中国科学: 地球科学 2023 年 第 53 卷 第 11 期: 2609 ~ 2622
SCIENTIA SINICA Terrae earthcn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
论 文

顾及复杂时空特征及关系的地球科学知识图谱自适应表达模型

诸云强[1,2*], 孙凯[1†], 王曙[1], 周成虎[1], 陆锋[1], 闾海荣[3], 邱芹军[4], 王新兵[5], 祁彦民[1,6]

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. 江苏省地理信息协同创新中心, 南京 210023;
3. 清华大学自动化系, 北京 100084;
4. 中国地质大学(武汉)计算机学院, 武汉 430074;
5. 上海交通大学电子信息与电气工程学院, 上海 200240;
6. 宁波诺丁汉大学计算机科学系, 宁波 315100
* 通讯作者, E-mail: zhuyq@igsnrr.ac.cn
† 通讯作者, E-mail: sunk@lreis.ac.cn

收稿日期: 2022-11-21; 收修改稿日期: 2023-07-09; 接受日期: 2023-08-21; 网络版发表日期: 2023-10-18
国家自然科学基金项目(批准号: 42050101)和国家重点研发计划项目(批准号: 2022YFB3904200、2021YFB00903)资助

摘要 地球科学(以下简称“地学”)知识图谱可以将各类地学知识组织成计算机可理解、可计算的语义网络, 是地学知识有效的组织和服务方式, 已经成为地学研究的热点和前沿之一. 地学知识包含众多的学科领域知识, 具有多尺度、多粒度、多维度等复杂的时空特征及关系. 因此, 建立符合地学知识特点的地学知识图谱表达模型, 是地学知识图谱构建与应用的基础和前提. 针对现有知识图谱表达模型采有固定元组, 难以充分表达复杂时空特征及关系的问题, 本文首先对地学知识分类、地学知识时空特征及关系进行了系统分析, 在此基础上, 提出了顾及复杂时空特征及关系的地学知识图谱自适应表达模型. 该模型在统一时空本体的约束下, 基于地学知识的时空关联度, 采用不同元组对不同类型的地学知识进行自适应表达. 通过该模型, 不仅可实现地学知识的高效表达, 避免时空特征表达的隔裂, 提升地学知识的检索精度和效率, 还可以通过时空本体, 实现时空信息的统一对齐转换与计算推理.

关键词 地球科学, 知识图谱, 表达模型, 时空特征, 时空关系

1 前言

知识图谱以结构化的图模式, 将知识组织成计算机可理解、可计算的一种语义网络, 具有强大的知识表示和推理能力, 已成为人工智能研究与应用的重要基础设施(Hogan等, 2022; Zhang等, 2022). 当前, 通用领域的知识图谱研究在知识表达与表示、抽取、对齐融合、补全等方面已取得较大进展(Ji等, 2022), 形成

中文引用格式: 诸云强, 孙凯, 王曙, 周成虎, 陆锋, 闾海荣, 邱芹军, 王新兵, 祁彦民. 2023. 顾及复杂时空特征及关系的地球科学知识图谱自适应表达模型. 中国科学: 地球科学, 53(11): 2609–2622, doi: 10.1360/SSTe-2022-0371
英文引用格式: Zhu Y, Sun K, Wang S, Zhou C, Lu F, Lv H, Qiu Q, Wang X, Qi Y. 2023. An adaptive representation model for geoscience knowledge graphs considering complex spatiotemporal features and relationships. Science China Earth Sciences, 66(11): 2563–2578, https://doi.org/10.1007/s11430-022-1169-9

© 2023 《中国科学》杂志社

了DBpedia(https://www.dbpedia.org)(Lehmann等, 2015)、YAGO(https://yago-knowledge.org)(Pellissier等, 2020)、Freebase(Bollacker等, 2008)、Wikidata(Vrandečić和Krötzsch, 2014)等经典知识图谱, 并在智能问答、语义搜索、个性化推荐等领域得到了广泛应用(Reinanda等, 2020; Hui等, 2022; Zhao等, 2022). 2022年11月, OpenAI发布了对话生成预训练模型(Chat Generative Pre-trained Transformer, ChatGPT), 引起了全球广泛的关注. ChatGPT是在大规模数据上训练的神经网络大语言模型, 它通过大规模预训练、多任务提示学习、强化学习等阶段的处理, 具有强大的高质量对话、复杂推理、思维链、语境学习、跨任务泛化、代码理解/生成等能力. 当然, ChatGPT也存在不足以准确处理逻辑问题、产生事实不正确或有偏见的答案、不具备实时学习新知识并更新知识储备等问题. 尽管如此, ChatGPT基于庞大的数据和强大的计算能力, 正在改变传统知识生产和创造的认知与方式, 为知识图谱的进一步发展带来了新的机遇和挑战(Brown等, 2020; Schulman等, 2022; Zhou等, 2023).

地学知识图谱可视为知识图谱技术在地学领域的应用, 目前已经逐步成为地学大数据集成共享与挖掘分析、地学知识发现以及地学与人工智能融合发展的基础设施(陆锋等, 2017; 高松, 2020; Ma, 2022; 周成虎等, 2021; 诸云强等, 2023; Li等, 2023), 推动和促进了地学新兴和交叉学科的发展(姚玉鹏和郭正堂, 2021). 例如, 新兴的信息地理学(熊巨华等, 2021; 李新等, 2021), 就更加强调对自然地理和人文地理信息普遍性和特殊性规律的认知和形式化表达(刘瑜, 2022), 从而促进物理世界、人文世界向信息世界的映射转换, 以及通过信息世界构建的虚拟空间场景, 仿真模拟和影响改变物理世界和人文世界(闾国年等, 2022); 通过信息地理学从本体论、认识论、方法论三个层面促进地理学范式的转换(李新等, 2022). 通用知识图谱的发展, 为地学知识图谱的研究与应用提供了重要的基础和参考. 然而, 地学知识涉及大气、海洋、地理、地质、地球物理、地球化学等多个学科(孙鸿烈, 2017), 包含地学对象、要素、现象、过程、事件等的空间格局、演变过程、形成规律及其相互作用等方面的知识, 具有多尺度、多粒度、多维度的复杂时空特征及关系. 因此, 如何面向不同学科和类型的地学知识, 建立顾及复杂时空特征及关系的地学知识图谱表达模型, 是地学知识图谱构建与应用的基础和前提(张雪英等, 2020; 周成虎等, 2021).

通用知识图谱表达模型采用“节点-边”有向图的形式对知识进行表达, 其中: 节点代表概念或实体, 边表示节点的属性或节点间的相互关系. 在实现上, 利用资源描述框架(Resource Description Framework, RDF)三元组数据模型, 即主语(Subject, S)-谓语(Predicate, P)-宾语(Object, O){S, P, O}, 或头部(Head, H)-关系(Relation, R)-尾部(Tail, T){H, R, T}来表达(陆锋等, 2017; 刘知远等, 2020; 高嘉良等, 2020; 周园春等, 2020); 其中: 主语表示节点, 谓语表示边, 宾语表示与主语具有边关系的另一个节点. 在现有知识图谱表达模型中, 地学知识的时空特征及关系作为一般的语义信息表达为不同的三元组(石顺中等, 2020; 齐浩等, 2020), 即将对象内容元组与时空元组分开表达(如: 采用下面不同的三元组表达北京市不同时期的人口: 北京市人口为1961.9万, 人口时间为2010年; 北京市人口为1363.6万, 人口时间为2000年), 容易造成同一地学知识的时空特征及关系在表达上的隔裂, 难以追溯分析不同时空条件下地学知识的状态及演变情况, 影响地学知识检索的效率与精度甚至产生错误. 尽管可以通过元组唯一标识, 实现人口数与时间的关联, 但由于人口数和时间的分开存储, 并不是强关联, 往往需要跨元组的关联检索, 容易造成检索信息遗漏而导致人口数表达的不准确. 同时, 现有知识图谱表达模型也未对时空特征及关系的描述进行规范化的约定或引用申明, 导致不同知识图谱采用不同的时空术语及谓词进行表达(表1), 难以实现地学知识统一的时空计算推理等.

为了克服现有知识图谱表达模型存在的问题, 已有研究进一步将时间(T)或空间(L)信息添加到通用三元组中, 形成诸如{S, P, O| T}、{S, P, O| L}等形式, 可有效表达时间或空间信息的四元组(Hoffart等, 2013; Zheng等, 2021). 同时, 也构建了包含时空本体在内的一些地学领域本体, 以此对地学知识进行规范化的定义(Raskin和Pan, 2005; Galton, 2009; Bittner等, 2009; Buccella等, 2011; Ma和Fox, 2013; Cox和Richard, 2015). 地理信息领域提出了基于“过程-关系”的地理知识图谱表示方法, 将同一地理实体节点扩展为不同状态的分节点, 以地理实体状态为核心, 关联地理实体时空、属性、过程和行为特征等, 依据不同状态的

表 1　代表性知识图谱采用的主要时空谓词

知识图谱	时间谓词	空间谓词
DBpedia	*directlyFollows*、*directlyPrecedes*、*follows*、*hasTimeInterval*、*includesTime*、*isObservableAt*、*isTimeIntervalOf*、*isTimeOfObservationOf*、*sameSettingAs*	*farFrom*、*hasCommonBoundary* 、*hasComponent*、*hasConstituent*、*hasConstraint*、*hasLocation*、*hasRegion*、*isComponentOf*、*isLocationOf*、*isRegionFor*、*nearTo*、*overlaps*、*coparticipatesWith*
YAGO (Hoffart等, 2013)	*startedOnDate*、*wasBornOnDate*, *diedOnDate*、*occursSince*、*occursUntil*、*happenedOnDate*、*asDestroyedOnDate*、*hasDuration*、*wasCreatedOnDate*、*occursUntil*、*occursSince*	*happenedIn*、*isLocatedIn*、*hasLatitude* *hasLongitude*、*livesIn*、*wasBornIn*、*diedIn*、*worksAt*、*hasArea*
SWEET (Raskin和Pan, 2005)	*hasStart*、*hasEnd*、*hasBeginning*、*hasTimeReference*	*above*、*across*、*adjacentTo*、*under*、*east*、*near*、*north*、*oppositeTo*、*outside*、*overlaps*、*separatedFrom*、*south*、*surroundedBy*、*upperPartOf*、*west*、*hasArea*、*hasLatitudeBand*、*hasLatitudeLine*、*hasLocation*

时序关系，反映地理实体的演变过程(Wang等，2019; 张雪英等，2020). 上述研究，在一定程度上解决了时空信息表达隔裂以及时空演变分析难的问题，但仍然存在以下问题：由于没有考虑时空信息在不同类型地学知识中的作用或关联程度，在实际应用时很难确定采用三元组还是四元组；与现有知识图谱表达模型相同，也缺乏对时空特征及关系描述的规范化约定或引用申明，对地学知识本身的版本及变更情况也没有记录, 难以支撑地学知识演化的研究分析.

针对上述问题，本文开展顾及复杂时空特征及关系的地学知识图谱表达模型研究，旨在建立在统一时空本体约束下，依据地学知识中时空信息的关联程度，高效地对复杂时空特征及关系进行描述的地学知识图谱自适应表达模型，从而为跨学科地学知识图谱的统一构建与计算推理、地学知识的演化分析等提供有效支撑.

2　地学知识分类

地学知识图谱表达模型研究首先要识别地学知识的类别及其时空特征及关系. 现有研究已从不同视角对知识进行了分类，包括：按内容分类的事实性、概念性、程序性和元认知知识，按获取途径分类的直接和间接知识，按认知深度分类的感性和理性知识，按表达角度分类的显性和隐性知识等. 在地学领域，已有研究将地理知识分为常识性和事实性知识(陆锋等, 2017)，事实型、规则及控制型、决策型知识(张雪英等, 2020)，以及数据性、概念性和规律性知识(王志华等, 2021). 俞肇元等(2022)提出了面向三元世界的地理信息分类与描述框架，认为地理信息应能承载和汇聚包括时间、地点、人物、事物、事件、现象、场景在内的信息“七要素”，并应能从地理语义、空间位置、几何形态、属性特性、要素关系、演化过程和作用机制“七维度”上对地理对象和现象进行综合性、集成性的表达和描述. 本文从地学知识图谱的视角, 对地学知识进行科学分类.

地学知识图谱的目标是实现地学知识的结构化组织与形式化关联，进而实现知识的广泛共享与智能应用. 从组成层次来看，地学知识图谱包括模式层和实例层(漆桂林等，2017; 诸云强等，2023). 模式层(又称本体层)主要包括地学领域的术语、概念及其属性、关系、规则等. 实例层(又称数据层)是在模式层指导下，对客观世界存在的各类地学知识的实例进行描述. 从知识功能的角度，地学知识可分为概念性、事实性、规则性知识和元知识.

地学概念性知识主要对应地学知识图谱的模式层, 包含地学领域的基础概念, 以及描述和表达地学实体/要素、现象、事件、场景等的概念. 地学事实性知识主要对应地学知识图谱的实例层，是客观存在的地学实体/要素、现象、事件和场景等的事实知识. 地学规则性知识则是地学概念和事实知识存在的规则、计算模型、定理和公理等，依附并作用于概念性和事实性知识. 同时，还需要完整记录上述三类地学知识创建及更新的人员、时间及具体内容等的元知识，以便支持地学知识的演化及溯源分析. 地学规则性知识和元知识也分模式层和实例层，模式层是对规则、元知

识的定义要求，实例层是依据模式层描述表达的各类具体规则和元数据. 地学知识分类及其相互间的关系如图1所示.

3 地学知识时空特征及其关系分析

3.1 地学知识时间特征

地学既有对地质时期古生物、古地理、古气候、古环境等的研究，也有对现代地理条件、水文气象、资源环境格局与过程，以及未来气候变化模拟预测等的研究；既有对漫长的大地构造、沉积、矿物演化等的研究，也有对瞬间暴发的地震、火山、地质灾害等的研究. 因此，地学知识的时间特征极其多样和复杂，具有多尺度、波动性、方向性和多标度性、不确定性等特征(侯志伟等, 2015).

从时间过程来看, 既有地质时间也有现代时间. 地质时间通常用地质年代表来表达，现代时间可采用不同的纪年方式(例如, 公历、农历、历史朝代)表示. 不同纪年方式对同一时间的表达不一样，但可以通过映射关系进行转换. 从时间表达对象来看，既有地学对象、现象、过程、事件本身发生/形成、发展/演化、消亡/结束等的时间，也有数据采集获取、生产加工、建库应用等的时间. 从时间相对性来看, 既有绝对时间也有相对时间. 相对时间往往可以通过参考点的时间转换为绝对时间. 从时间分辨率来看，既有百万年、万年等长尺度的时间，千年、百年等中尺度的时间，也有年月日甚至分秒级的高精度时间. 从时间长度来看，既有时间点、时间段也有周期性时间. 相同时间在不同分辨率下，其时间长度类型会发生变化，例如, 在年分辨率下, 某一年就是一个时间点, 但在日分辨率下就变成了时间段. 从时间方向来看，既有正向时间(从过去、现在到未来)也有负向时间，正向和负向时间都应有一个相对时间的原点，例如，距今多少年、公元前(BC)多少年、公元后(AD)多少年.

为了能够准确表达地学知识时间的上述特性，必须通过统一的时间框架(Wang等, 2022)，对不同年代、不同尺度、不同分辨率的时间进行统一的表达，以便不同的时间能够转换到统一的基准上，从而进行计算分析. 地学统一时间框架包括纪年、历法和计时三部分(图2).

在图2中，纪年是时间框架的基准，用于确定年代的起点和命名等的方法. 地学纪年主要包括地质年代

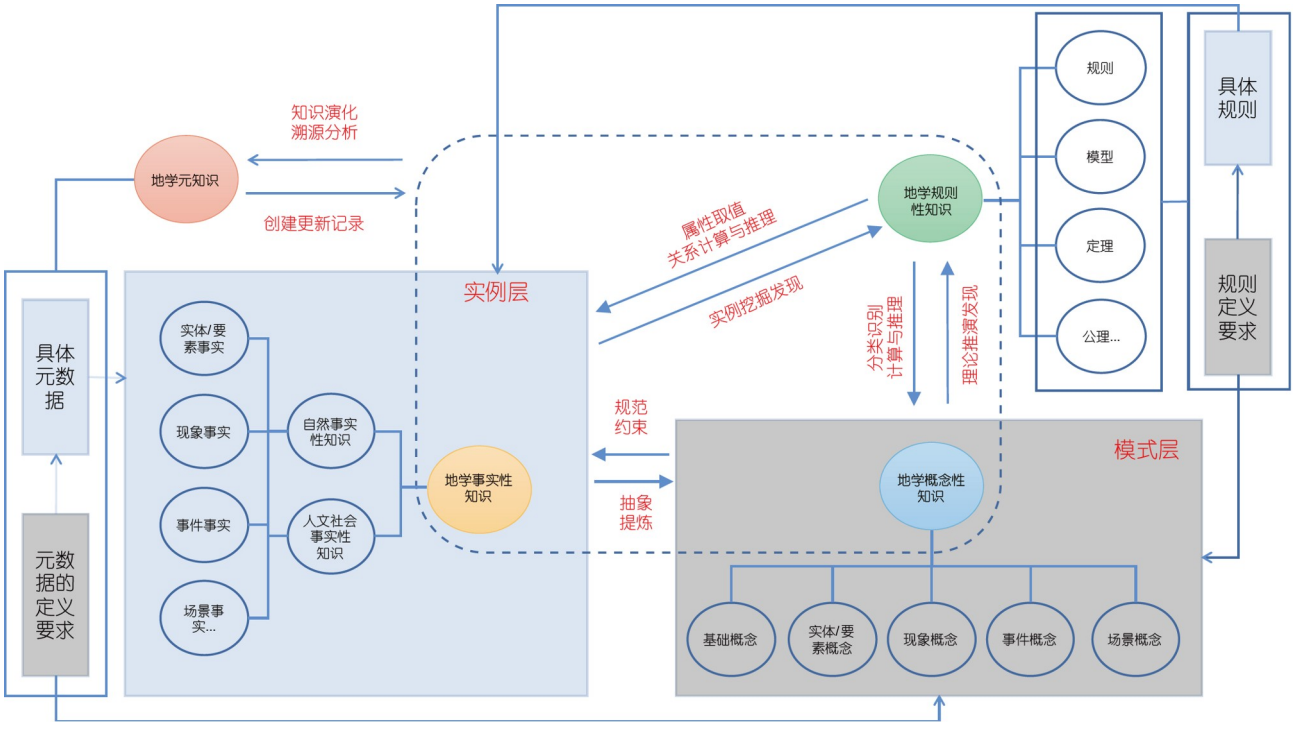

图 1 地学知识分类及相互关系

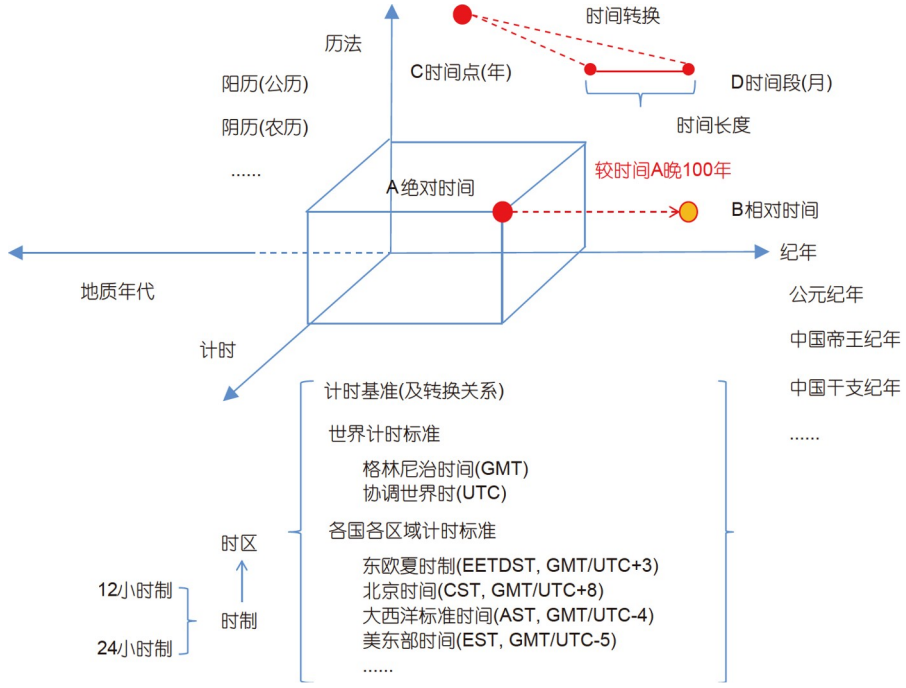

图 2　**统一的地学知识时间框架**

纪年、公元纪年、天文纪年、岁星纪年，以及中国的帝王纪年、干支纪年等. 历法是用来计算年、月、日的方法，目前常用的有阳历、阴历以及各国家采用的历法. 现代计时主要采用时、分、秒，还有毫秒、微秒等单位. 现代计时还需要考虑计时基准及其与世界计时基准的偏移量, 时制(12或24小时制), 以便各国时间能够相互转换. 世界计时基准, 包括早期的格林尼治时间(Greenwich Mean Time, GMT)和现在的协调世界时(Universal Time Coordinated, UTC); 各国本地计时基准, 如美东标准时间(Eastern Standard Time, EST)、北京时间(China Standard Time, CST), EST、CST与世界计时基准的偏移量分别为–5和+8个小时，即EST可表示为GMT或UTC-5, CST可表示为GMT、UTC+8.

因此, 构建地学知识图谱时, 必须在模式层建立完整的时间本体, 形成统一的时间表达和计量基准, 计算机才能准确理解各类时间实例并支持其相互间的映射转换和计算推理.

3.2　地学知识空间特征

地学研究的空间范围既有全球范围甚至行星际空间尺度，也有洲际、国家、城市甚至居民地、兴趣点等尺度；既有地球深部、表层的研究，也有大气空间、外层空间的研究. 因此，地学知识的空间特征也极其多样和复杂，具有覆盖范围广、尺度变化大、空间异质性与不确定性等特征(王东旭等, 2016).

从空间方向来看, 包括水平空间和垂向空间. 水平空间主要使用坐标来表示，垂向空间主要用海拔高度来表达. 由于地球是个不规则的椭球体，水平空间坐标需要指定参考坐标系(球面坐标或平面坐标系，其中：平面坐标系还需要指定投影方式)，垂向海拔高度需要指定参考的高程基准面. 水平空间通常也用地名来表示，但经常存在一名多地或一地多名等问题；从空间相对性来看，既有绝对空间又有相对空间. 如前，绝对空间通常用坐标和海拔高度来表示，而水平相对空间通常用方位、距离来表达，垂向相对空间通常用高度、深度来表达. 相对空间必须明确参考物，通过参考物的绝对位置，就可计算其绝对位置. 在三维空间中同时包含水平和垂向位置，其相对空间必须指明相对参考物的方位角、倾角和距离. 从演化过程来看，既有古地理空间也有现代空间. 古地理空间指地质时期海陆及其自然区域和地带的空间位置，现代空间指地球当前存在的各类自然和人文要素的空间位置. 古

地理空间通常依据岩石、地层、生物、地球化学、地球物理等手段测定, 但必须同时指明古地理空间位置所处的地质时间; 现代空间通过坐标系和高程系来表达. 从表达对象来看, 既有地学实体、现象、事件本身所处的空间位置, 也有样本采集、测站、钻孔等的空间位置. 从表达形式来看, 既有测站、钻孔、地震震中等点位置, 断层、剖面、河流等线位置, 平原、盆地、自然保护区等面位置, 又有地质体等三维空间. 地学统一空间框架由特定空间参考系下的水平空间坐标和垂向海拔高度构成(图3).

因此, 构建地学知识图谱时, 必须在模式层建立完整的空间坐标系、投影方式、高程系及其对应的参考椭球体和基准面、空间单位, 以及全球地名本体等, 计算机才能准确理解各类空间实例并支持其相互间的映射转换和计算推理.

3.3 地学知识关系分析

地学知识图谱的节点主要表达地学概念性知识和事实性知识, 每个概念性知识都有属性, 而每类事实性知识都归属于某个概念. 因此, 地学知识关系可分为属性与概念的属性关系、实例与概念的实例关系、概念与概念以及实例与实例之间的关联关系(如图4所示). 关联关系主要通过概念属性来表达, 主要包括语义关系、时间关系、空间关系、组成关系、作用关系、计算关系等. 时间关系、空间关系、计算关系是地学知识图谱区别于一般知识图谱的重要关系. 各类关系具有对称性(是否互斥)、继承性和传递性等的性质, 以及无向、双向和单向等方向(如表2所示). 其中, 对称性: A与B存在关系R, 则B与A也存在关系R. 如果A和B的语义类型不能同时存在, 则称A和B互斥(例如, “左”“右”在方位上是反义词, 其语义是互斥的; 林地与草地, 都是土地覆被类型, 具有兄弟关系, 但两者互斥, 即是林地就不能是草地), 否则不互斥(如: 两个地方具有相邻关系, 可以同时为城市类型). 继承性: A与B存在关系R, A′是A的子类(实例)、B′是B的子类(实例), 则A′与B′也存在关系R. 传递性: A与B存在关系R, B与C存在关系R, 则A与C也存在关系R. 由于地学知识涉及不同学科领域的知识, 其关系非常复杂, 对于本文图4及表2未能列举到的特殊关系, 可通过“其他关联关系”进行表达或对特殊关系类型进行扩展定义. 关系扩展定义时, 需要指明关系的类型、规范命名(不得与现有的关系名称重复)、标签、关系性质和方向.

属性关系是属性与概念的从属关系, 即概念包含属性, 其方向由概念指向属性, 一个概念可以有多个属性, 每个属性的出现次数可以是0(可选)、1(必选)或者*N*(多选). 属性的类型主要包括数据属性和对象属性两种类型, 其中: 数据属性是指可直接赋值的属性(其类型通常是基本的数值、字符和布尔型等), 对象属性是指由其他对象类(概念)定义的复杂属性, 即引用其他对象类. 根据属性的特征, 若干数据属性或对象属性还可进一步归类, 聚合形成复合属性. 如, 土壤具有物理和化学属性, 其中, 物理属性包括颜色、厚度、容重、通气性、透水性、黏性、松软程度等, 化学属性包括各种测量元素的含量、酸碱性(PH值)、重金属含

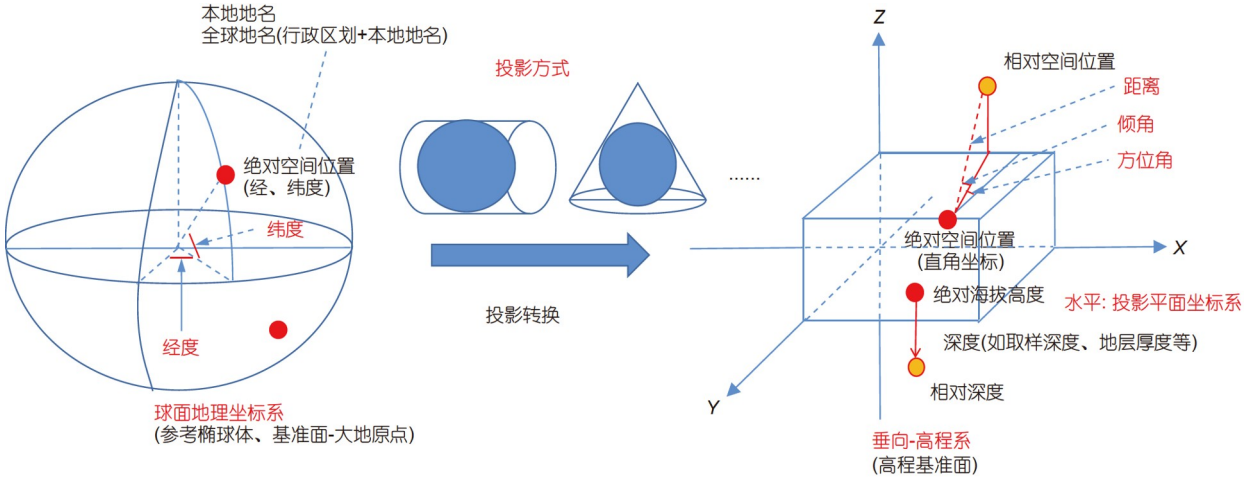

图 3　统一的地学知识空间框架

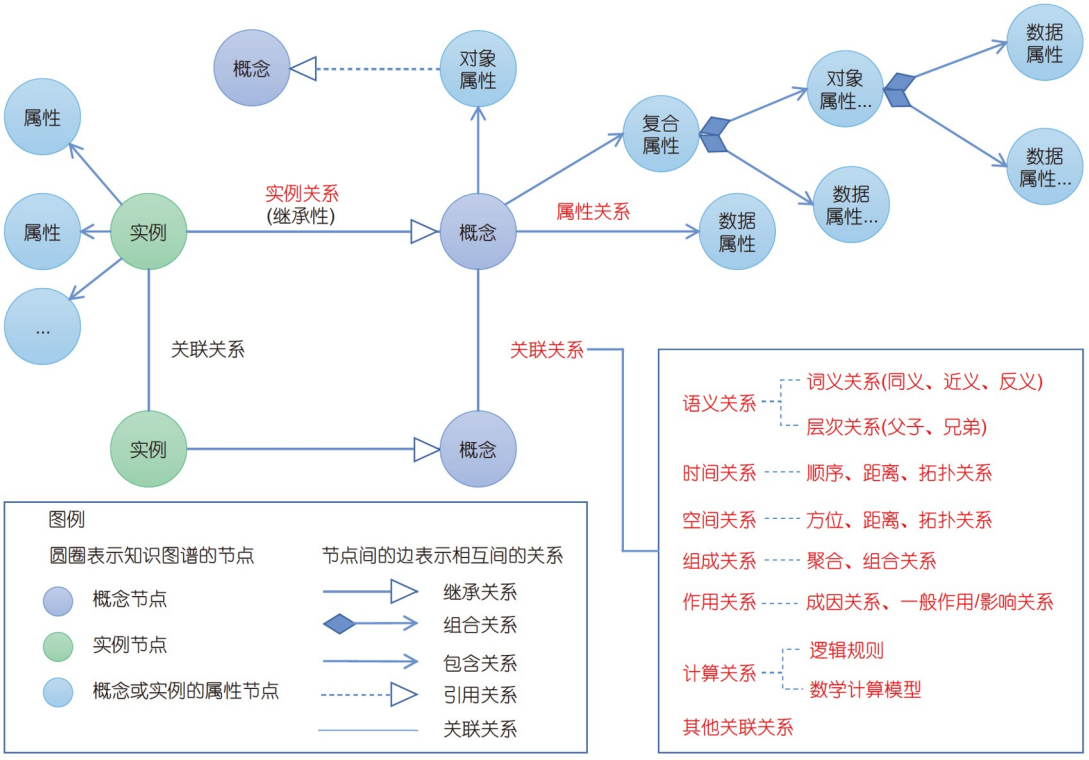

图 4　地学知识的关系类型

量等; 自然保护地具有地理位置、空间范围、面积、气象水文、土壤等自然属性, 以及保护地类型(国家公园、自然保护区、自然公园)、等级、保护生物、主管部门等社会属性; 沉积岩颜色、粒度、密度、孔隙度等可归结为沉积岩的物理属性, 而SiO_2、CaO、Al_2O_3、Fe_2O_3等元素含量可归结为其化学属性. 因此, 属性关系又可细分为数据属性关系、对象属性关系和复合属性关系. 其中, 数据属性关系需要指定数据类型及长度, 对象属性关系需要指定引用的对象类, 复合属性关系需要指定包含的属性集合. 为了实现计算机的可理解、可计算、可推理, 还需要进一步对属性概念的语义及其取值规则等进行约定.

实例关系表示实例与概念之间对象与类的继承关系. 实例必须继承其所属概念的核心属性和取值规则, 以及与其他概念之间的关系和计算推理规则. 由于所处的地理环境时空条件不一样, 同一类别的不同地学实例可能具有不同的属性值和关系值.

语义关系表示地学概念(或其实例)主题内容之间的词义关系和层级关系. 词义关系主要有同义、近义、反义关系. 同义、近义关系具有对称性和传递性, 反义关系具有对称性但不具有传递性; 层级关系主要有父子关系、兄弟关系. 父子关系(如: 沉积岩与下级的石灰岩、砂岩、页岩)表达下位概念是上位概念的子类, 具有传递性和继承性; 兄弟关系具有互斥性.

时间关系表示地学概念(或其实例)在时间上的相互关系, 主要包括时间方向关系、拓扑关系和距离关系. 时间方向关系是指两个时间之间的先后关系(早于/先于、晚于/迟于). 时间拓扑关系是指两个时间点或时间段之间的逻辑关系, 包括时间相交关系、包含/被包含关系、相邻关系等. 时间距离关系表示两个时间之间的距离, 时间距离的计算必须首先将两个时间统一到同一基准和尺度上, 而相同两点的时间距离也将随着时间单位的不同而不同.

空间关系表示地学概念(或其实例)在空间上的相互关系, 主要包括空间方位关系、距离关系和拓扑关系. 空间方位关系是两个地学实体在二维或三维空间中的相对方向, 包括二维的东西南北方向、三维的上中下方向及其组合, 方位可进一步用方位角和倾角进

表 2　地学知识关系类型及其性质[a)]

关系类型	子类型	关系名称		规范命名	关系性质			关系方向
					对称性/互斥	继承性	传递性	单向→双向↔
属性关系	属性关系	属性关系		*hasProperty*		√(继承核心属性)	√	概念→属性
实例关系	实例关系	实例关系		*hasInstance*			√	概念→实例
关联关系	语义关系	词义关系	同义关系	*isSynonymous*	√不互斥	√	√	节点↔节点
			近义关系	*isSynonym*	√不互斥	√	√	节点↔节点
			反义关系	*isAntonymy*	√互斥	√		节点↔节点
		层次关系	父子关系	*isPaternity*		√	√	父节点→子节点
			兄弟关系	*isBrother*	√互斥		√	节点↔节点
	时间关系	方向关系		*hasTDirection*		√	√	节点→节点
		拓扑关系	相交	*isTIntersect*	√不互斥	√		节点↔节点
			包含/被包含	*isTContain/isT-Within*		√	√	节点→节点
			相邻	*isTAdjacent*	√不互斥	√		节点↔节点
			相离	*isTDisjoint*	√不互斥	√		节点↔节点
		距离关系		*hasTDistance*	√不互斥	√		节点↔节点
	空间关系	方位关系		*hasLOrientation*		√	√	节点→节点
		拓扑关系	相交	*isLIntersect*	√不互斥	√		节点↔节点
			包含/被包含	*isLContain/isL-Within*		√	√	节点→节点
			相邻	*isLAdjacent*	√不互斥	√		节点↔节点
			相离	*isLDisjoint*	√不互斥	√		节点↔节点
		距离关系		*hasLDistance*	√不互斥	√		节点↔节点
	组成关系	聚合关系		*aggregationOf*		√		节点→节点
		组合关系		*combinationOf*		√		节点→节点
	作用关系	成因关系		*formedBy*		√		节点→节点
		一般作用/影响关系		*functionTo*		√		节点→节点
	计算关系	逻辑规则		*hasLogicalR*		√	√	节点→节点(单一规则及组合)
		数学计算模型		*hasMathCalR*		√	√	节点→节点(函数、模型及模型组合)
	其他关联关系							

a) 节点包含概念、属性、实例等

行定量表示; 空间距离关系表示两个地学实体之间的距离, 通常用二维的欧氏距离表示, 也可以是三维距离; 空间拓扑关系包括空间包含/被包含、邻近、相交和相离关系等.

组成关系包括聚合和组合两种关系. 聚合关系是一种弱的整体和部分的关系, 整体和部分可以相互独立, 例如, 岩石地层单位群由纵向上相邻两个或两个以上具有共同岩性特征的组联合而成; 而组合关系是一种强的整体和部分的关系, 整体和部分具有相同的生命周期.

作用关系包括成因关系与一般的作用/影响(被作用/被影响)关系, 其中成因关系是指地学对象与形成原因的关系. 例如, 地貌成因包括内营力和外营力两大类. 内营力又包括新构造运动、岩浆活动、热液活

动; 外营力又包括海成、湖成、流水、冰缘、风成、干燥、重力、生物和人为作用.

计算关系是指一个地学对象实体可由一个或若干个对象实体计算得到, 其计算关系方向由输入因子对象指向输出结果对象. 例如, 坡度可由数字高程模型(Digital Elevation Model, DEM)计算得到; 土壤生产潜力可由太阳总辐射、降水、温度、蒸发量、土壤肥力、土壤质量等计算得到. 这种计算关系可以是简单的逻辑规则、数学函数, 也可以是复杂的模型、模式. 具体实现时, 可根据计算方法的复杂程度, 采用不同的方式对计算方法进行表达, 例如, 对于简单的逻辑规则或数学函数, 可以直接将逻辑规则操作算子或计算函数作为"边"的属性; 对于复杂的计算模型, 可以考虑将具体的计算方法封装成可执行程序、软件包、网络服务(Web Service)或表征性状态转移接口(RESTful, Representational State Transfer)(Feng等, 2011), 再将它们与指定的关系边进行绑定(Qi等, 2022).

4 地学知识图谱自适应表达模型

4.1 自适应表达模型

已有研究采用固定元组, 将时空信息作为一般的语义信息或附加信息进行表达, 容易造成同一地学知识时空特征及关系在表达上的隔裂, 影响地学知识检索的效率与精度, 甚至产生错误, 难以追溯分析不同时空条件下地学知识的状态及演变情况. 同时, 由于缺乏统一的时空本体支持, 当不同地学知识采用不同的时空基准及方式表达时空信息时, 将很难进行时空的计算与推理(计算推理结果可能产生错误). 为此, 本文提出基于时空关联度的地学知识图谱自适应表达模型(如公式(1)所示), 即根据地学知识中时空关联程度的不同自动采用最合适的元组对地学知识进行表达, 其基本思路是: 在表达地学知识内容基本元组的基础上, 根据地学知识蕴含时空特征的关联度, 自动确定是否需要单独表达时间或空间信息; 而时空信息的表达, 必须遵循统一的时空本体, 进而确保不同地学知识能够在相同时空基准和表达方式上进行准确的时空计算推理; 而整个地学知识的创建更新情况可以采用元知识进行记录, 进而实现基于元知识的地学知识演化分析与追溯.

$$GeoKGAM = \left\langle S, P, O, \left\{ [P_T, T] \middle| T_o, [P_L, L] \middle| L_o \right\} \middle| R_{T,L} \right\rangle \leftrightarrow MetaK, \tag{1}$$

式(1), $GeoKGAM$为地学知识图谱自适应表达模型; S、P、O分别为表达地学知识内容的基本元组, 通常采用"主语、谓语和宾语"的形式进行表达, 主语和宾语主要用来描述地学知识对象(地学实体、要素、现象、事件、场景等的概念及实例)及其属性, 谓语主要用来描述地学对象及属性之间的关系; P_T、P_L分别为表示时间和空间的谓词, T、L分别为时间和空间信息, 时空谓词及时空信息为可选项, 主要依据时空关联度进行确定; T_o和L_o分别是时间和空间本体; $R_{T,L}$为地学知识时空关联度, 具体度量方法见下节; $MetaK$为地学元知识.

为了支持统一的时空计算, 新构建的地学知识图谱的时空信息必须按照统一的时空本体(T_o, L_o)进行表达. 已构建的地学知识图谱, 需要基于统一的时空本体对知识图谱中时空信息所采用的时空基准及表达方式(例如, 纪年方式、计时基准、所属时区编号、时间单位, 坐标参考系、坐标单位及表达方式等)进行标注, 进而支持不同知识图谱或不同地学知识间时空信息的转换对齐及准确的计算推理. 根据3.1和3.2节的分析, 时空本体T_o、L_o的具体形式分别如式(2)和(3)所示:

$$T_o = TCn, TDt\{Dt1, [Dt2], [Unt]\} \mid TdRef, \tag{2}$$

$$L_o = LCn, LCd\{Cd1, Cd2, ...Cdn\} \mid CdRef, \tag{3}$$

式中, TCn为时间名称, TDt为时间值. 在时间本体中, 就是要建立时间名称与时间值及其时间参考系的关联, 以此支持统一的时间计算. 时间值TDt可以是时间点($Dt1$), 也可以是时间段($Dt1$, $Dt2$); 对于地质时间还需要有不确定度(Unt), 以此表示时间的可信度. 例如, 深时数字地球国际大科学计划(Deep-time Digital Earth, DDE)(Wang等, 2021), 已经建立了1893~2022年60多个版本的地质年代本体(含地质年代名称、起止绝对时间、不确定度等), 可支持不同版本的地质年代表进行对齐转换和计算(Wang等, 2022); $TdRef$表示时间参考系, 具体形式如式(4)所示:

$$TdRef = Chron, Calen, Tunit, Timing(Zone, TSys), \tag{4}$$

式中, $Chron$表示纪年方式, 包括地质年代纪年、公元纪年、天文纪年等, 缺省采用公元纪年(也是不同纪年

方式的转换基准), 地质时间则采用地质年代纪年; *Calen*表示历法, 主要包括阳历、阴历、各国家或地区采用的历法, 缺省采用阳历(也是不同历法之间的转换基准); *Tunit*为时间单位, 不同时间值需要转换到相同的时间单位才能进行计算; *Timing*为计时方式, 包括时区(*Zone*)和时制(*TSys*), 中国的时区为东八区, 缺省的时制采用24小时制.

式(3)中*LCn*为地理位置名称, *LCd*为坐标值或高程值. 在空间本体中, 就是要建立地理位置名称与空间坐标及其空间参考系的关联, 以此支持统一的空间计算. 例如, DDE已经利用GeoNames和OpenStreetMap的对齐融合与处理, 建立了5,965,864个地名及其对应的点坐标与面坐标, 可支持全球地名的统一空间计算与关系推理; *LCd*根据空间对象的几何类型, 如果是点可以是单值, 如果是线则为坐标序列, 如果是面则是坐标序列且第一个坐标和最后一个坐标相同; *CdRef*表示空间参考系, 具体形式如式(5)所示:

$$CdRef = CoordSys\,/\,EleSys, [Prjcode], CdUnit, \tag{5}$$

式中, *CoordSys*、*EleSys*和*Prjcode*分别表示坐标系、高程系和投影方式. 在空间本体中, 可参照使用原欧洲石油调查组织(European Petroleum Survey Group, EPSG)发布的坐标参照系统的参数及坐标转换描述数据集. EPSG对已有的椭球体、投影坐标系等不同的组合都赋予不同的标识号(EPSG code). *CdUnit*表示不同的坐标单位, 有地理坐标的度分秒, 也有直角坐标的米、千米等单位. 值得注意的是, 即使是同一坐标单位, 如地理坐标也可能被表达为不同的形式, 例如, 南纬(S)北纬(N)、东经(E)西经(W)的度分秒或十进制度的形式. 为此, 在空间本体中, 也需要对坐标表达形式进行规范化定义.

*MetaK*为地学知识的元知识, 通常用来记录地学知识的版本号(*VerId*)、创建或更新的内容(*Content*)、创建或更新时间(*UpDt*)、地点(*UpSite*, 可选)、责任方(*resPart*, 可以是个人也可以是组织, 包括责任人及联系方式), 以及创建或更新的说明(*Desc*, 创建或更新修改的动机及目的等)等, 具体形式如式(6)所示:

$$MetaK = VerId, Content, UpDt, [UpSite], resPart, Desc. \tag{6}$$

通过*MetaK*可追溯和分析地学各类知识的演变情况, 包括变化的时间、内容及其变化的原因, 编辑修改知识的责任人等.

4.2 不同时空关联度的地学知识图谱表达模型

基于地学知识图谱自适应表达模型, 可以自动形成与其时空特征关联程度相一致的表达模型, 以此完整、准确地表达地学知识, 提高地学知识元组检索的效率与准确率. 不同时空关联度地学知识的具体表达模型如下.

直接表达时间和空间实体的地学知识图谱表达模型分别如式(7)和(8)所示, 其中: P_{T}、P_{L}分别为表示时间和空间的谓词(具体类型参见表2中的各类时空关系).

$$GeoKGAM = \langle S, P_{\mathrm{T}} | T_{\mathrm{o}}, O \rangle \leftrightarrow MetaK, \tag{7}$$

$$GeoKGAM = \langle S, P_{\mathrm{L}} | L_{\mathrm{o}}, O \rangle \leftrightarrow MetaK. \tag{8}$$

与时间、空间都强关联的地学知识图谱表达模型如式(9)所示. 其中: P为主题内容谓词(具体类型参见表2中的语义关系、组成关系、成因关系、计算关系等), P_{T}、P_{L}分别表示时间和空间的关系.

$$GeoKGAM = \langle S, P, O, (P_{\mathrm{T}}, T) | T_{\mathrm{o}}, (P_{\mathrm{L}}, L) | L_{\mathrm{o}} \rangle \leftrightarrow MetaK. \tag{9}$$

仅与时间或空间强关联的地学知识图谱表达模型分别如式(10)和(11)所示:

$$GeoKGAM =< S, P, O, (P_{\mathrm{T}}, T) | T_{\mathrm{o}} \rangle \leftrightarrow MetaK, \tag{10}$$

$$GeoKGAM =< S, P, O, (P_{\mathrm{L}}, L) | L_{\mathrm{o}} \rangle \leftrightarrow MetaK. \tag{11}$$

时空中度关联的地学知识图谱表达模型如式(12)所示, 其中: A_{T}为时间属性, A_{L}为空间属性, A_n为其他属性, 即时间和空间作为一般的属性信息分别记录在不同的元组中.

$$\begin{aligned} &GeoKGAM = \{\langle S, P, O \rangle, \\ &P = \{A_{\mathrm{T}}, A_{\mathrm{L}} ... A_n\} \mid T_{\mathrm{o}}, L_{\mathrm{o}}\} \leftrightarrow MetaK. \end{aligned} \tag{12}$$

弱时空关联的地学知识基本上与时空信息无关, 其知识图谱表达模型如式(13)所示:

$$GeoKGAM = \langle S, P, O \rangle \leftrightarrow MetaK. \tag{13}$$

综上, 根据时空关联度的不同, 可通过不同的元组表达地学知识图谱, 一方面可实现地学知识的准确表达和存储; 另一方面可支持地学知识的高效、精准检索与使用. 同时, 通过与时空本体的关联, 可实现时空信息的统一对齐转换与计算推理.

4.3　地学知识时空关联度计算

地学知识时空关联度是指某一条地学知识与时间和空间信息的紧密程度, 大致可分为以下几种: 一是时空弱关联, 即地学知识与时空信息关联程度非常小, 与时空信息无关或几乎不受时空条件的限制或约束; 二是时空中度关联, 即地学知识与时空信息有着比较紧密的关联, 具有时空属性信息; 三是时空强关联, 即地学知识与时空信息有着非常紧密的关联, 时空信息是地学知识的主体信息, 一旦发生变化, 其地学知识表达的内涵也随之改变; 四是超强关联, 即时间和空间实体对象本身的描述知识.

对于地学基础术语/概念, 基本上与时空不相关, 属于时空弱关联知识. 例如: “地貌学”作为一个基础概念, 本身与时空并不相关, 属于时空弱关联知识. 但对于实例术语/概念, 由于是实例知识的抽象描述, 往往具有时空属性及关系, 属于时空中度关联知识. 例如:“高原”, 作为一个实例概念, 由于需要对具体的实例(事实)知识进行规范化约束, 往往需要定义地理位置、形成年龄等时空属性, 以及“高原”与其他实例概念(如“河流”)可能存在的时空关系, 属于时空中度关联知识. 地学事实性知识存在于一定的时空条件, 时空信息变化后, 其属性特征也将随之变化. 因此, 事实性知识属于时空强关联知识. 如:“黄土高原”, 是一个事实性知识, 是客观世界存在的实体. 描述黄土高原时, 必须指明其地理位置、空间范围、形成过程等, 属于时空强关联知识. 作用于概念性知识的规则往往与时空无关, 属于时空弱关联知识; 而作用于事实性知识的规则通常只适用于该事实知识, 具有一定的时空限制性, 属于时空中度关联知识. 元知识需要记录地学知识创建或更新的时间和位置信息, 特别是时间信息, 属于时间强关联知识.

时空关联度的计算, 可采用基于规则和深度学习两种方法. 基于规则的方法首先确定地学知识与其时空特征关联程度的因子, 然后分别计算这些因子的时空关联程度并通过加权求和法计算总关联度, 最后设置阈值并根据阈值对地学知识时空关联度进行分类. 特征因子包括: (1) 形态学方面的因子, 如共现频率特征(地学知识与时间、空间信息在语料中共同存在的频次)、距离特征(地学知识与其上下文中时空信息的距离远近); (2) 语言学方面的因子, 如知识类别特征(地学知识在知识分类体系中的类别)、头实体类别特征(地学知识三元组中头实体的要素类别)、谓词语义特征(地学知识三元组中谓词的语义信息)等. 基于深度学习的方法则是直接训练深度学习模型, 将地学知识与其时空特征的关联程度进行离散分类, 将地学知识时空关联度计算视为一个多分类任务, 即以地学知识三元组为输入, 以其关联类别为输出. 基于深度学习的时空关联度计算包括训练数据准备、深度学习模型构建、模型训练与验证三个步骤. 深度学习模型通常可综合利用预训练语言模型(例如, 双向编码器表征变换模型, Bidirectional Encoder Representation from Transformers, BERT), 以及双向长短时记忆神经网络(Bi-directional Long Short-Term Memory, BiLSTM)等的组合, 来提升时空关联度分类的精度. 时空关联度计算方法具体实现将另行撰文讨论.

4.4　地学知识图谱自适应表达模型应用流程

地学知识图谱自适应表达模型应用流程如下图5所示. 首先对拟表达的多学科、多类型的地学知识进行时空关联度的计算. 时空关联度计算可采用前述提到的基于规则或深度学习模型的方法; 根据时空关联度(直接表达时空信息的地学知识, 以及与时空特征强关联、中度关联、弱关联的地学知识), 在地学知识图谱自适应表达模型的基础上, 自动选择与时空关联度匹配的更为紧凑和精准的表达模型. 这些表达模型既有共性的主题内容元组、元知识元组的表达以及统一时空本体的支持, 又有个性化地依据时空关联度的时空信息的表达. 采用统一的描述语言和图数据库, 如网络本体语言(Web Ontology Language, OWL)和JanusGraph图数据库, 可对地学知识图谱进行统一的存储管理. 自适应表达模型可根据时空关联度的不同, 将地学知识灵活表达为三元组或紧密关联时空信息的四元组及五元组. 因此, 可利用SPARQL(SPARQL Protocol and RDF Query Language)、GeoSPARQL或Gremlin等查询语言, 实现地学知识更为高效和精准的检索与计算推理.

5　讨论与结语

地学知识图谱通过将各类地学知识组织成计算机可理解、可计算的语义网络, 可实现地学知识的统一认知、精准关联、计算推理与智能服务, 是当前最有

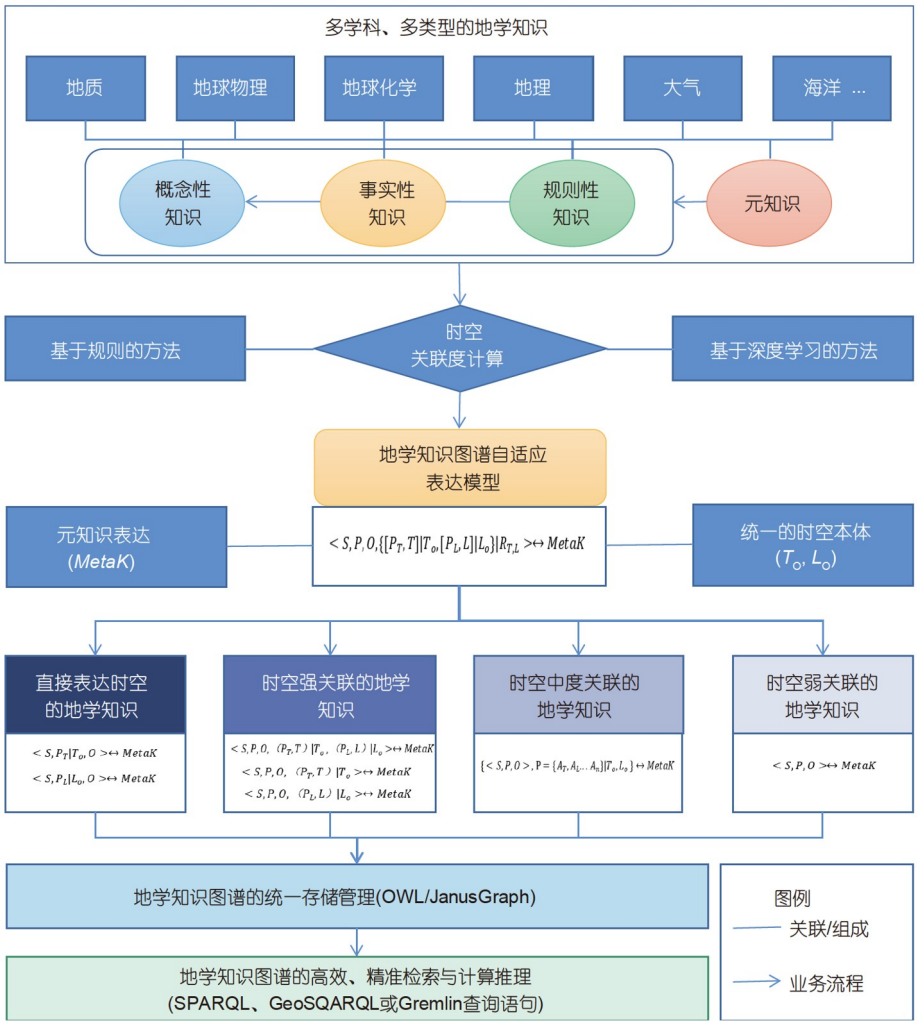

图 5　地学知识图谱自适应表达模型应用流程

效的地学知识组织和服务方式，已经成为基于大数据和人工智能的现代地学研究的基础，正成为地学研究的前沿和热点．地学知识包含众多的学科领域知识，具有复杂的时空特征及关系，呈现出多尺度、多粒度、多维度等特点．因此，面向不同学科和类型的地学知识，建立符合地学知识特点并顾及复杂时空特征及关系的地学知识图谱表达模型，是地学知识图谱构建与应用的基础和前提．

本文首先分析了现有知识图谱表达模型的现状与不足，然后对地学知识类别、地学知识时空特征及关系等进行了系统的剖析．在此基础上，提出可充分表达时空特征与关系的地学知识图谱自适应表达模型．该模型的核心是在统一时空本体的约束下，基于地学知识的时空关联度，采用不同元组对不同类型的地学知识进行表达．通过自适应表达模型，一方面可以实现地学知识的高效表达和存储；另一方面可支持地学知识高效精准地检索与使用；同时，通过与时空本体的关联，可实现时空信息的统一对齐转换与计算推理．

地学知识图谱自适应表达模型研究工作刚刚开始，并在深时数字地球国际大科学计划(DDE)地质年代知识图谱等相关研究中(Wang等, 2022; Qi等, 2022)进行了初步试验，但在实际应用中还涉及到其他很多的问题，未来还需要在以下几个方面进行深入的研究，并加强应用验证:

(1) 地学知识图谱自适应表达模型依赖于时空关联度的计算，当前我们主要采用人工的方式定性对地

学知识时空关联程度进行判断，正在开展基于深度学习模型，利用地学知识类型及上下文语境特征等自动对地学知识时空关联度进行定量化计算的研究.

(2) 地学知识图谱自适应表达模型需要统一时空本体的约束，这是多源异构地学知识进行计算和推理的基础，我们将在已有地理空间数据本体(包含时间本体、空间本体、数据形态和来源本体等)研究成果基础上(诸云强和潘鹏, 2019; Sun等, 2019), 不断优化和完善时空本体.

(3) 不同于地学数据的元数据, 地学元知识还没有得到学术界的足够重视，目前还没有相关的标准规范，各类知识图谱也都缺乏元知识的记录，这将极大影响多源异构以及同源但不同版本地学知识的对齐融合，特别是地学知识的演变分析与溯源等. 因此, 必须加快研究地学元知识标准，并在地学知识图谱建设中同步生成地学元知识.

(4) 随着地学知识图谱研究和应用的深入, 地学知识图谱将越来越多，如何建立大规模、高质量的地学知识图谱，进而充分发挥地学知识图谱的作用，赋能现代化地学研究，是地学知识图谱研究的关键. 为此，需要尽快加强自适应表达模型在地学知识图谱中的应用，进行模型合理性、适用性以及模型的存储效率、检索性能等方面的测试比对，为大规模、高效率地学知识图谱的构建提供强有力的支撑.

致谢　本文工作得到了深时数字地球国际大科学计划(Deep-time Digital Earth Big Science Program, DDE)的支持，感谢DDE首席科学家王成善院士、成秋明院士等的指导和讨论.

参考文献

高嘉良, 仇培元, 余丽, 黄宗财, 陆锋. 2020. 基于旅游知识图谱的可解释景点推荐. 中国科学: 信息科学, 50: 1055–1068

高松. 2020. 地理空间人工智能的近期研究总结与思考. 武汉大学学报-信息科学版, 45: 1865–1874

侯志伟, 诸云强, 高星, 潘鹏, 罗侃, 王东旭. 2015. 时间本体及其在地学数据检索中的应用. 地球信息科学学报, 17: 379–390

李新, 袁林旺, 裴韬, 黄昕, 刘广, 郑东海. 2021. 信息地理学学科体系与发展战略要点. 地理学报, 76: 2094–2103

李新, 郑东海, 冯敏, 陈发虎. 2022. 信息地理学: 信息革命重塑地理学. 中国科学: 地球科学, 52: 370–373

刘瑜. 2022. 地理信息科学: 地理学的核心或是外缘? 中国科学: 地球科学, 52: 377–380

刘知远, 韩旭, 孙茂松. 2020. 知识图谱与深度学习. 北京: 清华大学出版社

陆锋, 余丽, 仇培元. 2017. 论地理知识图谱. 地球信息科学学报, 19: 723–734

闾国年, 袁林旺, 俞肇元. 2022. 信息地理学: 地理三元世界的新支点. 中国科学: 地球科学, 52: 374–376

漆桂林, 高桓, 吴天星. 2017. 知识图谱研究进展. 情报工程, 3: 4–25

齐浩, 董少春, 张丽丽, 胡欢, 樊隽轩. 2020. 地球科学知识图谱的构建与展望. 高校地质学报, 26: 2–10

石顺中, 闾海荣, 董少春, 李艳, 唐小芳, 周成虎. 2020. 地球科学知识体系编辑平台. 高校地质学报, 26: 384–394

孙鸿烈. 2017. 地学大辞典. 北京: 科学出版社

王东旭, 诸云强, 潘鹏, 罗侃, 侯志伟. 2016. 地理数据空间本体构建及其在数据检索中的应用. 地球信息科学学报, 18: 443–452

王志华, 杨晓梅, 周成虎. 2021. 面向遥感大数据的地学知识图谱构想. 地球信息科学学报, 23: 16–28

熊巨华, 王佳, 张晴, 耿豪鹏, 官冬杰, 史云飞, 李飞, 张新林, 王天富. 2021. 地理科学的学科体系构建与内涵. 科学通报, 66: 153–161

姚玉鹏, 郭正堂. 2021. 落实地球科学战略规划, 优化学科资助格局. 科学通报, 66: 143–144

俞肇元, 袁林旺, 吴明光, 周良辰, 罗文, 张雪英, 闾国年. 2022. 地理学视角下地理信息的分类与描述. 地球信息科学学报, 24: 17–24

张雪英, 张春菊, 吴明光, 闾国年. 2020. 顾及时空特征的地理知识图谱构建方法. 中国科学: 信息科学, 50: 1019–1032

周成虎, 王华, 王成善, 侯增谦, 郑志明, 沈树忠, 成秋明, 冯志强, 王新兵, 闾海荣, 樊隽轩, 胡修棉, 侯明才, 诸云强. 2021. 大数据时代的地学知识图谱研究. 中国科学: 地球科学, 51: 1070–1079

周园春, 王卫军, 乔子越, 肖濛, 杜一. 2020. 科技大数据知识图谱构建方法及应用研究综述. 中国科学: 信息科学, 50: 957–987

诸云强, 潘鹏著. 2019. 地理空间数据本体概念、技术方法与应用. 北京: 科学出版社

诸云强, 孙凯, 胡修棉, 闾海荣, 王新兵, 杨杰, 王曙, 李威蓉, 宋佳, 苏娜, 牟兴林. 2023. 大规模地球科学知识图谱构建与共享应用框架研究与实践. 地球信息科学学报, 25: 1215–1227

Bittner T, Donnelly M, Smith B. 2009. A spatio-temporal ontology for geographic information integration. Int J Geogr Inf Sci, 23: 765–798

Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. 2008. Freebase: A collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. 1247–1250

Brown T, Mann B, Ryder N, Subbiah M, Kaplan J D, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei

D. 2020. Language models are few-shot learners. Adv Neural Inf Process Syst, 33: 1877–1901

Buccella A, Cechich A, Gendarmi D, Lanubile F, Semeraro G, Colagrossi A. 2011. Building a global normalized ontology for integrating geographic data sources. Comput Geosci, 37: 893–916

Cox S J D, Richard S M. 2015. A geologic timescale ontology and service. Earth Sci Inform, 8: 5–19

Feng M, Liu S, Euliss Jr. N H, Young C, Mushet D M. 2011. Prototyping an online wetland ecosystem services model using open model sharing standards. Environ Model Software, 26: 458–468

Galton A. 2009. Spatial and temporal knowledge representation. Earth Sci Inform, 2: 169–187

Hoffart J, Suchanek F M, Berberich K, Weikum G. 2013. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artif Intelligence, 194: 28–61

Hogan A, Blomqvist E, Cochez M, D'amato C, Melo G D, Gutierrez C, Kirrane S, Gayo J E L, Navigli R, Neumaier S, Ngomo A C N, Polleres A, Rashid S M, Rula A, Schmelzeisen L, Sequeda J, Staab S, Zimmermann A. 2022. Knowledge Graphs. ACM Comput Surv, 54: 1–37

Hui B, Zhang L Z, Zhou X, Wen X, Nian Y H. 2022. Personalized recommendation system based on knowledge embedding and historical behavior. Appl Intell, 52: 954–966

Ji S X, Pan S R, Cambria E, Marttinen P, Philip S Y. 2022. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Trans Neural Netw Learn Syst, 33: 494–514

Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes P N, Hellmann S, Morsey M, van Kleef P, Auer S, Bizer C. 2015. DBpedia—A large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web, 6: 167–195

Li X, Feng M, Ran Y H, Su Y, Liu F, Huang C L, Shen H F, Xiao Q, Su J B, Yuan S W, Guo H D. 2023. Big Data in Earth system science and progress towards a digital twin. Nat Rev Earth Environ, 4: 319–332

Ma X G. 2022. Knowledge graph construction and application in geosciences: A review. Comput Geosci, 161: 105082

Ma X G, Fox P. 2013. Recent progress on geologic time ontologies and considerations for future works. Earth Sci Inform, 6: 31–46

Pellissier T T, Weikum G, Suchanek F. 2020. Yago 4: A reason-able knowledge base. In: Harth A, et al. The Semantic Web: 17th International Conference, ESWC 2020. Heraklion, Crete, Greece. Springer International Publishing. 583–596

Qi Y M, Du H S, Farjudian A, Zhu Y Q. 2022. Representing computational relations in knowledge graphs. In: 15th International Conference on Spatial Information Theory (COSIT 2022). 29: 1–29:7

Raskin R G, Pan M J. 2005. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET). Comput Geosci, 31: 1119–1125

Reinanda R, Meij E, de Rijke M. 2020. Knowledge graphs: An information retrieval perspective. FNT Inf Retrieval, 14: 289–444

Schulman J, Zoph B, Kim C, Hilton J, Menick J, Weng J, Ceron Uribe J F, Fedus L, Metz L, Pokorny M, Lopes R G, Zhao S, Vijayvergiya A, Sigler E, Perelman A, Voss C, Heaton M, Parish J, Cummings R, Ryder N. 2022. Chat GPT: Optimizing language models for dialogue. https://openai.com/blog/chatgpt

Sun K, Zhu Y Q, Pan P, Hou Z W, Wang D X, Li W R, Song J. 2019. Geospatial data ontology: The semantic foundation of geospatial data integration and sharing. Big Earth Data, 3: 269–296

Vrandečić D, Krötzsch M. 2014. Wikidata: A free collaborative knowledgebase. Commun ACM, 57: 78–85

Wang C S, Hazen R M, Cheng Q M, Stephenson M H, Zhou C H, Fox P, Shen S Z, Oberhansli R, Hou Z Q, Ma X G, Feng Z Q, Fan J X, Ma C, Hu X M, Luo B, Wang J L, Schiffries C M. 2021. The deep-time digital Earth program: Data-driven discovery in geosciences. Natl Sci Rev, 8: nwab027

Wang S, Zhang X Y, Ye P, Du M, Lu Y X, Xue H N. 2019. Geographic knowledge graph (GeoKG): A formalized geographic knowledge representation. Isprs Int J Geo-Inf, 8: 184

Wang S, Zhu Y Q, Qi Y M, Hou Z W, Sun K, Li W R, Hu L, Yang J, Lv H R. 2022. A unified framework of temporal information expression in geosciences knowledge system. Geosci Front, 14: 101465

Zhang X Y, Huang Y, Zhang C J, Ye P. 2022. Geoscience knowledge graph (GeoKG): Development, construction and challenges. Trans GIS, 26: 2480–2494

Zhao F, Li Y G, Hou J, Bai L. 2022. Improving question answering over incomplete knowledge graphs with relation prediction. Neural Comput Appl, 34: 6331–6348

Zheng K, Xie M H, Zhang J B, Xie J, Xia S H. 2022. A knowledge representation model based on the geographic spatiotemporal process. Int J Geogr Inf Sci, 36: 674–691

Zhou J, Pei K E, Qiu X P, Huang M L, Zhang J P. 2023. ChatGPT: Potential, prospects, and limitations. Front Inform Technol Electron Eng, doi: 10.1631/FITEE.2300089

(责任编委: 李新)