

引用格式: 陆锋, 诸云强, 张雪英. 时空知识图谱研究进展与展望[J]. 地球信息科学学报, 2023, 25(6): 1091-1105. [ Lu F, Zhu Y Q, Zhang X Y. Spatiotemporal knowledge graph: Advances and perspectives[J]. Journal of Geo-information Science, 2023, 25(6): 1091-1105. ] DOI: 10.12082/dqxxkx.2023.230154

# 时空知识图谱研究进展与展望

陆 锋<sup>1,2,4,5\*</sup>, 诸云强<sup>1,2,4</sup>, 张雪英<sup>3,4</sup>

1. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101; 2. 中国科学院大学, 北京 100049; 3. 南京师范大学 虚拟地理环境教育部重点实验室, 南京 210023; 4. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023; 5. 政务大数据应用省部共建协同创新中心, 福州 350003

## Spatiotemporal Knowledge Graph: Advances and Perspectives

LU Feng<sup>1,2,4,5\*</sup>, ZHU Yunqiang<sup>1,2,4</sup>, ZHANG Xueying<sup>3,4</sup>

1. State Key Lab of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, Nanjing 210023, China; 4. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 5. Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou 350003, China

**Abstract:** The continuous generalization of geographic information poses a huge challenge to the classic geographic information analysis modes. Networked knowledge services will gradually become a new mode for geographic information applications, facilitating to transform the form of geographic computing into social computing. Geographic knowledge services need to connect people, institutions, natural environments, geographical entities, geographical units and social events, so as to promote knowledge assisted data intelligence and computational intelligence. Facing the urgent need for spatiotemporal knowledge acquisition, formal expression and analysis, this paper firstly introduces the concepts and characteristics of spatiotemporal knowledge graph. The spatiotemporal knowledge graph is a directed graph composed of geographic spatiotemporal distribution or geo- locational metaphors of knowledge that is a knowledge graph centered on spatiotemporal distribution characteristics. Secondly we proposes a research framework for spatiotemporal knowledge graph. The framework includes various levels from multimodal spatiotemporal big data to spatiotemporal knowledge services that contain ubiquitous spatiotemporal big data layer, spatiotemporal knowledge acquisition technique layer, spatiotemporal knowledge management layer, spatiotemporal knowledge graph layer, software/tools layer, and industrial application layer. Thirdly this paper introduces relevant research progress from text implied geographic information retrieval, heterogeneous geographic semantic web alignment, spatiotemporal knowledge formalization and representation learning. Combined with application practice, we then enumerate the construction and application approaches of domain oriented spatiotemporal knowledge graph. Finally, it discusses the key scientific issues and technical bottlenecks currently faced in the research of spatiotemporal knowledge

收稿日期: 2023-03-27; 修回日期: 2023-04-19.

基金项目: 国家重点研发计划项目(2022YFB3904200, 2021YFB3900900); 国家自然科学基金项目(41631177, 42050101)。

[ **Foundation items:** National Key Research and Development Program of China, No.2022YFB3904200, 2021YFB3900900; National Natural Science Foundation of China, No.41631177, 42050101. ]

作者简介: 陆 锋(1970—), 男, 新疆乌鲁木齐人, 博士, 研究员, 主要从事地理空间智能、地理大数据挖掘、时空知识图谱研究。

E-mail: luf@lreis.ac.cn

graph. It is argued that in the era of large models, constructing explicit spatiotemporal knowledge graph and conducting knowledge reasoning to meet domain needs is still the only way for spatiotemporal knowledge services.

**Key words:** spatial-temporal knowledge; knowledge graph; GeoAI; knowledge modeling; spatial-temporal big data; information extraction; geo-semantic alignment; advances

**\*Corresponding author:** LU Feng, E-mail: luf@lreis.ac.cn

**摘要** 地理信息的不断泛化对经典的地理信息分析模式提出了巨大挑战,网络化的知识服务将逐渐成为地理信息应用的新模式,助力地理计算到社会计算的形态转变。地理知识服务需要打通人、机构、自然环境、地理实体、地域单元、社会事件之间的关联,促进知识辅助下的数据智能与计算智能。本文聚焦地理时空知识获取与形式化表达及分析的迫切需求,首先分析了时空知识图谱的基本概念与特征,认为时空知识图谱是指具有地理时空分布或位置隐喻的知识构成的有向图,即以时空分布特征为核心的知识图谱;然后提出了时空知识图谱的研究框架,该框架可实现时空大数据到时空知识服务应用的转变,包括泛在时空大数据、时空知识获取、时空知识管理、时空知识图谱、软件系统及行业应用等多个层次;接着从文本描述地理信息抽取、异构地理语义网对齐、时空知识表达与表示学习等方面,介绍了相关研究进展;结合应用实践,介绍了面向行业的时空知识图谱构建与应用途径;最后,讨论了时空知识图谱研究目前面临的关键科学问题与技术瓶颈,提出在大模型时代,构建显式的时空知识图谱,并针对行业需求开展知识推理,仍是时空知识服务的必由之路。

**关键词** 时空知识;知识图谱;地理空间智能;知识表达;时空大数据;信息抽取;地理语义对齐;研究进展

## 1 引言

随着地球观测与导航及信息通讯技术的迅猛发展,各种显式和隐式的时空数据爆发式增长,已经成为泛在地理时空大数据的重要组成部分。泛化的地理时空大数据处理技术,超越了传统测绘/地理/数据库技术范畴,使得地理信息科学与技术(GIS)向广义化、社会化、知识驱动方向发展,地理空间智能(Geo-spatial Artificial Intelligence, GeoAI)成为研究热点,推动了城市信息学、计算社会科学的涌现,并融入电子商务、文化传媒、体育竞技等行业。GIS的发展目标也从聚焦地理空间实体/过程的几何形态过渡到侧重语义关联,从信息服务过渡到网络化的知识服务,从地理计算过渡到社会计算。

GeoAI是地理空间科学与人工智能相结合的交叉研究方向,通过研究与开发机器的空间智能,提升对于地理现象和地球科学过程的动态感知、智能推理和知识发现能力,并寻求解决人类和地球环境系统相互作用中的重大科学和工程问题<sup>[1]</sup>。GeoAI针对影像分类、目标探测、场景分割、仿真与插值、链接预测、基于自然语言的地理信息抽取与问答、实时数据集成、地理语义充实等应用需求,利用数据处理技术与数据体制文化的进步,支持更智能的地理信息以及方法、系统与服务的创建<sup>[2]</sup>。面对不断变化的复杂地理环境与社会系统,GeoAI需要打通人、机构、自然环境、地理实体、

地域单元、社会事件之间的关联,促进知识辅助下的数据智能与计算智能。当前流行的大数据工程化驱动的计算智能技术,能力的提升主要依赖于数据规模和计算速度的增长。而大数据“红利”效应正在逐渐减弱。计算智能技术的单点突破也难以以为大数据驱动的智能应用提供持续支撑<sup>[3]</sup>。脱离知识驱动的计算过程,智能化程度终究有限,需要从认知角度确立思维方式,实现知识汇聚与知识推理,才能充分发挥已有经验知识的价值。

GeoAI的成功应用需要系统化、形式化、规范化的地理时空知识支撑。一方面,大量隐藏在结构化或非结构化的地理时空大数据中的新知识等待发现;另一方面,GeoAI需要与地理时空分布相关的可解释、可复用、可推理的知识作为基础设施,辅助人工智能理解地理空间规律。目前基础地理服务普遍存在着“数据海量、信息爆炸、知识难求”现象,原因主要在于以往业界重点关注基础地理数据的空间载体作用,虽然研发了元数据查询、数据浏览下载、按需专题制图、API调用等服务功能,但对基础地理数据的知识存量挖掘不够,提供的地理空间知识服务极少<sup>[4]</sup>。

在此背景下,时空知识图谱应运而生。时空知识图谱作为一种高效的时空知识组织和表示方式,可为上述问题提供解决方案。时空知识图谱的构建过程,即如何自动化地探测地理实体间的时空关系与语义关系,实现地理信息的自动聚合过程<sup>[5]</sup>。时空知识图谱是地理信息科学领域的前沿科学问

题,也是提升时空大数据价值,促进地理信息服务能力和产业发展的必然要求。

## 2 时空知识图谱概念与特征

### 2.1 知识的形成与传承

知识这一概念源于人类不断递进的认知过程,一般遵从数据(Data)-信息(Information)-知识(Knowledge)-智慧(Wisdom)的金字塔模型(DIKW)<sup>①</sup>,如图1所示。数据是使用约定俗成的关键词,对客观事物数量、属性、位置及相互关系的抽象表示。信息是指具有时效性、有一定含义,有逻辑、经过加工处理、对决策有价值的数据流。知识是不断凝练并与已有知识库进行结构化整合的有价值的信息。智慧是基于已有知识,针对客观世界运动过程中产生的问题,根据获得的信息进行分析、对比、演绎,进而找出解决问题的能力。

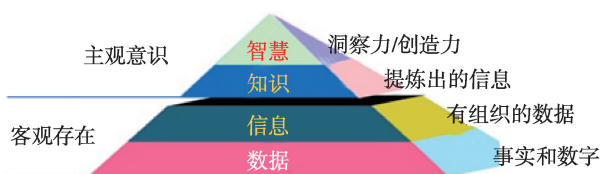


图1 人类认知过程的金字塔模型

Fig. 1 DIKW pyramid model for human cognition

更详细地,知识可以定义为:在人类感知世界和实践过程中,对已获取信息进行提炼、归纳,并认为正确的、具有普适性的认识、观点、经验、方法、技能和流程等,是人类探索物质世界和精神世界逐渐累积形成的结果总和。知识包括事实、信息的描述或在教育和实践中获得的技能。知识分类方法很多,可参考图书馆学与知识工程相关文献。从认知心理学角度,知识可分为描述型知识和操作型知识。描述型知识用来回答事物“是什么”、“怎么样”的问题,与人们日常使用的知识概念内涵较一致,包括事实、概念和原理,可以通过语言传授(如“珠穆朗玛峰高程8 848.86 m”、“早高峰由于通勤交通流剧增造成主干道拥堵”等)。描述型知识通过符号系统传播,包括语言、图形、图像、文字(竹简、石碑、书籍、网络文本等各种文字形式)、地图(地表事物和现象分布情况)、数学公式、音视频等,具有隐性、动态、时效性、主观、可复

制、可拓展、情境相关等基本特征。操作型知识是关于开展某项工作必需的知识和技能,一般须通过实践才能获得(如机动车驾驶技术等)。本文所述知识主要指描述型知识。

### 2.2 知识的形式化表达需求

知识表达形式多样,但绝大多数知识表达是服务人类学习的需求,不是服务机器理解的需求。故而知识的数字化过程多限于存储,以便人类用户可以更便捷地检索和学习知识。知识服务的内涵模糊,从图书馆到网络百科、文献共享,乃至基于已有知识的信息推荐,都可称为知识服务。目前各种流行的知识服务系统,大多是从各种知识资源中提炼内容,建立知识网络,为用户提出的问题提供已有知识资源或解决方案的信息服务过程。这些知识服务过程本质上是已有信息汇聚与人机交互查询过程。随着科学研究范式发展到数据驱动型,即科学研究的第四范式,使得可以在不依靠模型和假设的情况下,直接通过数据分析,发现很多以往的研究方法无法发现的新模式和新规律。然而第四范式在揭示事物本质规律方面存在固有的局限性,单独靠第四范式的数据驱动方法,不仅要消耗大量的计算资源,也难以真正预测未来的趋势与变化。目前的计算智能在存储、搜索、感知、确定性问题求解等方面性能表现优越,但在高级认知和复杂问题决策方面与人类智能相差很远<sup>[3]</sup>。从方法论来看,知识发现过程所需的人工智能,终究需要形式化的、计算机可读的知识体系支撑,实现深度学习与知识辅助的有机融合。因此,从早期的专家系统,到当前流行的知识图谱,业界一直在努力推动知识的形式化表达,以便让计算机能够充分利用已有知识,实现知识的自动推理。

知识图谱(Knowledge graph)概念由谷歌公司于2012年提出,旨在实现更智能的搜索引擎,并且于2013年以后开始在学术界和工业界普及。知识图谱是通过有向图的方式表达实体、概念及其相互之间语义关系的数据组织形式或产品,本质上是一种语义网络(Semantic network)。知识图谱中的“图谱”(Graph)源自图论,其中节点代表实体或者概念,边代表实体/概念的属性或者彼此之间的语义关系<sup>[5]</sup>。业界开发的通用知识图谱产品包括YAGO、DBpedia、Freebase、NELL、Zhishi.me等。这些知识

① <https://baike.baidu.com/item/DIKW模型/50904093>。



图谱通过本体(Ontology)、资源描述框架(Resource Description Framework, RDF)等方法实现非结构化或半结构化知识的表达,将人类已有的知识系统中的海量知识点,存储在数据库中,并开发一系列搜索引擎实现访问与查询,辅助网络搜索问答、客服机器人、智能推荐等。

严格意义上,“Knowledge Graph”一词译为“知识图”更合适,因为一般而言“谱”多具有“序列”的内涵,如电磁波谱、基因图谱等。虽然 Knowledge Graph 中实体或概念之间的逻辑关系结构也具有多尺度和可追溯性,但“序列”内涵较弱。然而,知识图谱的译称已得到业界广泛接受,在此不再进行严格界定。

### 2.3 时空知识图谱

时空知识图谱,顾名思义,是具有地理时空分布或位置隐喻的知识构成的有向图,即以时空分布特征为核心的知识图谱(Spatial-temporal Knowledge Graph, 或 Geo-related Knowledge Graph, 以下简称 GeoKG)。GeoKG 旨在通过计算机规范化表达与存储与地理时空分布相关的知识集合,进而支持地理时空分布或位置相关知识检索与知识推理。其中的图(谱)指地理时空分布或位置要素(概念、人类群体/个体、地理实体、事件、现象等)之间的语义联系形成的网状拓扑结构。

以往专业领域的关注点更多在于地理或地学专业领域知识,因此,常采用地理知识图谱或地学知识图谱的术语<sup>[5-7]</sup>。在追求自然系统、社会系统统一综合描述的背景下,涉及人类群体/个体、交通工具、建筑楼宇、兴趣点(Point of Interests, POI)以及其他具有社会系统属性与地理空间分布隐喻的概念和实体,无疑也是知识的巨大承载体,应用需求广泛。因此,为了避免与专业化的地理知识或地学知识混淆,本文将时空知识更广泛地定义为一切具有时空位置和动态变化特征的相关知识。时空知识是人类对地理空间上人、物、事件、现象分布、演变过程和相互作用规律的认知结果。除了传统的地理/地学应用外,时空知识也是消费领域基于位置的服务(Location-based Services, LBS)、新型基础设施建设和智慧城市管理的核心支撑。时空知识也可以参考地理知识或地学知识分类方法,分为事实型、规则及控制型、决策型知识3种类型<sup>[6]</sup>,或者分为数据性、概念性和规律性

知识3个层次<sup>[7]</sup>。业界已有诸多研究介绍了时空知识图谱的基本构建流程<sup>[5-6,8-9]</sup>。

时空知识图谱是地理信息分析向纵深拓展的关键技术。例如,在信息提取方面,传统的遥感分类和信息提取算法思维,主要是基于数据特征,利用样本学习获取遥感信息的过程。这种算法思维导向的方法可扩展性欠佳,在大区域应用时需要不断地补充新知识,或者使用海量样本训练学习来弥补地表异质性造成的特征差异<sup>[7]</sup>。业界已充分认识到,缺乏已有专业知识的支撑,很难在海量遥感大数据中获取更新、更广、更深层次的地学知识。在遥感大数据时代背景与深度学习不断发展壮大的趋势下,多源异质的大规模遥感知知识图谱构建以及基于图神经网络的知识图谱学习与计算成为内在需求<sup>[10]</sup>。一方面,知识图谱可以构建遥感知知识模型与复杂遥感场景,为知识—数据驱动的数据挖掘提供知识基础。另一方面,利用知识图谱中丰富的语义关系,可以实现面向语义需求的智能检索、增强智能遥感服务。耦合知识图谱和深度学习的新一代遥感影像解译范式,可以承载不同类型和层次的地学知识,使其具有查询检索、知识推理、动态更新、知识拓展等功能,为新时代地学知识驱动下的遥感大数据高精度智能解译服务,有效提升遥感影像的解译性能,并最终实现地学知识的自动化精准和积累<sup>[7,11]</sup>。在数据集成方面,数字孪生和智慧城市管理等需要实现自然环境/社会环境的多维度刻画与实时更新,从不同维度感知人-地关系,这也是全息地图思想的核心,即以全要素、全信息、全内容、全视角来开展地理空间对象的描述、表达、分析和服 务,实现物理系统和社会系统要素在任何时间、任何地点的信息无缝链接与交互<sup>[12]</sup>。从数据集成的视角,本文认为全息地图可以理解为时空场景测度+知识图谱,即以位置和时空场景实体为载体的异构数据实时关联与聚合平台,强调物理环境变化(水土气生、建筑环境、声光电磁)、社会环境动态(Rich POI/人群流动、地理空间交互、社交关系)和移动对象轨迹/语义的综合集成。场景要素间的动态联系是全息地图的精髓,有助于将地图转化为地理或位置相关知识的表达工具。在数据服务方面,由于各种可获取数据异构性极强,需要在统一时空本体的支撑下,基于地理空间数据的多维度特征,建立地理时空大数据之间的高精度关联,以及面向

计算模型的地理空间数据的自动推荐和匹配<sup>[13-15]</sup>。在数据分析方面,随着机器学习技术不断成熟,其中的深度学习模型,依靠海量的训练数据和强大的算力,在人脸识别、机器翻译、网络问答等任务上性能已经超过人类。然而,深度模型均为面向具体任务的隐性模型,难以迁移,也较难解释。而知识图谱作为显性模型,目前在知识储备量上已经超过人类个体,可广泛适用于不同任务,且采用类似人类的思考方式,具有很好的可解释性。作为人工智能技术的两大驱动力,深度学习和知识图谱需要深度融合<sup>②</sup>,才能满足城市智能计算等诸多场景需求,即实现城市多源时空数据的实时学习与时空知识的自动化积累。这也是城市建设与运维体现智慧的核心所在。

需要说明的是,时空知识图谱和地学信息图谱(Geo-information Tupu)是截然不同的2个概念。地学信息图谱是《地球信息科学学报》创刊人陈述彭院士提出的重要学术思想,源于中国历史早已有之的图谱概念。图谱是一种运用图形语言来进行表达与分析的方式。地学图谱是应用于地学分析的系列多维图解。其中图指地图,描述区域、现象的空间范围和分布格局,反映各地地理要素的空间配置和相互作用;谱指可从时间上追根溯源的动态系列表达,描述对象的发展方向与演化过程,大都以树状结构图表描述。图谱合一,则是空间与时间动态变化的统一表述。随着GIS技术的发展,地学图谱自然演化为由征兆图、诊断图和实施图组成的地学信息图谱<sup>[16-18]</sup>。地学信息图谱是认知、方法和动态地图三者的综合与统一。认知是人类对地学对象在图形思维上的观察能力;方法是信息的获取和分析的手段;地图是信息的交互表达形式,它反映的是地学对象的规律<sup>[19]</sup>。

地学信息图谱源于地理学与地图学,强调基于专业知识图形化描述的形象思维。而时空知识图谱源于计算机科学,强调基于知识表达形式化描述的逻辑思维。时空知识图谱的目标是辅助发现与利用新的与地理位置相关的知识。从这一点上看,就地学分析而言,时空知识图谱和地学信息图谱的目标是一致的。站在今天大数据、人工智能以及高分遥感等新技术的发展背景下,地学信息图谱就是通过信息图谱、知识图谱等方法,实现从数据到信息再到知识的跨越,进而发现和表达地学规律<sup>[19]</sup>。

### 3 时空知识图谱研究框架

从国家战略和行业需求的角度来看,自然资源规划利用、生态环境监管、智慧城市建设、公共健康监测等部门,正在全面推进行业治理体系和治理能力现代化,提升智能化管理能力和智慧化服务水平。然而,行业部门普遍存在数据资源整合集成、业务统筹协同难等问题。作为地理空间智能的基石,时空知识图谱是组织和利用大规模时空知识的有效手段。然而,现有的时空知识图谱主要依靠领域专家人工构建,存在行业领域知识不全面、不系统、不精准、更新困难等问题,迫切需要基于泛在时空大数据,在时空知识获取与管理技术支撑下,研发时空知识图谱管理和知识服务的基础软件,形成大规模、高质量的时空知识图谱,进而支撑行业领域时空知识服务应用。时空知识图谱总体研究框架如图2所示。

如何建立符合时空知识特点的时空认知与知识图谱表达方法,形成多维度的时空知识分类体系和统一的时空本体,发展顾及复杂时空特征及关系的时空知识图谱自适应表达模型,是时空知识组织管理、更新与计算推理、时空知识表示学习的理论基础。

时空知识蕴含在不同模态和类型的时空数据中,具有多粒度、异构、高维低密度、样本稀疏等特征,蕴含的时空知识内容、丰富程度和时效性各不相同,质量参差不齐。自动获取这些多模态、多类型的泛在时空数据,统一评估这些数据源所蕴含时空信息的质量,并在时空信息多维度特征和多层次关系综合认知指导下,实现基于统一时空本体的时空信息抽取与融合,为时空知识图谱构建与更新提供可靠语料,是时空知识图谱构建的数据基础。

现有时空知识图谱构建方法自动化程度极为有限,且存在时空知识本体与实例脱节、时空信息利用不充分问题,基于知识图谱的时空知识计算与推理研究更为罕见。亟需发展人工智能支持下群智协同的时空本体构建、基于海量开放语料和深度学习的时空知识图谱自动构建、“本体-实例”迭代融合的时空知识元组补全、时空知识表示学习与链接预测等技术,为大规模时空知识图谱更新与知识推荐提供技术支撑。

② 王仲远. 美团大脑 智享生活. <https://m.weike.fm/lecture/10038952> (访问时间 2023.01.26)。

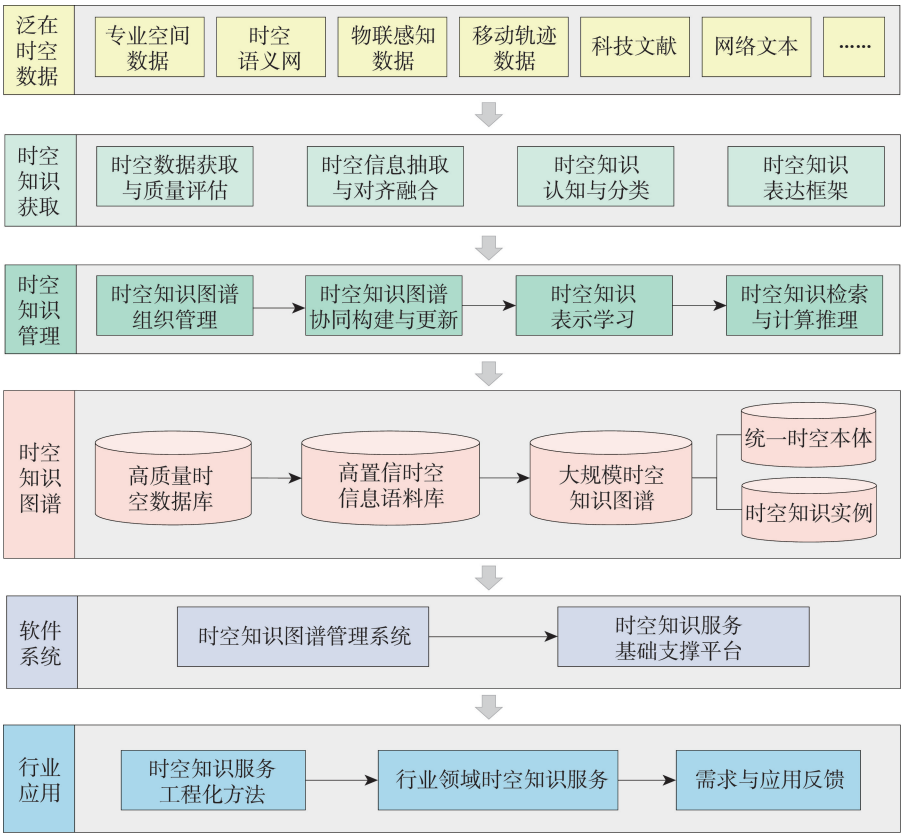


图2 时空知识图谱研究框架

Fig. 2 Framework of spatial-temporal knowledge graph research

4 时空知识图谱研究进展

文本是知识沉淀和传播的载体。各种形式的文本是目前构建时空知识图谱的主要数据源。当前的知识图谱研究也主要是围绕自然语言文本开展工作,以得到地理时空知识元组的规范化显式表达。其他形态的地理信息,如各种格式的专业空间数据、物联感知数据、移动轨迹数据等,在当前的知识图谱研究中,其知识发现结果也基本转化提炼为显式的文本(字符串)形式三元组,才可纳入知识图谱的知识计算过程。

近年来,在国家自然科学基金、国家重点研发计划等项目支持下,业界围绕自然语言文本蕴含地理信息抽取、异构地理语义网对齐、时空知识图谱构建、时空知识图谱领域应用等主题,开展了系统性研究工作。

4.1 地理信息标注语料库构建

地理信息标注语料库是开展文本蕴含地理信息提取与理解的重要基础,旨在为文本信息的时空

语义理解提供标准化的训练和测试数据。长期以来,标注语料库的匮乏一直是阻碍文本蕴含地理信息解析的关键因素。针对这一问题,业界设计了文本蕴含地理信息的标注规范,提出了地理实体及关系语料库自动构建方法<sup>[20-21]</sup>,弥补了人工定义标注体系缺乏可扩展性的不足,为语料库快速扩充和开放式地理信息抽取提供了新思路。

(1)地理实体标注语料库迭代式构建

在文本蕴含地理实体识别与语义消歧基础上,可设计人机交互式标注平台,对文本进行地理实体标注和校正<sup>[20]</sup>。当标注的文本语料达到阈值时,将初始训练语料与标注语料进行融合,输入到地理实体识别模型中再次训练,优化模型参数,提升识别效果,同时将新增语料加入标注语料库;迭代上述步骤直至所构建语料库满足要求。结果显示,在基于相同的语料库情况下,这一方法较传统的基于条件随机场(Conditional Random Field, CRF)的地理实体标注方法,准确率、召回率、综合值均有提高。

(2)地理实体关系语料库自动构建

领域专家构建的地理实体关系标注规范通常



要求标注者深入理解上下文语义,且涉及复杂空间关系的标注。如果缺乏人工参与,计算机将无法理解标注规范,不能完成自动标注。此外,同一地理实体的不同表达形式在语料标注的多个环节均可能存在歧义,影响地理实体关系语料库质量。据此,业界提出了基于回标技术的地理实体关系语料库自动构建方法<sup>[21]</sup>,首先利用词嵌入模型对网络百科中的地理实体标签类别进行语义聚类,通过地理实体词条的百度信息框属性名称和属性值筛选出地理实体关系,形成地理实体关系分类体系;其次,如果地理实体词条信息框的属性值也为地理实体,则把记录转化为地理实体关系种子三元组。在实体消歧后,对句子集合和种子三元组集合进行主客体匹配和关系匹配,实现种子三元组映射到句子的定量化评价,筛选出最优匹配项作为地理实体关系回标结果。这种基于百科信息框的回标方法,充分利用百科平台中数据庞大且关系类型丰富的地理实体词条作为构建地理实体关系语料库的数据源,且构建的地理实体关系标注体系涵盖了语义关系和空间关系。

#### 4.2 文本蕴含地理实体识别与语义消歧

文本蕴含地理实体识别与语义消歧是将文本中涉及的地理实体名称进行切分并予以标记的过程,旨在识别出文本中表示地理实体名称的符号,消除“同名异体”或“异名同体”的地理实体歧义现象,获取地理实体对应的位置或空间分布信息<sup>[22-23]</sup>。文本蕴含地理实体识别与语义消歧可通过实体共现与主题建模方法实现<sup>[24]</sup>,也可通过基于规则的地名词典间地理实体匹配方法实现<sup>[25]</sup>。以往多采用隐马尔科夫模型、支持向量机、最大熵、条件随机场等机器学习方法开展地理实体识别<sup>[26]</sup>。近年来,深度学习方法被引入地理实体识别领域,以减少人工方式的地理实体描述语言特征归纳过程,或突破字符串相似性度量指标难以处理由于语言和文化差异形成的地名字符变更问题,例如基于深度神经网络(Deep Neural Network, DNN)的方法<sup>[27]</sup>、基于深度信念网络(Deep Belief Network, DBN)的方法<sup>[28]</sup>。DNN方法采用循环节点来表达与待匹配字符串对应的字节序列,并将其与前馈节点相结合,从而决定地名对是否匹配。方法显著优于单一相似度匹配指标及基于监督机器学习的多重指标组合方法<sup>[27]</sup>。

DBN方法通过DBN模型的深层次网络结构,从输入数据中解析更易分类的高维度语言特征,计算每个字符属于地理实体名称组成部分的概率,并采用相邻字符组合方法,得到输入数据中描述地理实体信息的字符串。研究结果显示DBN模型与条件随机场模型的地理实体识别结果具有显著互补性,将二者进行融合有望大幅度提升地理实体识别性能<sup>[28]</sup>。文本中地理实体描述具有歧义现象,同一地理实体可能对应多个地理位置,同一位置也可能有多个地理实体描述。此外,文本中的地理实体描述还存在显著的离散化和稀疏性特征。针对这一问题,可基于已有的开放型百科知识平台(如维基百科、百度百科等)和自然语言词向量方法,进行地理实体语义消歧<sup>[29]</sup>。

#### 4.3 文本描述地理实体空间关系抽取

地理实体之间具有密切的空间关系。文本中的地理实体空间关系描述是对地理实体间精确空间关系的抽象化。自然语言理解领域中的关系抽取研究,以往主要集中于语义关系、因果关系、时态关系抽取等,针对空间关系抽取的研究较少<sup>[30]</sup>。然而,文本蕴含地理实体空间关系抽取对地理位置相关应用至关重要,包括空间查询<sup>[31]</sup>、空间参考<sup>[32]</sup>、数据预测<sup>[33]</sup>等。

已有的空间关系抽取方法可分为传统机器学习方法<sup>[34-35]</sup>和神经网络方法<sup>[36-37]</sup>。前者高度依赖人工特征或明确的句法结构,后者基于神经网络,通过依存句法分析产生候选关系,并利用双向长短时记忆网络模型(Bi-directional Long Short-Term Memory, BiLSTM)进行分类,或者利用预训练自然语言模型BERT和CRF模型抽取空间关系。这些方法本质上均属于分类任务,对非空角色关系抽取表现优异,但由于信息缺失的限制,难以胜任空角色(null-role)关系抽取,即难以处理空间位置关系三元组中元组要素缺失的情况。因此,业界提出将生成模型和分类模型相结合,利用分类模型抽取非空角色关系,利用生成模型引入先验知识捕获空角色关系语义,并使得结果具有更好的可解释性<sup>[30]</sup>。但这种方法在非地理实体之间动态空间关系抽取方面可能有效,对于地理实体之间静态空间关系抽取的有效性尚需验证。此外,也有研究提出融合地理语义的文本描述地理实体空间关系抽取模型<sup>[38]</sup>,将地理实

体类型与空间关系特征词两项地理语义与深度学习模型融合,以提高空间关系抽取结果的准确率与召回率,并提升模型对拓扑、方向、距离等空间关系分类的泛化能力。

随着网络文本爆发式增长,预定义的关系类型难以全面覆盖不断变化的地理实体间新型关系,限制了地理语义理解和空间推理。无监督学习方法具有捕获文本的动态变化特征并发现新增关系类型的能力,对动态扩展时空知识图谱中知识三元组类型极具潜力。例如,有学者提出了语境增强的无监督地理实体关系抽取方法<sup>[39]</sup>,首先进行地理实体识别,借助分类知识库确定实体类别,然后基于预训练BERT词向量知识库合并上下文中语义相似的词汇,增强相同类别的地理实体对的上下文语境,最后利用给定的句法特征排序增强语料中各词语的重要性,确定给定地理实体对的关系关键词。结果显示,相较于Domain Frequency和Entropy词频统计的抽取方法,语境增强的无监督地理实体关系抽取方法的关键词提取准确率大幅提升。同时,该方法具有识别新增关键词的能力,可有效扩展地理知识三元组的关系种类。业界进一步利用Stanford OpenIE工具实现了网络文本蕴含地理实体开放关系的实时抽取<sup>[40]</sup>,基于通用知识库的地理实体开放关系过滤方法,从OpenIE抽取的地理实体开放关系中筛选出高质量结果。相比OpenIE的原始抽取结果,该方法大幅度滤除了开放关系抽取结果中的噪声,并最大程度召回正确的地理关系三元组。

#### 4.4 文本描述地理事件信息抽取

不同类型地理事件在文本描述中差异显著,而且地理事件往往涉及多个地理实体及空间位置。针对这一问题,有学者将地理实体关系识别引入事件抽取过程,采用CRF模型和支撑向量机(Support Vector Machine, SVM)模型进行角色和语义关系标注,并以交通为例,通过正则表达式匹配识别事件的时间要素,实现了社交媒体文本中涉及多个地理实体描述的交通事件位置信息抽取<sup>[41]</sup>。此外,文本中存在大量多种类型地理事件表达共存的现象。基于触发词的方法容易将描述近似但类型不同的地理事件错误分类,而使用传统的文本全局特征向量作为监督分类方法的输入,模型学习的特征大多数与事件主题无关,导致训练分类器对

事件分类检测失效。针对这一问题,有研究以环境污染事件为例,提出了基于联合主题特征的文本蕴含地理事件检测方法<sup>[42]</sup>,首先计算文本中主题词频来构建主题分布特征向量,利用事件语料和维基百科语料训练的Word2Vec词向量对初始主题词库进行同义词扩展,然后基于TF-IDF向量表示文本的全局特征向量,计算词汇在各种事件主题类型的词频分布作为文本的主题分布特征向量,最后构建联合主题特征向量,进而采用逻辑回归模型与SVM分类模型提取了地理事件信息。这一方法兼顾文本的全局特征和主题分布特征,使用同义词林和知网改善了多类型地理事件信息检测分类效果。

地理事件具有时空动态变化特征。现有的模式匹配和机器学习算法很难迁移适用。针对这一问题,可在事件抽取过程中结合事件特征和语境特征<sup>[43]</sup>,考虑词汇的事件特征和上下文信息,采用句法分析快速标注事件语料,基于BiLSTM-CRF模型,从文本中抽取事件的时空属性信息,进而支撑地理事件知识库构建。

#### 4.5 文本描述地理空间范围计算

空间位置与几何形态共同刻画出地理实体、地理事件占据的空间范围。文本描述的空间范围具有显著模糊性。基于文本描述计算地理实体、事件的空间范围,需要明确地理空间位置的认知机制与描述方式,界定地理实体空间位置与几何形态的关系,将文本上下文中特定地理实体的空间关系描述,简化为二元空间关系的组合描述,实现地理实体、事件空间范围描述的近似计算。

针对不同语境中位置描述模糊性差异显著而难以定位的问题,有学者提出了基于超赋值语义的定位方法<sup>[44]</sup>,从多个层次分析位置描述中模糊性的来源,以及在各要素表达形式上的体现,然后从空间对象、距离关系和方向关系3个方面构建定位模型。通过对多个位置描述上下文进行超赋值,从上下文语义中获得观察值阈值,进而实现位置描述的精确化定位。

量化模糊空间边界有助于通过空间计算理解人类活动与地理空间的交互特征。业界提出了基于上下文空间关系的模糊地名空间建模方法<sup>[45]</sup>、基于核密度分析的地名模糊空间建模方法<sup>[46]</sup>、基于网络文本地理实体共现关系特征的认知范围方法<sup>[47]</sup>



等。上述方法充分利用了文本中的空间关系词汇信息,但是忽略了不同类型地理实体的特征差异。因此,也可以考虑采用地理时空知识图谱和上下文相结合的方法,利用知识图谱中相关地理实体的时空特征、属性和关系,扩大相关文本数据源,同时提升文本中相关信息获取性能,从而提高地理实体、事件空间范围计算精度。

#### 4.6 异构地理语义网对齐

互联网上有着大量开放的地理语义网(地理知识库)资源,可以作为构建时空知识图谱的重要基础。然而,这些地理语义网产品中不同类型的实体存在词法、结构和空间等方面的异质性,导致不同语义网中知识间的链接关系不可见,冗余数据难识别等问题。为此,业界开展了地理实体对齐形式化表达研究,针对不同应用场景提出了基于词嵌入增强的机器学习和基于空间映射变换的地理实体及类别对齐方法<sup>[48-51]</sup>,可以在地理实体缺乏空间参考或者信息不完整的情况下有效完成实体和类别对齐,并构建了异构地理语义网对齐框架,充分利用多维信息度量地理实体相似性,并集成不同类型的地理实体对齐任务,一次完成地理实体从概念、属性到实例的完整对齐。

##### (1) 词嵌入增强的地理实体对齐

现有文本蕴含地理实体对齐方法严重依赖地理实体的字符距离相似度,无法准确捕捉地理实体的深层次语义相关性,使得对齐性能难以提高。为此,业界提出了基于词嵌入增强的地理实体对齐方法<sup>[48]</sup>,借助大规模文本中丰富的词汇间上下文关系,通过训练和学习,基于词嵌入模型,将文本中的词汇表示为低维实数向量。该向量隐含了从文本中学习到的语义信息以及词汇间的深层次语义相关性。相比于基于字符距离的地理实体相似度,由于词向量隐含了语义信息,可以更准确地捕捉语义相关性,度量其相似程度。

##### (2) 基于空间映射变换的地理实体对齐

现有地理实体对齐方法通常计算实体名称相似度,利用绝对空间坐标来计算空间距离,并利用计算得到的实体名称相似度和空间距离来实现对齐。然而,矢量型地理实体可能存在缺乏空间参考或者信息不完整的情况,严重影响对齐效果。据此,业界提出了基于空间映射变换的地理实体对齐方法<sup>[49]</sup>,通过橡皮页空间映射方式校正待对齐的两

个数据集的相对坐标,使得数据集可以实现空间叠置,叠置后的数据集相对坐标可用于计算空间距离等,实现实体对齐。其中,空间映射通过基于控制点的橡皮页变换实现。控制点可通过数据集中部分地理实体对齐和对齐的实体计算得到。

##### (3) 基于词嵌入的地理实体类别对齐

针对已有的地理实体类别对齐主要基于字符或结构相似度方法进行,难以利用深层语义信息的问题,业界提出了地理语义增强的词嵌入地理实体类别对齐方法<sup>[50]</sup>。首先构建语料库,输入词嵌入模型进行训练,生成能够充分表达地理实体类别语义信息的词向量,然后根据生成的词向量,获得来自不同地理知识库的实体类别的相似度,根据相似度计算结果生成地理实体类别的对齐结果。这种方法可以有效度量类别间的相关性。需要注意的是,词向量维度提升虽然增强了向量对语义相关性的表示精度,提升了对齐结果,但过大的维度却可能导致查全率的降低。需要根据查准率和查全率的均衡确定合适的词向量维度。

##### (4) 异构地理语义网整体对齐框架

目前业界已有大量公开的地理语义网资源,如 GeoNames Ontology、OSM Semantic Network Network、GeoWordNet、LinkedGeoData 等。然而不同组织发布的地理语义网标准不同,存在一词多义和同形异义的现象,整体上分散独立。如何整合这些异构分散的知识资源成为亟待解决的难题。通用对齐系统大多缺乏实例对齐功能,需要建立集成框架以支持各种类型的对象对齐。此外,许多地理空间数据使用扁平式存储,缺乏明确的层级结构,和通用链接数据存在本质差异。独立研究各种类型的地理实体对齐方法,不但增加了计算资源,还缺乏相似性度量指标的自动整合机制,限制了对齐方法的可扩展性。鉴于此,业界构建了地理空间数据整体对齐框架,充分利用了多维信息度量地理实体相似性,集成了不同类型的地理实体对齐任务,可一次完成概念、属性到实例的匹配<sup>[51]</sup>。其中相似性计算考虑了地理实体的空间相似性、词法相似性、结构相似性和扩展信息相似性;采用初始匹配和增强匹配的两阶段匹配策略。相较于著名的集成对齐系统 PARIS,这一集成框架不仅利用了更多类型的信息来协同增强对齐的效果,而且聚合多种相似性度量方法时避免了人工调参,降低了方法对指定信息的依赖,适用于信息分布失衡的数据对齐。

#### 4.7 时空知识表达与表示学习

综合考虑时间和空间2种要素的独特性,业界提出了“以地理实体为核心”的时空知识表示模型,发展了基于描述逻辑语言的时空知识形式化描述方法,建立了时空知识图谱的统一表达框架,解决了“如何将时空知识表示为计算机可接受的符号和可存储数据结构”的问题<sup>[6,52]</sup>,并以网页文本为数据源,探索了基于机器学习模型的时空知识图谱自动生成方法<sup>[53]</sup>。针对现有知识图谱中地理实体间链接稀疏,无法为已有知识嵌入表示的学习过程提供充足训练数据,导致地理知识嵌入表示效果不佳的问题,提出了顾及地理距离特征的时空知识图谱嵌入表示方法<sup>[54]</sup>,结合时空知识图谱具有空间隐喻的独特性,将隐含的地理实体和地理关系间的空间距离模式引入知识嵌入表示学习过程,优化地理实体和关系在低维向量空间的表示效果,以更好地表达地理实体间稀疏链接的知识图谱三元组,服务时空知识图谱补全等应用场景。

#### 4.8 时空知识图谱应用案例

通用知识图谱作为现代搜索引擎的底层资源,已成为互联网信息服务的核心支撑。与地理位置相关的城市设施与人类个体/群体知识资源,如POI画像与消费者行为轨迹模式等,也已在智慧城市管理、城市电商和零售领域取得了广泛应用。针对行业需求的其他时空知识资源,尚处于应用起步阶段。

例如,针对台风灾害应急管理决策需求,业界提出了台风灾害知识图谱构建方法<sup>[55]</sup>,通过梳理台风灾害知识来源,建立了台风灾害事件知识体系,提出了台风灾害事件知识表达模型,实现了基于“事件-对象-状态-属性”的台风灾害事件知识融合,构建了台风灾害知识图谱,分析了知识图谱在台风灾害对象查询、灾情应对措施推理等场景中的应用途径;针对大气污染事件追溯和执法需求,提出了以事件为核心的大气污染执法知识本体模型,建立了大气污染发现、现场检测、处罚决定、整改监督4个阶段的事件本体,确定了本体涉及的对象要素的核心属性及其相互关系<sup>[56]</sup>;采用BERT和CRF模型组合,实现了大气污染事件分类及其文本中事件对象要素等的自动化精准抽取<sup>[57]</sup>,构建了大气污染执法事件知识图谱,开展了大气污染事件关联

性、发展态势、执法量裁准确度分析等;针对铁路隧道智能化施工管理需求,提出了铁路隧道钻爆法施工安全质量进度知识图谱构建方法<sup>[58]</sup>,根据铁路隧道施工过程中与安全质量进度关联的人机料法环5类关键要素的概念与语义关系,设计了模式层自上而下和数据层自下而上双向协同的构建方式,抽取了实体及关系并进行融合,构建了模式-数据关联的知识图谱,精细刻画了影响安全、质量和进度的关键要素属性、要素间语义关联关系以及互馈作用关系等,为铁路隧道施工安全质量进度管理提供了支撑;针对国土资源不同数据产品难以有效管理与快速应用的问题,利用图数据库对公开土地覆盖数据集进行语义层面的结构化存储,建立了可动态更新的中国国土资源知识图谱<sup>[59]</sup>,提出了基于知识图谱的感兴趣图节点快速检索算法,充分利用多源土地覆盖数据产品间的信息,缩短了数据预处理时间,并发现了覆盖全国的县级行政单元的产品数据一致性误差。针对旅游管理与市场分析需求,基于旅游信息门户的用户生成内容(User Generated Contents, UGC),通过专家知识本体与预训练语言模型ERNIE的迁移学习,实现了旅游相关知识抽取,并集成多源信息构建了全方位的旅游知识图谱,支持游客行为模式和出游偏好分析<sup>[60]</sup>,并基于知识图谱,利用旅游评论文本训练Word2Vec词嵌入模型,实现文本语义增强,捕捉旅游知识图谱中蕴含的关系语义,度量旅游景点的相似性<sup>[61]</sup>;通过构建知识图谱,解决了当前推荐系统面临的推荐精度低和推荐结果缺乏可解释性问题<sup>[62-63]</sup>。

### 5 发展趋势与展望

时空知识图谱方兴未艾。针对时空知识图谱所涉及的时空信息抽取、空间计算、语义网对齐、知识形式化表达与表示学习等主题,业界已经开展了大量研究,取得了诸多研究成果,为时空知识图谱的深度应用奠定了基础。然而,时空知识服务所需的知识获取、推理计算与主动推荐过程,依然存在一些关键科学问题与技术瓶颈需要突破,包括:

#### (1) 多模态时空知识获取

知识图谱中的形式化知识元组,当前主要基于各种形态的文本数据源处理得到。然而,除文本



外,各种对地观测、传感器网络与社会感知手段获取的图像、视频、移动对象轨迹与传感网监测数据中同样蕴含丰富的时空知识。基于不同模态数据获取的时空知识在系统性、完整性上存在较大差异。需要探索基于多模态数据的时空知识获取与融合方法,弥补单一数据来源的不足,为时空知识图谱提供多源异构信息支持。

### (2) 复杂时空知识表达

知识图谱通常采用RDF模型进行知识点表达,强调对象间的语义关系。RDF模型本质上是一种“二元一阶谓词逻辑”的知识表示形式,虽然能够表示大部分简单事件或实体属性,然而在表示复杂时空知识时却束手无策。面对客观世界中普遍存在的各类复杂时空知识,RDF模型难以保证知识的准确性与完备性。如何突破RDF模型的缺陷,发展可充分反映多维度时空特征及关系,兼容不同时空对象、尺度和粒度的复杂时空知识自适应表示模型,及对应的数据库实现方法,需要进一步研究。

### (3) 面向时空知识图谱的知识推理

面向知识图谱的知识推理旨在基于已有知识图谱中的事实,推理出新的事实或识别错误的知识,一方面维护知识图谱的完整性与有效性,另一方面,利用知识推理过程的可追溯性,增强推理结果的可解释性。可以说,知识推理贯穿知识图谱从构建到应用的全过程。面向知识图谱维护的知识推理包括知识图谱补全和知识图谱去噪两个主要任务。知识图谱补全通过对已有事实的归纳演绎,基于逻辑、规则和上下文,填补存在缺失的知识三元组要素,或者对知识元组进行新的扩充,增加知识存量,包括连接预测、实体预测、关系预测、属性预测等任务。知识图谱去噪通过对已有知识元组进行一致性检验,或者根据新获取的知识,识别已有知识图谱中知识元组的误差甚至错误,对知识元组进行修订更新<sup>[64]</sup>。

面向通用知识图谱的知识推理已在垂直搜索、智能问答等应用领域发挥了重要作用。而时空知识推理目前还处于萌芽状态。时空知识推理是实现时空数据智能向时空知识智能跃迁的技术瓶颈,直接决定了时空知识图谱的应用广度与深度。如何构建时空知识高效计算与推理机制,并提升时空知识推理结果的可解释性,需要深度借鉴面向通用知识图谱的知识推理技术,结合时空知识表示学

习、实体空间位置语义与空间关系表达,采用地理实体多尺度层次关系推理机制与关系传递概率图模型等,进行深入研究。

### (4) 行业时空相关知识抽取与工程应用

知识图谱的核心资源是高质量的知识。对于应用领域而言,行业的深层知识属于知识中最具含金量的内容,其表现形式都是非结构化、隐性的。目前仍然缺乏有效的行业知识抽取手段。实际应用中多依赖人工/半人工的方式进行,效率和准确性有待提升。

此外,时空知识在空间格局解析、问题诊断、演变模式识别、趋势预测等不同应用中的作用不尽相同。现有的时空知识服务主要实现了相关数据、文献及工具的整合,提供跨类型资源语义检索及简单的知识导航等。时空知识的行业应用,需要基于时空知识图谱自适应表达模型,在前述关键技术方法基础上,研发时空知识图谱管理系统和时空知识应用服务支撑平台,并与自然资源、生态环境、应急管理、城乡规划与建设等行业领域业务深度融合,发展适应不同应用场景的多层次时空知识服务工程化应用技术,支撑相关行业对时空知识服务的迫切需求。

目前各大互联网企业竞相推出基于大规模语料和机器学习大模型的问答系统,如OpenAI ChatGPT、Google Bard、百度文心一言等,成为全球关注热点。这些产品利用自然语言模型在大规模语料上训练巨大的神经网络模型,能够通过理解和学习人类语言来进行互动<sup>[65]</sup>。基于海量语料和机器学习大模型的知识提取与产生能力,让很多以往依赖人工完成的基础性、模式化工作变得无关紧要。虽然这些产品被诟病不具备实时搜索功能,无法学习新知识并更新知识储备,但是可以预见,这些问题将很快得到解决。我们认为,在科学探索方面,碳基生命的个性化特征也将不断被硅基生命的普适化特征所取代。除非这些个性化特征具有强烈的创新内涵,在当时的科技背景下,人工智能难以表达与复现。

然而,这些系统从大规模语料库中学习到的很多知识,据推测是存储在分布式表示的难以解释的黑盒模型中,并非知识图谱要求的形式化表达的显式知识元组。这与人类直接认识世界内在机理的朴素需求是相悖的。并且,训练一个大型语言模型



费时费力,成本巨大,这种解决方案是否适合行业领域的应用需求,值得商榷。行业领域知识异构性极强,如何把行业领域知识融入到大模型中极具挑战。此外,这些系统并不构建显式的知识图谱,而是通过大规模文本语料的学习过程,建立隐式的知识图谱,然后根据这个隐式知识图谱回答问题。换句话说,它们并不是真正理解了人类的知识传承,进而也难以实现很多需要因果关系而不是相关关系的知识推理过程。因此,我们认为,在很长一段时间内,构建显式的时空知识图谱,并与深度学习技术有机融合,针对行业领域业务需求进行知识组织和计算推理,仍将是时空知识服务的必由之路。

#### 参考文献(References):

- [1] 高松.地理空间人工智能的近期研究总结与思考[J].武汉大学学报·信息科学版,2020,45(12):1865-1874. [Gao S. A review of recent researches and reflections on geospatial artificial intelligence[J]. Geomatics and Information Science of Wuhan University, 2020,45(12):1865-1874. ] DOI:10.13203/j.whugis20200597
- [2] Janowicz K, Gao S, McKenzie G, et al. GeoAI: Spatially explicit artificial intelligence techniques for geographic knowledge discovery and beyond[J]. International Journal of Geographical Information Science, 2020,34(4):625-636. DOI:10.1080/13658816.2019.1684500
- [3] 程学旗,梅宏,赵伟,等.数据科学与计算智能:内涵、范式与机遇[J].中国科学院院刊,2020,35(12):1470-1481. [Cheng X Q, Mei H, Zhao W, et al. Data science and computing intelligence: Concept, paradigm and opportunities[J]. Bulletin of Chinese Academy of Sciences, 2020,35(12):1470-1481.] DOI:10.16418/j.issn.1000-3045.20201116005
- [4] 陈军,刘万增,武昊,等.基础地理知识服务的基本问题与研究方向[J].武汉大学学报·信息科学版,2019,44:38-47. [Chen J, Liu W Z, Wu H, et al. Basic issues and research agenda of geospatial knowledge service[J]. Geomatics and Information Science of Wuhan University, 2019,44:38-47. ] DOI:10.13203/j.wuugis20180441
- [5] 陆锋,余丽,仇培元.论地理知识图谱[J].地球信息科学学报,2017,19(6):723-734. [Lu F, Yu L, Qiu P Y. On geographic knowledge graph[J]. Journal of Geo-information Science, 2017,19(6):723-734. ] DOI:10.3724/SP.J.1047.2017.00723
- [6] 张雪英,张春菊,吴明光,等.顾及时空特征的地理知识图谱构建方法[J].中国科学:信息科学,2020,50(7):1019-1032. [Zhang X Y, Zhang C J, Wu M G, et al. Spatio-temporal features based geographical knowledge graph construction[J]. Sci Sin Inform, 2020,50(7):1019-1032.] DOI:10.1360/SSI-2019-0269
- [7] 王志华,杨晓梅,周成虎.面向遥感大数据的地学知识图谱构想[J].地球信息科学学报,2021,23(1):16-28. [Wang Z H, Yang X M, Zhou C H. Geographic knowledge graph for remote sensing big data[J]. Journal of Geo-information Science, 2021,23(1):16-28. ] DOI:10.12082/dqxxkx.2021.200632
- [8] 蒋秉川,万刚,许剑,等.多源异构数据的大规模地理知识图谱构建[J].测绘学报,2018,47(8):1051-1061. [Jiang B C, Wang G, Xu J, et al. Geographic knowledge graph building extracted from multi-sourced heterogeneous data[J]. Acta Geodaetica et Cartographica Sinica, 2018,47(8):1051-1061. ] DOI:10.11947/j.AGCS.2018.20180113
- [9] 刘俊楠,刘海砚,陈晓慧,等.面向多源地理空间数据的知识图谱构建[J].地球信息科学学报,2020,22(7):1476-1486. [Liu J N, Liu H Y, Chen X H, et al. The Construction of knowledge graph towards multi-source geospatial data[J]. Journal of Geo-information Science, 2020,22(7):1476-1486. ] DOI:10.12082/dqxxkx.2020.190565
- [10] 张永军,王飞,李彦胜,等.遥感知识图谱创建及其典型场景应用技术[J].遥感学报,2023,27(2):249-266. [Zhang Y J, Wang F, Li Y S, et al. Remote sensing knowledge graph construction and its application in typical scenarios [J]. National Remote Sensing Bulletin, 2023,27(2):249-266. ] DOI:10.11834/jrs.20210469
- [11] 李彦胜,张永军.耦合知识图谱和深度学习的新一代遥感影像解译范式[J].武汉大学学报·信息科学版,2022,47(8):1176-1190. [Li Y S, Zhang Y J. A new paradigm of remote sensing image interpretation by coupling knowledge graph and deep learning[J]. Geomatics and Information Science of Wuhan University, 2022,47(8):1176-1190. ] DOI:10.13203/j.whugis20210652
- [12] 周成虎,孙九林,苏奋振,等.地理信息科学发展和技术应用[J].地理学报,2020,75(12):2593-2609. [Zhou C H, Sun J L, Su F Z, et al. Geographic information science development and technological application[J]. Acta Geographica Sinica, 2020,75(12):2593-2609.] DOI:10.11821/dlxb202012004
- [13] Zhu Y Q, Zhu A X, Feng M, et al. A similarity-based automatic data recommendation approach for geographic models[J]. International Journal of Geographical Information Science, 2017,31(7):1403-1424. DOI:10.1080/13658816.2017.1300805
- [14] Zhu Y Q, Zhu A X, Song J, et al. Multidimensional and

- quantitative interlinking approach for linked geospatial data[J]. *International Journal of Digital Earth*, 2017,10(9): 923-943. DOI:10.1080/17538947.2016.1266041
- [15] Zhu Y Q and Yang J. Automatic data matching for geospatial models: a new paradigm for geospatial data and models sharing[J]. *Annals of GIS*, 2019,25(4):283-298. DOI: 10.1080/19475683.2019.1670735
- [16] 陈述彭,岳天祥,励惠国.地学信息图谱研究及其应用[J]. *地理研究*,2000,19(4):337-343. [Chen S P, Yue T X, Li H G. Studies on Geo-informatic Tupu and its application [J]. *Geographical Research*, 2000,19(4):337-343. ] DOI: 10.11821/yj2000040002
- [17] 陈述彭.地学信息图谱探索研究[M].北京:商务印书馆, 2001. [Chen S P. Research on Geo-information Tupu[M]. Beijing: The Commercial Press, 2001. ]
- [18] 陈述彭.地学信息图谱的概念[M]//陈述彭.地学的探索 (第六卷):地球信息科学.北京:科学出版社,2003:209-228. [Chen S P. The concept of Geo-information Tupu [M]//Chen S P. Exploration of Geosciences (Volume VI): Earth Information Science. Beijing: Science Press, 2003: 209-228. ]
- [19] 张洪岩,周成虎,闫国年,等.试论地学信息图谱思想的内涵与传承[J].*地球信息科学学报*,2020,22(4):653-661. [Zhang H Y, Zhou C H, Lv G N, et al. The connotation and inheritance of Geo-information Tupu[J]. *Journal of Geo-information Science*,2020,22(4):653- 661. ] DOI:10.12082/dqxxkx.2020.200167
- [20] 张春菊,张雪英,王曙,等.中文文本的事件时空信息标注 [J]. *中文信息学报*,2016,30(3):213- 222. [Zhang C J, Zhang X Y, Wang S, et al. Annotation of spatial-temporal information of event in Chinese text[J]. *Journal of Chinese Information Processing*, 2016,30(3):213-222. ]
- [21] 王姬卜,陆锋,吴升,等.基于自动回标的地理实体关系语料库构建方法[J].*地球信息科学学报*,2018,20(7):871-879. [Wang J B, Lu F, Wu S, et al. Constructing the corpus of geographical entity relations based on automatic annotation[J]. *Journal of Geo-information Science*, 2018, 20(7):871-879. ] DOI:10.12082/dqxxkx.2018.180032
- [22] Gao S, Li L, Li W, et al. Constructing gazetteers from volunteered big geo-data based on Hadoop[J]. *Computer, Environment and Urban Systems*, 2017,61:172-186.
- [23] Hu Y. Geo-text data and data-driven geospatial semantics [J]. *Geography Compass*, 2018,12(11):e12404
- [24] Ju Y, Adams B, Janowicz K, et al. Things and strings: Improving place name disambiguation from short texts by combining entity co-occurrence with topic modeling[C]. *European Knowledge Acquisition Workshop*, Bologna, Italy, 2016.
- [25] Acheson E, Volpi M, Purves R S. Machine learning for cross-gazetteer matching of natural features[J]. *International Journal of Geographical Information Science*, 2020, 34(4):708-734.
- [26] 张雪英,闫国年,杜咪,等.大数据驱动的地名信息获取与应用[J].*现代测绘*,2017,40(2):1-5. [Zhang X Y, Lv G N, Du M, et al. Acquisition and application on geographical names information based on large data driving[J]. *Modern Surveying and Mapping*, 2017,40(2):1-5. ]
- [27] Santos R, Murrieta-Flores P, Calado P, et al. Toponym matching through deep neural networks[J]. *International Journal of Geographical Information Science*, 2018,32(2): 324-348.
- [28] Wang S, Zhang X Y, Ye P, et al. Deep belief networks based toponym recognition for Chinese text[J]. *ISPRS International Journal of Geo-information*, 2018,7(6):217. DOI:10.3390/ijgi7060217
- [29] 张春菊,陈玉冰,张雪英,等.基于交互式与迭代式学习的地名语料库智能构建方法[P].中国:ZL201911029961.4, 2023年1月18日. [Zhang C J, Chen Y B, Zhang X Y, et al. An intelligent construction method for place name corpus based on interactive and iterative learning[P].China: ZL201911029961.4, 2023-01-18. ]
- [30] Wang F, Li P F, Zhu Q M. A hybrid model of classification and generation for spatial relation extraction[C]. In *Proceedings of the 29<sup>th</sup> International Conference on Computational Linguistics*, 2022:1915-1924.
- [31] Zhang L, Wang R Z, Zhou J B, et al. Joint intent detection and entity linking on spatial domain queries[C]. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020:4937-4947.
- [32] Yang T Y, Lan A S, Narasimhan K. Robust and interpretable grounding of spatial references with relation networks [R]. 2020, DOI:10.48550/arXiv.2005.00696
- [33] Song C, Lin Y F, Guo S N, et al. Spatial-temporal synchronous graph convolutional networks: A new framework for spatial-temporal network data forecasting[C]. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020,34(1):914- 921. DOI:10.1609/aaai.v34i01.5438
- [34] Nichols E, Botros F. SpRL-CWW: Spatial relation classification with independent multi-class models[C]. In *Proceedings of the 9<sup>th</sup> International Workshop on Semantic Evaluation (SemEval 2015)*, 2015:895- 901. DOI:10.18

- 653/v1/S15-2150
- [35] D'Souza J, Ng V. Sieve-based spatial relation extraction with expanding parse trees[C]. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 758-768, Lisbon, Portugal. Association for Computational Linguistics. DOI:10.18653/v1/D15-1087
- [36] Ramrakhiani N, Palshikar G, Varma V. A simple neural approach to spatial role labelling[C]. In: Azzopardi L, Stein B, Fuhr N, et al. (eds) Advances in Information Retrieval. ECIR 2019. Lecture Notes in Computer Science, vol 11438. Springer, Cham. DOI:10.1007/978-3-030-15719-7\_13
- [37] Shin H J, Park J Y, Yuk D B, et al. BERT-based spatial information extraction[C]. In Proceedings of the Third International Workshop on Spatial Language Understanding, 2020:10-17. DOI:10.18653/v1/2020.splu-1.2
- [38] Qiu Q, Xie Z, Ma K, et al. Spatially oriented convolutional neural network for spatial relation extraction from natural language texts[J]. Transactions in GIS, 2022,26(2): 839-866. DOI:10.1111/tgis.12887
- [39] Yu L, Qiu P Y, Gao J L, et al. A knowledge-based filtering method for open relations among geo-entities[J]. ISPRS International Journal of Geo- information, 2019,8(2):59. DOI:10.3390/ijgi8020059
- [40] 高嘉良,余丽,仇培元,等.基于通用知识库的地理实体开放关系过滤方法[J]. 地球信息科学学报,2019,21(9):1392-1401. [ Gao J L, Yu L, Qiu P Y, et al. A knowledge-based method for filtering Geo- entity relations[J]. Journal of Geo- information Science, 2019,21(9):1392- 1401. ] DOI: 10.12082/dqxkx.2019.190005
- [41] 仇培元,张恒才,余丽,等.微博客蕴含交通事件信息抽取的自动标注方法[J].中文信息学报,2017,31(2):107-116. [ Qiu P Y, Zhang H C, Yu L, et al. An automatic labeling method for extracting traffic event information from microblog messages[J]. Journal of Chinese Information Processing, 2017,31(2):107-116. ]
- [42] 黄宗财,仇培元,陆锋,等.基于联合主题特征的网络新闻文本蕴含环境污染事件检测[J]. 地球信息科学学报, 2019,21(10):1510-1517. [ Huang Z C, Chou P Y, Lu F, et al. Detection of environmental pollution events in news corpora based on joint thematic features[J]. Journal of Geo-information Science, 2019,21(10):1510-1517. ] DOI: 10.12082/dqxkx.2019.190037
- [43] 黄宗财,仇培元,王海波,等. 2019. 结合事件和语境特征的台风事件信息抽取方法[J]. 测绘科学技术学报,36(2): 209-214. [ Huang Z C, Qiu P Y, Wang H B, et al. Typhoon event information extraction method based on event and context characteristics[J]. Journal of Geomatics Science and Technology, 2019,36(2):209-214.] DOI:10.3969/j.issn.1673-6338.2019.02.018
- [44] Ye P, Zhang X Y, Zhang C J, et al. Positioning localities for vague spatial location description: A supervaluation semantics approach[J]. ISPRS International Journal of Geo-information, 2022,11(1):68. DOI:10.3390/ijgi11010068
- [45] Wang X N, Du S H, Feng C C, et al. Interpreting the fuzzy semantics of natural- language spatial relation terms with the fuzzy random forest algorithm[J]. ISPRS International Journal of Geo- information, 2018,7(2):58. DOI:10.3390/ijgi7020058
- [46] 王圣音,刘瑜,陈泽东,等.大众点评数据下的城市场所范围感知方法[J].测绘学报,2018,47(8):1105-1113. [ Wang S Y, Liu Y, Chen Z D, et al. Representing multiple urban places' footprints from Dianping.com Data[J]. Acta Geodaetica et Cartographica Sinica, 2018,47(8):1105- 1113. ] DOI:10.11947/j.AGCS.2018.20180110
- [47] Liu K, Qiu P Y, Gao S, et al. Investigating urban metro stations as cognitive places in cities using points of interest [J]. Cities, 2020,97:102561. DOI:10.1016/j.cities.2019.102561
- [48] Sun K, Zhu Y Q, Song J. Progress and challenges on entity alignment of geographic knowledge bases[J]. ISPRS International Journal of Geo- information, 2019,8(2):77. DOI: 10.3390/ijgi8020077
- [49] Sun K, Hu Y J, Song J, et al. Aligning geographic entities from historical maps for building knowledge graphs[J]. International Journal of Geographical Information Science, 2021,35(10):2078- 2107. DOI:10.1080/13658816.2020.1845702
- [50] 徐召华,诸云强,宋佳,等.基于词嵌入的地理知识库实体类别对齐方法研究[J]. 地球信息科学学报,2021,23(8): 1372-1381. [ Xu Z H, Zhu Y Q, Song J, et al. Word embedding-based method for entity category alignment of geographic knowledge base[J]. Journal of Geo- information Science, 2021,23(8):1372-1381. ] DOI:10.12082/dqxkx.2021.200566
- [51] Yu L, Qiu P Y, Liu X L, et al. A holistic approach to aligning geospatial data with multidimensional similarity measuring[J]. International Journal of Digital Earth, 2018,11(8): 845-862. DOI:10.1080/17538947.2017.1359688
- [52] Wang S, Zhang X Y, Ye P, et al. Geographic knowledge graph (GeoKG): A formalized geographic knowledge representation[J]. ISPRS International Journal of Geo-infor-



- mation, 2019,8(4):184. DOI:10.3390/ijgi8040184
- [53] Zhang X Y, Huang Y, Zhang C J, et al. Geoscience Knowledge Graph (GeoKG): Development, construction and challenges[J]. Transactions in GIS, 2022,26(6):1-15. DOI: 10.1111/tgis.12985
- [54] Qiu P Y, Gao J L, Yu L, et al. Knowledge embedding with geospatial distance restriction for geographic knowledge graph completion[J]. ISPRS International Journal of Geo-information, 2019,8(6):254. DOI:10.3390/ijgi8060254
- [55] 黄挚文.台风灾害事件知识图谱构建方法研究[D].南京: 南京师范大学,2021. [ Huang Z W, Research on the construction method of knowledge graph for the typhoon disaster events[D]. Nanjing: Nanjing Normal University, 2021. ]
- [56] 王晓爽,李吉东,徐海红,等.顾及时空特征的大气污染执法事理图谱构建方法研究[J].地理与地理信息科学,2022, 38(3):1-8 [ Wang X S, Li J D, Xu H H, et al. Logic graph construction of air pollution law enforcement event considering the spatiotemporal features[J]. Geography and Geo-information Science, 2022,38(3):1- 8.] DOI:10.3969/j.issn.1672-0504.2022.03.001
- [57] Wang X, Zhu Y Q, Zeng H Y, et al. Spatialized analysis of air pollution complaints in Beijing using the BERT + CRF model[J]. Atmosphere 2022,13:1023. DOI:10.3390/atmos13071023
- [58] 朱庆,王所智,丁雨淋,等.铁路隧道钻爆法施工智能管理的安全质量进度知识图谱构建方法[J].武汉大学学报·信息科学版,2022,47(8):1155-1164. [ Zhu Q, Wang S Z, Ding Y L, et al. A method of safety- quality- schedule knowledge graph for intelligent management of drilling and blasting construction of railway tunnels[J]. Geomatics and Information Science of Wuhan University, 2022, 47(8):1155-1164. ] DOI:10.13203/j.whugis20210573
- [59] 张永军,程鑫,李彦胜,等.利用知识图谱的国土资源数据管理与检索研究[J].武汉大学学报·信息科学版,2022,47 (8):1165-1175. [ Zhang Y J, Cheng X, Li Y S, et al. Research on land and resources management and retrieval using knowledge graph[J]. Geomatics and Information Science of Wuhan University, 2022,47(8):1165- 1175. ] DOI:10.13203/j.whugis20210714
- [60] 高嘉良,陆锋,彭澎,等.基于网络文本迁移学习的旅游知识图谱构建[J].武汉大学学报·信息科学版,2022,47(8), 1191-1200,1219. [ Gao J L, Lu F, Peng P, et al. Construction of tourism attraction knowledge graph based on web text and transfer learning[J]. Geomatics and Information Science of Wuhan University, 2022,47(8):1191- 1200,1219. ] DOI:10.13203/j.whugis20220120
- [61] Qiu P Y, Gao J L, Lu F. Identifying the relatedness between tourism attractions from online reviews with heterogeneous information network embedding[J]. ISPRS International Journal of Geo-information, 2021,10(12):797. DOI: 10.3390/ijgi10120797
- [62] 秦川,祝恒书,庄福振,等.基于知识图谱的推荐系统研究综述[J].中国科学:信息科学,2020,50(7):937-956. [ Qin C, Zhu H S, Zhuang F Z, et al. A survey on knowledge graph- based recommender systems[J]. Sci Sin Inform, 2020,50(7):937-956. ] DOI:10.1360/SSI-2019-0274
- [63] 高嘉良,仇培元,余丽,等.基于旅游知识图谱的可解释景点推荐[J].中国科学:信息科学版,2020,50(7):1055-1068 [ Gao J L, Qiu P Y, Yu L, et al. An interpretable attraction recommendation method based on knowledge graph[J]. SCIENTIA SINICA Informationis, 2020,50(7):1055- 1068. ] DOI:10.1360/SSI-2019-0268
- [64] 官赛萍,靳小龙,贾岩涛,等.面向知识图谱的知识推理研究进展[J].软件学报,2018,29(10):2966-2994. [ Guan S P, Jin X L, Jia Y T, et al. Knowledge reasoning over knowledge graph: A survey[J]. Journal of Software, 2018,29 (10):2966-2994.] DOI:10.13328/j.cnki.jos.005551
- [65] Zhou J, Ke P, Qiu X P, et al. ChatGPT: Potential, prospects, and limitations[J]. Frontiers of Information Technology & Electronic Engineering. 2023. DOI:10.1631/fi-tee.2300089