者画像的研究中，Aminer系统[19]提出了一系列解决办法，该系统采用条件随机场作为标注模型对来自异构平台的学者信息进行统一

标注，在兴趣挖掘阶段采用LDA主题模型对抽取到的兴趣关键词进行聚类，以期找出用户的兴趣主题，在论文引用数预测方面，采用循环神经网络[20](Recurrent Neural Network, RNN)和长短时记忆单元[21](Long Short Term Memory, LSTM)来构建预测模型。

2.6 学术资源共享

科研机构与各大高校是在线学术资源的主要需求集体，而在这些集体中教师群体或科研人员群体又是学术资源的主要拥有者，随着跨学科研究的不断深入，不同学科之间的教师、科研人员合作也越来越频繁，为了将这些资源数字化，为了更方便不同学科之间学术资源共享，已有研究应用了知识图谱的相关技术。

Nonthakarn等[22]设计了一个将教师的档案信息与教学资源联系起来的元数据模式，该模式基于都柏林核心应用纲要的Singapore框架，集成了包括教师简介、课程、学术工作、奖励和活动在内的五个实体，为了通过LOD云链接所有的数据，还使用了本体的一些通用属性，像DC、BIBO、EVENT、FOAF、LOM和VIVO等，最后依据提出的模型开发了一个名为OpenTeacher的在线教师组合作原型，用于教师之间分享彼此的教学资源并相互合作。这一研究对更广范围的科研人员专业合作具有普适和推广的意义。

3 学术知识图谱构建过程

学术知识图谱的构建方式主要有自顶向下(top-down)与自底向上(bottom-up)两种，前者是在做好上层本体和数据模式层(scheam)的基础上，将从海量数据中抽取出实体信息存放到知识库中，后者的顺序则反之。具体学术知识图谱构建过程如图1所示。

图中学术知识抽取是从底层异构的数据集中自动抽取出学术资源实体、关系及属性；学术知识融合是对所抽取出的三元组进行共链、消歧，最终整合在一起组成一系列事实表达；学术知识加工是经过本体构建、知识推理和质量评估后，将事实转变为可用的知识。目前针对学术知识图谱构建过程的研究也主要集中在以下几个方面。

3.1 学术知识抽取

知识抽取是学术知识图谱构建过程中的第一步，这一步涉及的关键技术包括实体抽取、关系抽取和属性抽取。

3.1.1 实体抽取

实体抽取是自动从文本语料中识别出命名实体，

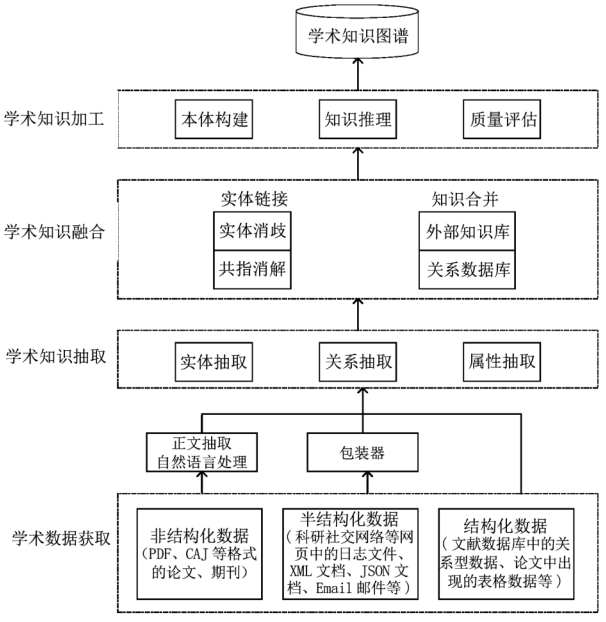

图1 学术知识图谱技术架构

实体抽取的效果将直接影响到最终知识图谱的质量，所以这一步是知识抽取中最基础也是最关键的一步，目前针对在线学术资源领域的实体抽取研究中，普遍采用的是语言学、统计学等方法。其中方俊伟等[23]通过基于先验知识TextRank的抽取算法来对文献类在线学术资源进行学术关键词的抽取；Khabsa等[24]通过使用多个提取器和输出概率来表示每个实体的置信度得分的方法来识别学术文献中的实体；除了文献类资源之外，还有大量的在线学术资源存在于各个学术网页上，由于这类资源结构更复杂、冗余度更高，所以相关研究侧重在提高实体抽取的精确性和准确度等上面，其中Yuan等[25]结合了机器学习方法和其他基于规则的方法，通过将网页分解为文本单元，然后使用分类器来确定单元的类型，最终根据类型来选择合适的技术进行实体抽取；Nie等[26]则通过在实体抽取中使用知识库里已有的结构化知识来提高实体抽取的精度。

实体抽取的单位除了关键词之外，还可以延伸到句子形式，就学术领域而言，很多情况下，一条完整的句子能够更好地表达一条知识的内涵。化柏林[27]通过句子级知识抽取的方法抽取出了学术论文中的方法知识元，并形成了方法知识元描述规则；王凯等[28]也是在句子级抽取方法的基础上，通过一种基于位置加权的核心知识挖掘方法，来抽取挖掘出能代表一个文本核心知识单元的句子；Hwang等[29]将词性(POS)标签标记学术期刊摘要部分的每个句子，通过分析各个标签之

间的搭配关系来抽取实体，并创建实体识别模型。

3.1.2 关系抽取和属性抽取

关系抽取和属性抽取是指在一个句子中识别实体对的语义关系和实体对应的属性，是将抽取出来的零散实体连接起来的语义纽带，这个过程在学术知识图谱的构建中起着重要的作用。目前关于在线学术资源中学术关系和属性的抽取方法中，已经开始从传统的依靠手动编制规则的抽取方法[30]向半监督以及无监督方法转变。其中Zheng等[31]借用了深度学习的思想，提出了一种基于神经网络的关系分类框架，它可以同时学习关系模式的信息和给定实体的语义属性，这一方法的提出弥补了当前大多数基于深度学习的方法主要集中在学习单个句子的语义表示而不能反映上下文关系的问题。

3.2 学术知识融合

知识融合是对经过多个数据源中抽取出的知识单元进行融合，知识融合涉及的技术主要包括实体链接和知识合并。

3.2.1 实体链接

实体链接是解决知识融合过程中经常会出现的实体歧义问题的一种技术，它能够实现将从文本中抽取得到的实体对象链接到知识库中对应的正确实体对象[15]。在在线学术资源领域的实体抽取中，也会经常出现同词异义或者同义不同词的现象，这一现象多发生在专家学者名与学术成果等信息的匹配问题上，针对这一问题，Cifariello等[32]通过实体链接成功地将基于文本数据的经典语言建模技术与维基百科知识图谱联系起来，提高了学术专家实体匹配的效果；武帅等[33]通过实体链接的方法融合了同一学者来自不同信息源的信息项，最终生成学者画像；Gao等[34]提出了从属关系名称和研究员名称消歧算法；邱爽[35]提出了一种基于高置信度特征属性的层次聚类方法和基于语义的作者相关话题模型。

3.2.2 知识合并

知识合并强调的是针对结构化数据（比如外部知识库和关系数据库）的整合，这一部分的数据相较半结构和非结构化数据处理起来更为容易，但也面临着不同数据库之间数据层和模式层不一致的问题。在针对在线学术知识库的合并过程中，Zhang Y等[36-37]通过使用自然语言处理技术（NLP）和众包技术，成功将Scopus摘要引文数据库中的数据合并进SKFM知识融合模型中，并在同年，在基于上述模型构建起的推荐系统中嵌入了知识融合和知识排序的模块，为学术知识图谱在帮助识别研究问题和相应解决方案方面提供了新的思路。

3.3 学术知识加工

经过知识抽取和知识融合之后得到一系列事实表达，需要经过知识加工之后才能得到最终结构化、网络化的知识体系。知识加工主要包括本体构建、知识推理和质量评估这三大技术。

3.3.1 本体构建

本体构建主要用于描述知识图谱的数据模式（Schema），在线学术资源中，大量可用的科学论文为本体构建的研究奠定了基础。其中González等[38]提出了一种支持大数据分析中知识管理的本体——BIG-OWL；Ren[39]将概念与人类专家的知识整合到本体构建中，并考虑了概念的时间属性；Zhu等[40]基于学术研究特征构建了本体FARO，该本体具有描述动态和静态研究特征以及建立社交网络的能力；温浩等[41]构建了一个“问题-方法-结果”三元组本体模型，并以句子为分析单位，以学术期刊文章为挖掘对象，实现了基于模式识别的文本知识点深度挖掘方法。以上这些研究都是基于文献资源的本体构建，结合了文献资源的一些特有属性，如专家学者、概念时间、研究特征等。

另外，还有一些相关研究侧重于利用本体构建语义框架。Gao等[34]构建了一个可以整合领域专家提供的分类法、术语以及从学术出版物的语料库中提取关系的本体框架；孙建军等[42]构建了一个面向学科领域学术文献的语义标注框架，该框架包括三部分：学术文献标注本体的构建、学科领域本体的构建和标注本体与领域本体的关联实例。由此可见，本体框架的搭建更有利于资源的整合。

除了可以基于文本类在线学术资源进行本体构建之外，图像资源中也存在着大量的语义信息，胡蓉等[43]通过构建VRAL（文内视觉资源）本体，实现了VRAL资源在底层视觉特征、高层语义特征、元数据三个层次上的描述与组织。

3.3.2 知识推理

知识推理能够从实体间发现新的关系，能够从已有的知识中发现隐含的知识，进而丰富和拓展已有的知识网络，是学术知识图谱补全的重要手段。

目前针对在线学术资源知识推理方面的研究还处于初期阶段，已有研究多集中在基于规则的推理上，其中Zheng等[44]重点研究了SWRL（Semantic Web Rule Language）的推理机制和实现策略，通过在SWRL的基

础上融合内容相似度计算来识别学术文献之间的内容关联，使本体中的语义联系得以扩展，推理结果更趋完善，同时提高了内容计算的准确率，对学术知识图谱、学术资源智能检索等应用获得高质量推荐结果起到了重要的作用。

3.3.3 质量评估

学术知识图谱的构建过程中很多环节都是机器自动完成的，不可避免地会存在很多错误信息，虽然学术知识图谱面对的是专业性较强的学术资源，但是也不能完全解决多语言、多平台、多种类型的异构数据所带来的冲突问题，所以进行质量评估在学术知识图谱的构建中也是相当重要的，有助于保证知识库的质量。

目前，针对在线学术资源知识库进行质量评估的方法有两类，一类是采用本体构建，其中晏归来等[45]以现有医学科技评价维度和科研本体为依据，构建了面向医学科技评价的本体模型，并抽取出其中的概念及属性；钱玲飞等[33,46]构建了学术创新力的概念本体，为学术创新力的自动测度提供了基础支持。另一类是采用评估函数，其中Fader等[47]在LDIF框架基础上提出的质量评估方法支持用户根据自身需求来定义评估函数，还可以通过对多种评估结果进行综合以得到最终的评分；而Knowledge Vault项目[48]则是根据指定数据信息的抽取频率对信息的可信度进行评分，然后利用从Freebase知识库中得到的先验知识对可信度进行修正。

4 知识图谱在在线学术资源应用中面临的问题与研究趋势

知识图谱是目前比较热也比较新的研究方向，在信息检索、人工智能、自然语言处理等学科领域中都有比较突出的成果，就具体的在线学术资源领域而言，由于学术知识具有较强的专业性和规范性，所以给学术知识图谱的构建带来了很多便利，然而随着当前在线学术资源涵盖范围的不断扩大，除了传统文献数据库之外，还有海量知识存在于众多学术社交网络等平台上，这些数据中有很多错乱、冗余的信息，为知识图谱的应用带来了很多阻力，所以今后相关研究仍面临着一些巨大的困难和挑战。

4.1 纯文本知识抽取面与方法有待拓展与改进

对论文、期刊等纯文本的抽取仍然存在很多没有解决的问题。首先，已有研究多是以文献的标题、摘要等部分为数据集进行知识抽取的，一定程度上会影响到结果的全面性和精确度，虽然这两部分本身就是一篇文献中心思想的集中体现，但有些细节还是要看全文才能发现；其次，现有的实体、关系抽取方法在针对非结构化学术资源上还存在很多不足，这也是当前科研社交网络中文本抽取还涉入未深的原因之一，科研社交网络中的交互文本信息相对传统的文献资源来说数量更多、结构更零散、语言更偏口语化、冗余度更高，但在时效性、新颖度方面却优势更明显，可以预见，对这方面的文本语义抽取进行研究会成为一个趋势；最后，由于基于以上这两类文本进行知识抽取和知识图谱构建的代价很大，需要花费的时间、经济成本很高，所以加入专业词典的参考会减少很多工作量，然而目前虽然像医学、农学等学科已经有做得比较权威的专业词典，但总体来看数量还很匮乏，对大规模学术知识图谱构建的帮助还远远不够，这也是每个专业领域学科建设层面需要考虑的宏观问题。

4.2 实体匹配精确度还有待提高

虽然学术领域与通用常识领域相比有较高的专业性，很多专业术语都有专门的名称，而且关系属性的描述也有相对明确的限制，但是对很多新兴的词汇和研究领域仍然存在多种命名的情况，除此之外同一个学者的个人信息在不同平台上也会存在差别，这些问题的存在都直接影响到了最终实体匹配环节的效果，虽然通过实体消歧和共指消解可以解决其中一部分，但考虑到学术研究的严谨性，所以同样是针对类似一词多义和信息缺失的问题，在学术领域还需要学科专家的介入，以确保最终实体匹配的科学性和准确性。

4.3 文本语义理解技术上还需要突破

在线学术资源的类型中，文本型数据占很大比重，然而，虽然知识图谱在文本型数据的语义理解上很有优势，但也深受数据冗余等问题的诟病，目前已经在视频、音频、图片等多媒体资源语义理解上有了很大突破的深度学习技术，吸引了很多研究者的注意，同样是分析语义信息，将深度学习与知识图谱技术结合，或许可以帮助学术知识图谱乃至更广范围的大型通用知识图谱在文本语义理解上得到质的飞越。

4.4 基于中文的学术知识图谱构建还很缺乏

作为目前学术知识图谱产品的代表，不管是国外的微软学术图谱还是国内的Aminer系统都是在基于英文文献和其他英文在线学术资源的基础上建立起来的，知识图谱的构建技术与自然语言处理密切相关，由于受到语种的限制，基于英语语料构建的一系列方法

并不能直接移植到其他语言的知识图谱上，而且中文表达的灵活性与特殊性导致当前针对中文进行的自然语言处理仍然面临很多挑战，这也是目前国内在中文学术知识图谱构建上研究成果并不是很多的一个重要原因，因此，解决问题应该从源头出发，进一步加快对中文语义关系和实体抽取自动化的研究，进一步完善和丰富知识库中中文在线学术资源实体关系集的数量，这些工作对今后该领域的深入发展将会有非常重要的意义。

4.5 知识图谱在学术信息服务领域的应用还有待深入开发

在学术信息服务领域中，基于知识图谱的应用已经开始有向多方面拓展的趋势发展，但目前应用比较好的还是主要局限在文献检索和学术资源推荐等服务上面，像学术知识发现、学术信息可视化分析等智能应用还处于探索阶段。随着相关技术的成熟，未来可以朝着精准描绘某一研究领域的发展脉络、分析论文引用模式以及预测新的研究热点等方向迈进，以促进在线学术资源利用的最大化，并帮助不同需求的学者在短时间内获得更多的学术信息服务。

5 结语

海量在线学术资源为学术研究工作和学术探讨提供了充足的研究基础，但庞大的数量中也存在着大量冗余和错误的信息，这些问题无形中为科研人员搜寻各自需要的知识增加了很多时间和精力成本，由于目前众多在线学术资源类型中主要以文本类为主，而且知识图谱在文本语义理解和分析上具有很强的优势，所以近年来有很多研究开始将知识图谱应用到在线学术资源中，以减少当前在线学术资源的信息压力，从而为科研人员的科研工作带来便利。目前，在科学文献推荐、学术知识图谱构建、学术知识发现、学术社交网络、学者画像和语义出版等方面都有了积极的尝试，但还处于初期阶段，本文对在线学术领域知识图谱的应用现状进行了分类分析，还对已有研究中存在的问题进行了阐述，并给予了展望，以期可以为今后该领域的研究提供借鉴意义。

参考文献

[1] AMIT S. Introducing the knowledge graph[R]. America: Official Blog of Google, 2012.

[2] 中国中文信息学会语言与知识计算专委会. 知识图谱发展报告(2018)[EB/OL].[2019-06-26]. http://cips-upload.bj.bcebos.com/KGDevReport2018.pdf.

[3] Huang S, Wan X. AKMiner: Domain-specific knowledge graph mining from academic literatures[C] .International Conference on Web Information Systems Engineering. Springer, Berlin, Heidelberg, 2013: 241-255.

[4] 秦 玥. 面向创业领域科技论文的知识图谱构建与应用研究[D]. 长春: 吉林大学,2018.

[5] Sadeghi A , Lange C , Vidal M E , et al. Integration of scholarly communication metadata using knowledge graphs[C]. International Conference on Theory & Practice of Digital Libraries. Springer, Cham, 2017.

[6] 汤 庸,陈国华,贺超波,等. 知识图谱及其在学术信息服务领域的应用[J]. 华南师范大学学报(自然科学版),2018,50(5): 110-119.

[7] 张 晔,贾雨葶,傅洛伊,等. AceMap学术地图与AceKG学术知识图谱——学术数据可视化[J]. 上海交通大学学报,2018,52(10):1357-1362.

[8] Qiao B, Fang K, Chen Y, et al. Building thesaurus-based knowledge graph based on schema layer[J]. Cluster Computing, 2017, 20(1): 81-91.

[9] 刘 康. 基于不确定图的知识差异论文推荐算法研究[D]. 哈尔滨: 哈尔滨工程大学,2017.

[10] 邹弘智,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.

[11] Ayala-Gómez F, Daróczy B, Benczúr A, et al. Global citation recommendation using knowledge graphs[J]. Journal of Intelligent & Fuzzy Systems, 2018, 34(5): 3089-3100.

[12] Huang Y, Bu Y, Ding Y, et al. Number versus structure: towards citing cascades[J]. Scientometrics, 2018, 117(3): 2177-2193.

[13] Dong Z, Lu J, Ling T W, et al. Using hybrid algorithmic-crowdsourcing methods for academic knowledge acquisition[J]. Cluster Computing, 2017, 20(4): 3629-3641.

[14] Vahdati S, Palma G, Nath R J, et al. Unveiling scholarly communities over knowledge graphs[C] .International Conference on Theory and Practice of Digital Libraries. Springer, Cham, 2018: 103-115.

[15] Traverso-Ribón I, Palma G, Flores A, et al. Considering semantics on the discovery of relations in knowledge graphs[C] .European Knowledge Acquisition Workshop. Springer, Cham, 2016: 666-680.

[16] 许 鑫,江燕青,翟姗姗. 面向语义出版的学术期刊数字资源聚合研究[J]. 图书情报工作,2016,60(17):122-129.

[17] Qiu J, Tang J, Ma H, et al. Deepinf: social influence prediction with deep learning[C] .Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, 2018: 2110-2119.

[18] 姚 远, 张 蕙, 郝 群, 等. 基于本体的用户画像构建方法

[C]. 中国计算机用户协会网络应用分会,2018:226-232.
[19] 袁 莎,唐 杰,顾晓韬.开放互联网中的学者画像技术综述[J].计算机研究与发展,2018,55(9):1903-1919.
[20] Pascanu R, Mikolov T, Bengio Y. On the difficulty of training recurrent neural networks[C] .International Conference on Machine Learning. 2013: 1310-1318.
[21] Xiao S, Yan J, Yang X, et al. Modeling the intensity function of point process via recurrent neural networks[C] .Thirty- First AAAI Conference on Artificial Intelligence. 2017.
[22] Nonthakarn C, Chawuthai R, Wuwongse V. An application profile for linked teacher profiles and teaching resources[C] .International Conference on Asian Digital Libraries. Springer, Cham, 2014: 138-148.
[23] 方俊伟, 崔浩冉, 贺国秀, 等. 基于先验知识TextRank的学术文本关键词抽取[J]. 情报科学, 2019, 37(3):77-82.
[24] Khabsa M, Giles C L. Chemical entity extraction using CRF and an ensemble of extractors[J]. Journal of Cheminformatics, 2015, 1(7): 1-9.
[25] Yuan P, Li Y, Jin H, et al. Self-adaptive extracting academic entities from World Wide Web[C] .2015 IEEE Conference on Collaboration and Internet Computing (CIC). IEEE, 2015: 270-277.
[26] Nie Z, Wen J R, Ma W Y. Statistical entity extraction from the web[J]. Proceedings of the IEEE, 2012, 100(9): 2675-2687.
[27] 化柏林. 学术论文中方法知识元的类型与描述规则研究[J]. 中国图书馆学报, 2016, 42(1):30-40.
[28] 王 凯, 孙济庆, 李 楠. 面向学术文献的知识挖掘方法研究[J]. 现代情报, 2017(5):49-53.
[29] Hwang S, Hong J, Nam Y. Towards effective entity extraction of scientific documents using discriminative linguistic features [J]. KSII Transactions on Internet and Information Systems (TIIS), 2019, 13(3): 1639-1658.
[30] Kambhatla N. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations[C] . Proceedings of the ACL 2004 on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2004: 22.
[31] Zheng S, Xu J, Zhou P, et al. A neural network framework for relation extraction: learning entity semantic and relation pattern [J]. Knowledge-Based Systems, 2016, 114: 12-23.
[32] Cifariello P, Ferragina P, Ponza M. Wiser: a semantic approach for expert finding in academia based on entity linking[J]. Information Systems, 2019, 82: 1-16.
[33] 武 帅, 罗 威, 钱 旭, 等. 基于文献大数据分析的人才创新能力感知方法研究[J]. 情报理论与实践,2018, 41(12):45-49.
[34] Gao Z, Gui Y, Zhu M, et al. An academic search and analysis prototype for specific domain[C] .Asia- Pacific Web Conference. Springer, Berlin, Heidelberg, 2012: 171-178.
[35] 邱 爽. 学术论文同名作者消歧问题研究[D]. 武汉: 湖北大学,2016.
[36] Zhang Y, Saberi M, Chang E. A semantic-based knowledge fusion model for solution-oriented information network development: a case study in intrusion detection field[J]. Scientometrics, 2018, 117(2): 857-886.
[37] Zhang Y, Saberi M, Chang E, et al. Solution and reference recommendation system using knowledge fusion and ranking [C] .2018 IEEE 15th International Conference on e-Business Engineering (ICEBE). IEEE, 2018: 31-38.
[38] Barba-González C, García-Nieto J, del Mar Roldán-García M, et al. BIGOWL: knowledge centered big data analytics[J]. Expert Systems with Applications, 2019, 115: 543-556.
[39] Ren F. Learning time-sensitive domain ontology from scientific papers with a hybrid learning method[J]. Journal of Information Science, 2014, 40(3): 329-345.
[40] Zhu H, Zeng Y, Yang Y. Research topics variation analysis and prediction based on FARO and neural networks[C] .2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2016: 000910-000915.
[41] 温 浩,温有奎,王 民.基于模式识别的文本知识点深度挖掘方法[J].计算机科学,2016,43(3):279-284.
[42] 孙建军,裴 雷,蒋 婷.面向学科领域的学术文献语义标注框架研究[J].情报学报,2018,37(11):1077-1086.
[43] 胡 蓉, 唐振贵, 朱庆华. 混合需求驱动的文内视觉资源移动视觉搜索框架[J]. 情报学报, 2018(3):285-293.
[44] 聂 卉.结合逻辑推理与内容计算实现面向学术网络的智能检索[J].现代图书情报技术,2013(1):22-29.
[45] 晏归来,安新颖,范少萍,等.面向医学科技评价的本体模型构建研究[J].中华医学图书情报杂志,2018,27(10):1-7.
[46] 钱玲飞,张吉玉,汪 荣,等.基于领域知识的学术创新力测度本体构建研究[J].现代情报,2019,39(5):30-37.
[47] Fader A, Soderland S, Etzioni O. Identifying relations for open information extraction[C] .Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011: 1535-1545.
[48] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion[C] .Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2014: 601-610.

[作者简介] 熊回香,女,1966年生,华中师范大学信息管理学院教授,博士生导师。
景紫薇,女,1994年生,华中师范大学信息管理学院硕士研究生。
杨梦婷,女,1996年生,华中师范大学信息管理学院硕士研究生。
收稿日期:2019-09-16