顾及时空特征的地理知识图谱构建方法

张雪英^{1,2*}, 张春菊³, 吴明光^{1,2}, 闫国年^{1,2}

1. 虚拟地理环境教育部重点实验室(南京师范大学), 南京 210023

2. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023

3. 合肥工业大学土木与水利工程学院, 合肥 230009

* 通信作者. E-mail: zhangsnowy@163.com

收稿日期: 2019-12-01; 接受日期: 2020-01-20; 网络出版日期: 2020-07-06

国家自然科学基金(批准号: 41971337, 41631177, 41671393)资助项目

摘要 地理知识是人类对地理现象或事物空间分布、演变过程和相互作用规律的认知结果. 当前大数据环境下的地理信息服务, 普遍存在“数据海量、信息爆炸、知识难求”现象. 地理知识图谱是一种利用语义网络对地理概念、实体及其相互关系进行形式化描述的知识系统, 在地理知识理解、地学问题求解、时空预测决策等方面具有巨大的应用潜力. 地理知识除了具有通用知识的内涵和特点之外, 还具有特定的时空特征和地学机理特点. 因此, 地理知识图谱构建和应用既具有一定的通用性, 同时具有一定的专业特殊性. 本文结合地理知识的时空特征和知识图谱的表达形式, 提出了一种顾及时空特征的地理知识图谱构建方法. 首先, 系统梳理了地理知识图谱构建的基本思路和技术流程, 并简要阐述了地理知识获取、地理知识抽象与表达、地理知识组织与管理 3 个关键环节的主要研究内容及其进展. 其次, 从地理学回答的基本问题出发, 对地理知识的内容特征进行概括和抽象, 构建了涵盖“地理概念—地理实体—地理关系”3 个层次的地理知识表达模型, 用于描述不同粒度地理知识语义单元的基本组成及其逻辑关系. 最后, 借鉴知识图谱采用的语义网络知识表示方法, 提出了基于“过程—关系”的地理知识表示方法. 该方法以时间和空间特征为基本条件, 以地理实体的状态划分为基础, 实现了地理实体演化过程以及复杂地理关系的形式化描述. 本文的研究成果有效解决了融合时空间维度的地理知识结构化表达和形式表示问题, 为地理知识获取、融合、推理与应用奠定了基础. 同时, 在地质、环境、气象等地学领域具有一定通用性, 对地学知识服务的推进具有重要参考价值.

关键词 地理实体, 时空特征, 地理知识表达模型, 地理知识形式化, 地理知识图谱

1 引言

知识图谱本质上是一种大型的语义网络, 用以描述客观世界中的概念、实体及其相互关系^[1]. 知识图谱以更加接近人的思维形式, 构建在计算机世界中表示信息和知识的有效载体, 成为人工智能应

引用格式: 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. 中国科学: 信息科学, 2020, 50: 1019–1032, doi: 10.1360/SSI-2019-0269
Zhang X Y, Zhang C J, Wu M G, et al. Spatio-temporal features based geographical knowledge graph construction (in Chinese). Sci Sin Inform, 2020, 50: 1019–1032, doi: 10.1360/SSI-2019-0269

图 1 地理知识图谱构建的技术流程

Figure 1 Technique framework of geographical knowledge graph

用的重要基础设施, 在语义搜索、智能问答、预测决策等方面凸显了越来越重要的应用价值 [2]. 2012 年, 谷歌推出基于知识图谱的搜索引擎, 推动了互联网搜索从字符 (string) 到事物 (thing) 的跨越式发展. 之后, 知识图谱迅速引起学术界和工业界的广泛关注, 比较有代表性的开放知识图谱包括 DBpedia, YAGO, NELL, Probase, Zhishi.me 等.

地理知识是人类对地理事物或现象空间分布、演变过程和相互作用规律的认识结果, 可以分为事实型知识、规则及控制型知识、决策型知识 3 种类型. 当前大数据环境下的地理信息服务, 普遍存在“数据海量、信息爆炸、知识难求”的现象 [3]. 地理知识图谱是一种对地理概念、实体及其相互关系进行形式化描述的知识系统 [4,5], 能够提供系统的、深层次的结构化地理知识. 因此, 地理知识图谱成为当前地理信息服务向地理知识服务拓展的关键, 在地理知识理解、地理知识搜索、时空预测等方面具有巨大的应用潜力.

近年来, 地理知识图谱逐步成为地理信息科学领域的研究热点, 但是尚处于概念探讨和初步实验阶段 [3]. 总体上, 地理知识图谱借鉴了现有知识图谱的通用技术, 重点解决地理信息语义理解、地理知识融合、地理知识质量评价等关键问题 [6]. 然而, 地理知识除了具有通用知识的内涵和特点之外, 还与地理现象或事物的演化过程密切相关, 具有特定的时空特征和地学机理特点 [7]. 因此, 如何从不同类型的地理数据中获取地理知识? 如何构建适应地理知识特点的地理知识表达模型? 如何利用知识图谱技术进行地理知识的管理、检索和推理? 如何挖掘地理知识图谱的应用场景? 这一系列问题, 无疑是当前地理知识图谱研究与应用面临的挑战. 其中, 融合时空维度的地理知识抽象与表达是地理知识获取、融合、推理与应用的基础, 也是地理知识图谱构建的核心科学问题之一.

2 基本思路与技术流程

知识图谱构建和应用具有较大程度的通用性, 同时也具有一定的专业特殊性. 本文在借鉴现有知识图谱通用技术的基础上, 综合考虑地理知识的独特性和地理知识来源的数据形式, 提出一种顾及时空特征的地理知识图谱构建方法. 主要内容包括地理知识获取、地理知识抽象与表达、地理知识组织与管理 3 个部分, 具体的技术流程如图 1 所示.

2.1 地理知识获取

地理学语言是人类理解、表达与传播地理信息的重要工具, 主要包括自然语言、地图和 GIS. 不同

地理学语言形成了地理知识的有效载体,分为自然语言和图形语言两种形式.其中,文本和地图是最基本的、最主要的两种地理知识来源.从数据结构形式来看,地理知识来源于各种形式的结构化数据、半结构化数据和非结构化数据.近年来,由于网络文本资源呈爆发式增长趋势,而且自然语言处理技术发展迅速,人们比较关注以文本为主的自然语言中地理信息理解问题^[6].相比较而言,地图中地理知识的计算机理解尚未有突破性的进展.现有的研究主要集中在基于图像处理的地图符号和文字注记识别方法^[8~15].因此,下文主要阐述文本数据中地理知识获取方法.

(1) 地理实体识别是指由计算机从文本中抽取地理实体名称.早期采用的地名词典匹配方法,具有容易实现、准确性高的优点,但受到词典完备性和时效性的限制^[16,17].之后,隐马尔可夫(Hidden Markov)、最大熵、支持向量机、条件随机场等机器学习模型被广泛应用于地理实体识别^[18~20],可以显著提高地理实体识别的性能,但是对语料库具有较强的依赖性^[21~24].研究表明,深度学习模型能够在深层结构上自动学习上下文特征,对语料库的依赖性明显低于其他机器学习方法,而且能够获得较好的地理实体识别性能^[25,26].此外,由于文本中同一地理实体名称可能指称多个物理空间位置,还需要借助地名词典、地理本体、上下文特征等进行地理实体消歧^[27].

(2) 时间信息贯穿于地理实体产生、发展、消亡的整个过程中,能够反映地理实体的状态和演变过程.时间信息可分为显性和隐性两种类型.其中,显性时间是指具有比较明确概念的通用时间描述,如“2018年”、“星期六”、“六月五日”等.因此,显性时间信息的语言特征比较明显,可以归纳总结特定的语法规则^[28,29].隐性时间是诸如“地震发生的时候”之类隐藏在语义之中的信息,主要采用条件随机场、最大熵、神经网络等模型进行信息抽取^[30~33].

(3) 自然语言中描述的空间信息主要包括地名、地址、坐标和空间关系等.其中,地名和地址信息抽取可以参考地理实体识别方法.在此基础上,以标准地名地址数据库为参考,借助理编编码技术,建立自然语言描述的空间信息与地理坐标之间的映射关系.自然语言中空间信息描述普遍存在一定的模糊性,可以通过认知实验^[34,35]、空间聚类^[17,36]、空间关系计算^[37,38]等方式进行模糊空间建模.

(4) 属性是对于地理实体某一特征的刻画,属性值是赋予属性的特定值.属性抽取需要获取到与地理实体相关的属性类型及其属性值,是挖掘有价值语义单元的重要手段.目前,地理属性信息抽取主要采用本体法、规则匹配法^[39]、监督学习法^[40,41]、弱监督学习法^[22].此外,从文本中抽取的属性信息,还需要进行属性值的标准化^[5,24,42~44].

(5) 地理关系抽取是从文本中辨别出两个地理概念或者实体之间存在的语义关系,如时间关系、空间关系、属性关系、状态关系等.现有的研究侧重于时间关系和空间关系抽取.由于时间作为一种天然有序的概念,本身具备推理和计算的能力.时间关系理论定义了“相同”、“重叠”、“之前”等13类时间关系^[45].因此,按照常识制定相关规则,可以有效实现文本中时间关系的抽取和计算.由于每种语言都有一套能够完整表达空间关系的词汇系统,空间关系抽取难度较大.在构建空间关系词典和句法模式库的基础上,采用模式匹配的方法可以实现以句子为单位的空间关系抽取^[16,46~49].机器学习方法可以有效避免规则的局限性,但是依赖大规模、高质量的空间关系标注语料库^[20,50~52].采用回标技术可以快速自动构建地理实体关系语料库,但是需要提供可靠、丰富的开放地理知识资源^[53].最新研究表明,深度学习模型能更有效地挖掘描述地理关系的句法信息、位置信息、上下文信息等复杂的语言特征^[54].同时,通过构建本体、事实和同义词相结合的地理关系知识库,可以用于地理关系抽取结果的过滤筛选^[55].

(6) 地理知识融合:不同数据来源的地理知识描述存在一定的互补性和差异性,如分类体系不统一、地理实体歧义、特征描述详略不同、实体关系冲突等信息冗余和不一致问题.具体来讲,可以从概念、实体、特征和关系4个层次进行地理知识融合.首先,按照统一的概念分类体系,采用语义网对

齐技术进行地理知识的概念映射, 解决地理知识的粒度和等级不一致问题; 其次, 综合考虑地理实体类型、文本相似度和属性特征等, 建立各个概念层次的地理实体链接, 解决不同数据来源中地理实体知识的歧义问题; 最后, 利用冲突检测、真值发现等技术, 对不同地理实体关系知识进行去重、关联和合并。

2.2 地理知识抽象与表达

在信息世界中, 需要将现实世界存在的各种地理知识描述转换成计算机可识别和处理的内容, 形成尽可能全面的地理知识表示, 使得机器可以通过学习这些知识, 表现出类似于人类的行为。因此, 地理知识抽象与表达对于地理知识图谱构建具有关键作用。主要内容包括两个方面: 一是构建地理知识表达模型, 通过概括和抽象地理知识的内容特征, 对地理知识语义单元的基本组成及其逻辑关系进行结构化描述; 二是面向知识图谱的地理知识表示, 特别是基于语义网络的地理知识表示。具体内容将在第 3 和 4 节详细阐述。

2.3 地理知识组织与管理

知识图谱可以采用基于表结构和图结构的存储方式。目前业界公认的图数据模型包括属性图、资源描述框架 (resource description framework, RDF) 和三元组超图, 其中前两种广泛应用到图数据库产品中。因此, 知识图谱可以采用关系型数据库和图数据库 (例如 Neo4j, HugeGraph, TigerGraph 等)。相比较而言, 前者的搜索效率较为低下, 而后者更利于数据读写、存储和查询。然而, 通用的知识图谱数据组织和管理技术并不能完全适用于地理知识图谱。通常情况下, 可以采用数据库扩展的方法。例如, 对 RDF 关系数据库引擎进行扩展, 实现空间信息的存储、索引和查询^[56,57]; 对 RDF 图数据引擎增加语义和空间混合索引, 开发融合空间信息的图数据库引擎^[58]。尽管如此, 仍然无法解决地理实体状态、过程和地理关系等复杂地理知识的存储、索引和查询问题。

3 层次化的地理知识表达模型

3.1 地理知识的内涵与内容分析

地理学研究地理要素和地理综合体的空间分异规律、时空演化过程及区域特征, 是研究人地关系的科学^[59]。具体来讲, 地理学是关于地球上不同区域特点、人和事物分布, 以及事件发生和发展规律的科学, 侧重考虑在特定地点和位置的人与环境的交互作用。其中, 地理对象包括地球上的不同区域、人、事物、事件、现象等。地理学主要回答 6 类关于地理对象的问题, 即“它在哪里? 它是如何发生的? 它是什么样子的? 它为什么在那里? 它产生了什么影响? 怎样使它有利于人类和自然环境?”^[60]。因此, 我们可以认为, 地理知识就是对地理对象的空间分布规律、演化过程及其相互作用的解释, 能够用于回答上述类型的地理学问题。地理对象具有时间、空间、属性等基本特征, 且这些特征之间存在一定的逻辑关系。其中, 时间和空间是地理对象存在的基本条件。属性是属于地理对象本质方面的特性, 狭义的属性仅指性质。但是, 在逻辑上属性还包括关系、状态、行为和过程等特性。因此, 为了区分各种属性及其逻辑关系, 地理对象应具有时间、空间、属性、状态、行为和过程等特征。图 2 描述了地理知识的组成及其作用。

3.2 地理知识表达模型

根据地理知识的类型、特征及其逻辑关系, 地理知识可以划分为 3 个层次: 地理概念层、地理实

图 2 地理知识的组成及其作用

Figure 2 Geographical knowledge elements and their functions

体层和地理关系层. 其中, 地理实体知识可以分为 3 个子层次, 包括基本特征 (时间、空间、属性、行为)、状态特征和过程特征; 地理关系知识可以分为两个子层次: 概念关系和特征关系. 按照地理知识的层次划分, 可以形成不同粒度的地理知识语义单元.

3.2.1 地理概念分类体系

由于地理事物或现象存在多样性和复杂性等特点, 可按照不同的标准对地理概念进行分类和定义. 例如, 按照学科领域可以分为自然地理和人文地理; 按照时空尺度分为区域地理、历史地理、城市地理等; 按照地理信息类型可以分为基础地理、专题地理、综合地理等. 通常情况下, 地理概念表达需要参照特定的分类体系. 鉴于地理概念分类体系的多样性, 本文对地理概念的分类体系不再赘述. 需要强调的是, 在同一地理知识图谱中, 需要采用统一的地理概念分类体系.

3.2.2 地理实体表达模型

地理实体主要包括时间、空间、属性、行为、状态、过程等特征. 地理实体表达模型主要用于对地理实体特征及其逻辑关系进行结构化描述 (图 3).

从图 3 可以看出, 依据地理实体的各种特征及其关系, 可以构成不同粒度、不同层次的地理知识单元. 每个知识单元可以根据实际应用需求进行扩展细化, 且不同知识单元之间存在继承或者聚合等形式. 其中, 时间和空间是地理实体存在的前提条件, 也是地理知识表达的基本框架; 属性特征用于表征地理实体本身性质的数据或数据量, 且不同类型的地理实体具有特定的属性特征; 行为特征用于描述地理实体的各种活动; 在特定的时间、空间条件下, 结合地理实体所具有的属性特征和行为特征, 形成了地理实体的不同状态; 依据不同状态的时序关系, 能够刻画地理实体的演变过程. 依据时空范围跨度可以将过程进行层次分级, 从大到小依次为过程整体、过程阶段和过程序列, 呈现倒金字塔结构特性 [61].

图 3 地理实体知识表达模型
Figure 3 Knowledge representation model of geographical entities

图 4 地理关系表达模型
Figure 4 Knowledge representation model of geographical relations

3.2.3 地理关系表达模型

地理关系可以分为概念关系和特征关系两种类型. 其中, 概念关系是指在同一分类体系中, 不同地理概念所属概念之间的语义关系, 包括等同关系、上位关系、下位关系、同位关系和相关关系等; 特征关系是不同地理实体特征之间的关系, 例如时间关系、空间关系、属性关系、状态关系、过程关系等. 地理关系表达模型用以描述地理实体的特征关系和地理实体之间的概念关系, 如图 4 所示.

4 面向知识图谱的地理知识表示

知识表示是把知识用计算机可接受的符号并以某种形式描述出来. 目前, 地理知识表示主要采用逻辑表示法和语义网络表示法.

4.1 逻辑表示法

逻辑表示法与人类自然语言比较接近, 是最早使用的知识表示方法之一. 常用的逻辑语言有命题

逻辑、谓词逻辑和描述逻辑,三者对知识的表达能力逐渐提升.在经典描述逻辑 ALC (attribute concept description language with complements) 的基础上,通过融入适用于空间、状态和过程描述的逻辑语言,可以实现空间和语义的知识表达和推理^[62,63];产生式表示法具备模块性、清晰性和自然性等特点,设计空间关系的产生式规则,可用于支持地理知识库或专家系统的构建^[64,65];框架表示法以框架理论为基础,能把陈述性知识与过程性知识相结合,例如文献^[66]提出的一种地理知识框架,由领域、对象、时间、GeoAgent 及关系等五元组构成.逻辑表示法容易造成知识间的语义关联性弱,不宜用来表示和求解地理事物或现象的时空特征、演化过程及其相互作用规律.

4.2 语义网络表示法

语义网络 (semantic network) 是一种采用网络形式表示思维知识、描述人类对事物认识的方法.本质上,语义网络是一种哲学认识论的方法,同时也是一种知识表达和处理的工具.语义网络侧重于对概念之间关系模型的构建,以描述逻辑为核心的本体知识表示体系.语义网 (semantic web) 通过给万维网上的文档添加能够被计算机所理解的语义“元数据”,从而使整个互联网成为一个通用的信息交换媒介.语义网实现了概念与概念之间的联系,构建起数据之间的链接.基于链接数据开发出一批高质量的知识库,例如 Google Knowledge Graph, DBpedia 和 YAGO 等,为谷歌知识图谱项目奠定了坚实基础.地理语义网作为语义网络的一个子集,将描述对象分为概念、属性和实例 3 类.其中,概念是对具有相同特征的一组对象的抽象定义,属性包括数据属性和对象属性,实例是指定概念的具象表达^[67].目前,比较有代表性的地理语义网有 Geo-Names Ontology, OSM Semantic Network, Linked-GeoData, GeoWordNet 等.

4.3 基于语义网络的地理知识表示

4.3.1 基于“实体 – 关系”的地理知识表示

知识图谱是一种基于语义网络的知识库系统,以一种统一的方式表示知识定义 (schema) 和知识实例 (instance) 两个层次的知识,使得知识获取、知识融合和知识推理在可操作性和可计算性方面具有显著优势.知识图谱中最基本的语义单元用 (节点 1, 关系, 节点 2) 的三元组形式表示,其中节点可以代表概念、实体和属性.图 5 为一种典型的地理知识图谱样例,其中“(中国, 首都, 北京)”为一个语义单元,“中国”和“北京”为两个实体,“首都”表示关系,每个实体还有人口、面积、经度和纬度等属性信息^[68].很显然,这种地理知识图谱表达存在以下问题:一是缺失对时间和空间特征的建模,不能体现地理知识特定的时空特征和地学机理特点.人们普遍认为,时间和空间被认为是一切事物存在和发展的基本条件.本质上,知识是主体关于事物所处状态及其变化规律的表述.因此,不论是通用的知识图谱还是特定领域的地理知识图谱,时间和空间应该是知识描述和表达的基本框架和重要维度.二是现有知识图谱中的“实体”是指具有可区别性且独立存在的某种事物,以及对应特定概念和实体的属性.因此,此“实体”并不能直接对应地理知识描述与表达的主体.三是现有的语义网络只能表达两个实体之间的一种关系,例如“北京是中国的首都”,但是,不能同时表达“北京”与“中国”的空间关系和属性关系(比如人口、面积、GDP 等).

4.3.2 “过程 – 关系”的地理知识表示

参考层级化的地理知识表达模型,可以构建基于“过程 – 关系”的地理知识图谱表达形式.语义单元采用 (节点 1, 关系, 节点 2) 的三元组形式,其中“节点”可以表示地理概念、地理实体、地理概念/地理实体的基本特征(包括时间、空间、属性和行为),“关系”表示地理实体的状态、过程和地理关

图 5 面向“实体 - 关系”的地理知识图谱样例

Figure 5 A sample of a geographical knowledge graph based on "entity-relation"

系. 具体表达规则如下: 首先, 将时间和空间特征作为地理实体状态划分的基本条件, 以刻画地理知识的时空特征; 其次, 通过区分地理实体的基本特征和复合特征 (状态和过程), 利用语义网络的关系表示形式可以描述地理实体自身的演化过程; 然后, 将地理概念/实体之间的边用于表达概念关系, 而将地理实体特征之间的边用于表达特征关系, 能够更加准确、全面地表达地理概念/实体之间的不同关系. 例如, 图 6 中的“北京”和“中国”两个节点, 两者之间的边表示的隶属关系是一种表示上下位的概念关系; 而 A1 和 A2 为两个表示人口属性的节点, 两者之间的边表示“北京”和“中国”在特定状态下的人口对比关系. 因为, “北京”和“中国”的人口数量是不断变化的, 只有在相同的时间条件下, 二者才具有可比性.

5 结论

知识图谱构建和应用技术属于知识工程领域的研究范畴, 被认为是人工智能的重要基础设施, 成为近年来人工智能领域和交叉学科领域的研究热点. 目前, 国内外相关研究主要集中在知识获取、知识融合、知识检索、知识推理等方面. 知识图谱构建和应用具有较大程度的通用性, 同时也具有一定的专业特殊性. 近年来, 地理知识图谱逐步成为地理信息科学领域的研究热点, 但是尚处于概念探讨和初步实验阶段, 亟待从地理学、语言学 and 人工智能等多个学科视角, 突破地理知识建模、非结构化地理知识获取、地理知识图谱管理和普适化应用的关键科学问题.

本文结合地理知识的时空特征和知识图谱的表达形式, 提出了一种顾及时空特征的地理知识图谱构建方法. 主要创新之处体现在两个方面: 一是从地理学回答的基本问题出发, 对地理知识的内涵进行了界定, 构建了涵盖“地理概念 - 地理实体 - 地理关系”3 个层次的地理知识表达模型, 为“地理数

图 6 面向“过程-关系”的地理知识图谱表示样例

Figure 6 A sample of a geographical knowledge graph based on “process-relation”

据-地理信息-地理知识”转换提供了基础框架;二是基于“过程-关系”的地理知识图谱表示方法,充分利用了现有知识图谱统一表示方式的可操作性和可计算性优势。同时,将同一地理实体节点扩展为不同状态的地理实体分节点,巧妙实现了地理实体过程特征的形式化描述。此外,本文提出的地理知识图谱构建方法具有一定通用性,可以推广应用于地质、环境、气象等地理学领域。

后续研究工作重点围绕两个方面展开:一是系统、深入地研究多模态数据中地理知识理解和融合方法;二是探索基于图数据库的地理知识图谱组织和管理技术,包括基于“过程-关系”的地理知识图谱数据模型和基于层次化地理知识表达模型的索引方法。总之,通过一系列相关理论、方法和技术的研究,有望实现地理知识图谱构建的系统化、智能化、工程化,以进一步揭示地理知识的认知机理和描述规律,推动地理知识资源的深度挖掘和开发利用。

参考文献

- 1 James P. Knowledge graphs. In: Linguistic Instruments in Knowledge Engineering. Amsterdam: Elsevier, 1992. 97-117
- 2 Zhao J, Liu K, He S Z, et al. Knowledge Map. Beijing: Higher Education Press, 2018. 293 [赵军, 刘康, 何世柱, 等. 知识图谱. 北京: 高等教育出版社, 2018. 293]
- 3 Chen J, Liu W Z, Wu H, et al. Basic issues and research agenda of geospatial knowledge service. Geomat Inform Sci Wuhan Univ, 2019, 44: 38-47 [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. 武汉大学学

- 报 (信息科学版), 2019, 44: 38–47]
- 4 Chen J Y, Deng S M, Chen H J. CrowdGeoKG: crowdsourced geo-knowledge graph. In: Proceedings of China Conference on Knowledge Graph and Semantic Computing, Singapore, 2017
- 5 Ballatore A, Bertolotto M, Wilson D C. A structural-lexical measure of semantic similarity for geo-knowledge graphs. ISPRS Int J Geo-Inf, 2015, 4: 471–492
- 6 Lu F, Yu L, Qiu P Y. On geographic knowledge graph. J Geo-Inf Sci, 2017, 19: 723–734 [陆锋, 余丽, 仇培元. 论地理知识图谱. 地球信息科学学报, 2017, 19: 723–734]
- 7 Lin H, You L, Hu C B, et al. Prospect of geo-knowledge engineering in the era of spatio-temporal big data. Geomat Inform Sci Wuhan Univ, 2018, 43: 2205–2211 [林珏, 游兰, 胡传博, 等. 时空大数据时代的地理知识工程展望. 武汉大学学报 (信息科学版), 2018, 43: 2205–2211]
- 8 Miao Q G, Xu P F, Li X L, et al. The recognition of the point symbols in the scanned topographic maps. IEEE Trans Image Process, 2017, 26: 2751–2766
- 9 Yang D. Research on map symbol recognition based on neural network. Dissertation for Master's Degree. Guiyang: Guizhou University, 2007 [杨丹. 基于神经网络的地图符号识别的研究. 硕士学位论文. 贵阳: 贵州大学, 2007]
- 10 Hoermann S, Henzler P, Bach M, et al. Object detection on dynamic occupancy grid maps using deep learning and automatic label generation. In: Proceedings of IEEE Intelligent Vehicles Symposium, Changshu, 2018. 826–833
- 11 Luo X Y, Zhang J, Cao X B, et al. Object-aware power line detection using color and near-infrared images. IEEE Trans Aerosp Electron Syst, 2014, 50: 1374–1389
- 12 Tian K. Research on point symbols recognition in color topographic map. Dissertation for Master's Degree. Xi'an: Xidian University, 2017 [田宽. 彩色地形图中的点状符号识别算法研究. 硕士学位论文. 西安: 西安电子科技大学, 2017]
- 13 Feldman J. Perceptual grouping by selection of a logically minimal model. Int J Comput Vision, 2003, 55: 5–25
- 14 Hancer E, Samet R. Advanced contour reconnection in scanned topographic maps. In: Proceedings of International Conference on Application of Information & Communication Technologies, New York, 2011. 1–5
- 15 Pezeshk A, Tutwiler R L. Automatic feature extraction and text recognition from scanned topographic maps. IEEE Trans Geosci Remote Sens, 2011, 49: 5047–5063
- 16 Le X Q, Yang C J, Yu W Y. Spatial concept extraction based on spatial semantic role in natural language. Geomat Inform Sci Wuhan Univ, 2005, 30: 1100–1103 [乐小虬, 杨崇俊, 于文洋. 基于空间语义角色的自然语言空间概念提取. 武汉大学学报 (信息科学版), 2005, 30: 1100–1103]
- 17 Jones C B, Purves R S. Geographical information retrieval. Int J Geograph Inf Sci, 2008, 22: 219–228
- 18 Qian J, Zhang Y J, Zhang T. Research on Chinese person name and location name recognition based on maximum entropy model. J Chin Mini-Micro Comput Syst, 2006, 27: 1761–1765 [钱晶, 张玥杰, 张涛. 基于最大熵的汉语人名地名识别方法研究. 小型微型计算机系统, 2006, 27: 1761–1765]
- 19 Jiang W M. Research on extraction of spatial orientation relations for chinese texts. Dissertation for Ph.D. Degree. Nanjing: Nanjing Normal University, 2010 [蒋文明. 面向中文文本的空间方位关系抽取方法研究. 博士学位论文. 南京: 南京师范大学, 2010]
- 20 Wu L, Liu L, Li H R, et al. A Chinese toponym recognition method based on conditional random field. Geomat Inform Sci Wuhan Univ, 2017, 42: 150–156 [鄂伦, 刘磊, 李浩然, 等. 基于条件随机场的中文地名识别方法. 武汉大学学报 (信息科学版), 2017, 42: 150–156]
- 21 Feng Y Y, Sun L, Zhang D K, et al. Study on Chinese named entity recognition using small scale character tail hints. Acta Electron Sin, 2008, 9: 1833–1838 [冯元勇, 孙乐, 张大鲲, 等. 基于小规模尾字特征的中文命名实体识别研究. 电子学报, 2008, 9: 1833–1838]
- 22 Zhang C J. Study on the analysis method of event space-time and attribute information in Chinese text. Dissertation for Ph.D. Degree. Nanjing: Nanjing Normal University, 2013 [张春菊. 中文文本中事件时空与属性信息解析方法研究. 博士学位论文. 南京: 南京师范大学, 2013]
- 23 Tang X R, Chen X H, Zhang X Y. Research on toponym resolution in Chinese text. Geomat Inform Sci Wuhan Univ, 2010, 35: 930–935 [唐旭日, 陈小荷, 张雪英. 中文文本的地名解析方法研究. 武汉大学学报 (信息科学版), 2010, 35: 930–935]
- 24 Zhang C J, Zhang X Y, Chen Y T, et al. Extraction of geographical attribute-values in natural language. In: Proceedings of Advances in Computational Environment Science. Melbourne: Springer, 2012. 51–59

- 25 Zhang X Y, Ye P, Wang S, et al. Geological entity recognition method based on deep belief networks. *Acta Petrol Sin.* 2018, 34: 343–351 [张雪英, 叶鹏, 王曙, 等. 基于深度信念网络的地质实体识别方法. *岩石学报*, 2018, 34: 343–351]
- 26 Wang S, Zhang X Y, Ye P, et al. Deep belief networks based toponym recognition for Chinese text. *Isprs Int J Geo-Inf*, 2018, 7: 217
- 27 Zhang X Y. Spatio-temporal information extraction in Chinese text. *Commun China Comput Feder*, 2015, 11: 33–39 [张雪英. 中文文本的时空信息获取方法. *中国计算机学会通讯*, 2015, 11: 33–39]
- 28 Zhang C J, Zhang X Y, Li M, et al. Interpretation of temporal information in Chinese text. *Geogr Geo-Inf Sci*, 2014, 30: 1–7 [张春菊, 张雪英, 李明, 等. 中文文本中时间信息解析方法. *地理与地理信息科学*, 2014, 30: 1–7]
- 29 Ma L L, Li H C, Wei Y, et al. Chinese text temporal expression recognition and normalization method based on rules. *J Inf Eng Univ*, 2017, 18: 560–565 [马雷雷, 李宏传, 魏勇, 等. 基于规则的中文文本时间表达式识别和规范化方法. *信息工程大学学报*, 2017, 18: 560–565]
- 30 Yan Z F, Ji D H. Exploraton of Chinese temporal information extraction based on CRF and semi-supervised learning. *Comput Eng Desig*, 2015, 36: 1642–1646 [闫紫飞, 姬东鸿. 基于 CRF 和半监督学习的中文时间信息抽取. *计算机工程与设计*, 2015, 36: 1642–1646]
- 31 Liu Z J, Tang B Z, Wang X L, et al. CMedTEX: a rule-based temporal expression extraction and normalization system for Chinese clinical notes. *Annu Symp Proc*, 2016, 2016: 818–826
- 32 Jin B W. Study on Chinese temporal expression recognition and normalization. Dissertation for Master's Degree. Shenyang: Shenyang University of Aeronautics and Astronautics, 2018 [金博文. 中文时间表达式识别和标准化研究. 硕士学位论文. 沈阳: 沈阳航空航天大学, 2018]
- 33 Hao T Y, Pan X Y, Gu Z Y, et al. A pattern learning-based method for temporal expression extraction and normalization from multi-lingual heterogeneous clinical texts. *BMC Med Inform Decis Mak*, 2018, 18: 22
- 34 Montello D R, Goodchild M F, Gottsegen J, et al. Where's downtown? Behavioral methods for determining referents of vague spatial queries. *Spat Cogn Comput*, 2003, 3: 185–204
- 35 Liu Y, Yuan Y H, Zhang Y. A cognitive approach to modeling vague geographical features: a case study of Zhongguancun. *J Remote Sens*, 2008, 2: 370–377 [刘瑜, 袁一泓, 张毅. 基于认知的模糊地理要素建模 —— 以中关村为例. *遥感学报*, 2008, 2: 370–377]
- 36 Yan F. Estimation of spatial range of fuzzy location based on interest points. Dissertation for Master's Degree. Shenyang: Shenyang University of Aeronautics and Astronautics, 2018 [颜芬. 基于兴趣点的模糊地点空间范围估计. 硕士学位论文. 沈阳: 沈阳航空航天大学, 2018]
- 37 Voegelé T, Schlieder C, Visser U. Intuitive modelling of place name regions for spatial information retrieval. In: *Spatial Information Theory Foundations of Geographic Information Science*. Berlin: Springer, 2003. 239–252
- 38 Zhang X Y, Zhu S N. Contextual spatial relations based spatial modelling of vague place names. In: *Proceedings of 2010 2nd IITA International Conference on Geoscience and Remote Sensing*. New York: IEEE Computer Society, 2010. 23–26
- 39 Liu S Y. Extracting landslide disaster information from web pages. Dissertation for Master's Degree. Chengdu: Southwest Jiaotong University, 2015 [刘思源. 面向 Web 文本的滑坡灾害信息提取技术研究. 硕士学位论文. 成都: 西南交通大学, 2015]
- 40 Imran M, Elbassuoni S, Castillo C, et al. Extracting information nuggets from disaster-related messages in social media. In: *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management*, 2013. 791–801
- 41 Yu L, Lu F, Zhang H C. Extracting geographic information from web texts: status and development. *J Geo-Inf Sci*, 2015, 17: 127–134 [余丽, 陆锋, 张恒才. 网络文本蕴涵地理信息抽取: 研究进展与展望. *地球信息科学学报*, 2015, 17: 127–134]
- 42 Matsuo Y, Sakaki T, Uchiyama K, et al. Graph-based word clustering using a web search engine. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, 2006. 542–550
- 43 Su F L, Xie Q H, Qiu J Y, et al. Study on word clustering for attribute extraction based on deep learning. *Micro Comput ITS Appl*, 2016, 35 : 53–55, 59 [苏丰龙, 谢庆华, 邱继远, 等. 基于深度学习的领域实体属性词聚类抽取研究. *微型机与应用*, 2016, 35: 53–55, 59]
- 44 Jiang H J. Slot filling via deep learning. Dissertation for Master's Degree. Zhejiang: Zhejiang University, 2017 [蒋焕剑. 基于深度学习的属性抽取技术研究. 硕士学位论文. 浙江: 浙江大学, 2017]

- 45 Allen J F. Towards a general theory of action and time. *Artif Intell*, 1984, 23: 123–154
- 46 Zhu S N, Zhang X Y, Zhang C J. Syntactic pattern recognition of geospatial relations described in natural language. In: *Proceedings of 2010 International Conference on Broadcast Technology and Multimedia Communication*, Hong Kong, 2010. 367–370 [朱少楠, 张雪英, 张春菊. 地理空间关系描述的句法模式识别. 见: *广播技术与多媒体通信国际会议*, 香港, 2010. 367–370]
- 47 Ma L B, Gong J Y. Research on spatial database query oriented natural language. *Comput Eng Appl*, 2003, 39: 16–19 [马林兵, 龚健雅. 面向自然语言的空间数据库查询研究. *计算机工程与应用*, 2003, 39: 16–19]
- 48 Zheng Y, Long Y, Ming X N, et al. Natural language description of geographic location considering various spatial relations with different references. *J Geo-Inf Sci*, 2011, 13: 465–471 [郑玥, 龙毅, 明小娜, 等. 多种空间关系组合的地理位置自然语言描述方法. *地球信息科学学报*, 2011, 13: 465–471]
- 49 Yu L, Lu F, Liu X L. A bootstrapping based approach for open geo-entity relation extraction. *Acta Geod Cartogr Sin*, 2016, 45: 616–622 [余丽, 陆锋, 刘希亮. 开放式地理实体关系抽取的 Bootstrapping 方法. *测绘学报*, 2016, 45: 616–622]
- 50 Zhang X Y, Zhu S N, Zhang C J. Annotation of geographical named entities of chinese text. *Acta Geod Cartogr Sin*, 2012, 41: 115–120 [张雪英, 朱少楠, 张春菊. 中文文本的地理命名实体标注. *测绘学报*, 2012, 41: 115–120]
- 51 Zhang X Y, Zhang C J, Zhu S N. Annotation of geographical named entities of chinese text. *Acta Geod Cartogr Sin*, 2012b, 41: 468–474 [张雪英, 张春菊, 朱少楠. 中文文本的地理空间关系标注. *测绘学报*, 2012b, 41: 468–474]
- 52 Zhang X Y, Zhang C J, Du C L, et al. SVM based extraction of spatial relations in text. In: *Proceedings of Spatial Data Mining and Geographical Knowledge Services*. Beijing: Institute of Electrical and Electronics Engineers, 2011. 529–533
- 53 Wang J B, Lu F, Wu S, et al. Constructing the corpus of geographical entity relations based on automatic annotation. *J Geo-inf Sci*, 2018, 20: 871–879 [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. *地球信息科学学报*, 2018, 20: 871–879]
- 54 Jin Y. Research and implementation of geographical relationship extraction system based on deep learning. Dissertation for Master's Degree. Beijing: Beijing University of Posts and Telecommunications, 2019 [靳滢. 基于深度学习的地理关系抽取系统研究与实现. 硕士学位论文. 北京: 北京邮电大学, 2019]
- 55 Gao J L, Yu L, Qiu P Y, et al. A knowledge-based method for filtering geo-entity relations. *J Geo-Inf Sci*, 2019, 21: 1392–1401 [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. *地球信息科学学报*, 2019, 21: 1392–1401]
- 56 Brodt A, Nicklas D, Mitschang B. Deep integration of spatial query processing into native RDF triple stores. In: *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 2010. 33–42
- 57 Liagouris J, Mamoulis N, Bouros P, et al. An effective encoding scheme for spatial RDF data. *Proc VLDB Endow*, 2014, 7: 1271–1282
- 58 Wang D, Zou L, Feng Y, et al. S-store: an engine for large RDF graph integrating spatial information. In: *Proceedings of the 18th International Conference on Database Systems for Advanced Applications (DASFAA 2013)*. Berlin: Springer, 2013. 31–47
- 59 Fu B J. Geography: from knowledge, science to decision making support. *Acta Geogr Sin*, 2017, 72: 1923–1932 [傅伯杰. 地理学: 从知识、科学到决策. *地理学报*, 2017, 72: 1923–1932]
- 60 Commission on Geographical Education, International Geography Union. International Charter on Geography Education. 1992. <http://www.igu-cge.org/1992-charter/>
- 61 Claramunt C, Parent C, Theriault M. Design patterns for spatio-temporal processes. In: *Proceedings of Technology of Object-Oriented Languages and Systems*, California, 1999. 2–15
- 62 Haarslev V, Lutz C, Moller R. A description logic with concrete domains and a role-forming predicate operator. *J Logic Computat*, 1999, 9: 351–384
- 63 Wang S. Research on construction of geographic knowledge graph driven by natural language. Dissertation for Ph.D. Degree. Nanjing: Nanjing Normal University, 2018 [王曙. 自然语言驱动的地理知识图谱构建方法研究. 博士学位论文. 南京: 南京师范大学, 2018]
- 64 Frank A U. Qualitative spatial reasoning: cardinal directions as an example. *Int J Geogr Inform Syst*, 1996, 10: 269–290

- 65 Kondaveeti A, Runger G C, Liu H, et al. Extracting geographic knowledge from sensor intervention data using spatial association rules. In: Proceedings of IEEE International Conference on Spatial Data Mining & Geographical Knowledge Services, Beijing, 2011. 127–130
- 66 Yu C Q, Peuquet D J. A GeoAgent-based framework for knowledge-oriented representation: embracing social rules in GIS. *Int J Geograph Inf Sci*, 2009, 23: 923–960
- 67 Otero-Cerdeira L, Rodríguez-Martínez F J, Gómez-Rodríguez A. Ontology matching: a literature review. *Expert Syst Appl*, 2015, 42: 949–971
- 68 Jiang B C, Wan G, Xu J, et al. Geographic knowledge graph building extracted multi-sourced heterogeneous data. *Acta Geod Cartogr Sin*, 2018, 47: 1051–1061 [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. *测绘学报*, 2018, 47: 1051–1061]

Spatiotemporal features based geographical knowledge graph construction

Xueying ZHANG^{1,2*}, Chunju ZHANG³, Mingguang WU^{1,2} & Guonian LV^{1,2}

1. *Key Laboratory of Resources and Environmental (Nanjing Normal University), Ministry of Education, Nanjing 210023, China;*

2. *Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China;*

3. *School of Civil and Hydraulic Engineering, Hefei University of Technology, Hefei 230009, China*

* Corresponding author. E-mail: zhangsnowy@163.com

Abstract Geographical knowledge is the cognitive result of human beings on the spatial distribution, evolution, and interaction of geographical phenomena or things. At present, there is a common phenomenon, i.e., “mass of data, information explosion, and hard to find knowledge” in geographic information services in the context of big data. The geographic knowledge graph (GeoKG) is a knowledge-based system that uses the semantic network to describe geographic concepts, entities, and their relations. It has application prospects in geographic knowledge understanding, geological problem solution, as well as spatiotemporal prediction and decision-making. In addition to the characteristics of general knowledge, geographic knowledge also has specific characteristics of spatiotemporal and geoscience mechanism. Therefore, the construction and applications of GeoKGs have both generality and certain professional specialty. Based on the spatiotemporal characteristics of geographic knowledge and the structure of knowledge graphs, we propose a method for GeoKG construction that considers the characteristics of time and space. Firstly, we describe the basic ideas and technical processes of GeoKG construction and briefly explain the main research contents of geographic knowledge acquisition, geographic knowledge abstraction and representation, as well as geographic knowledge organization and management. Secondly, based on the analysis of basic questions answered by geography, we summarize and abstract geographic knowledge elements and then explore a geographic knowledge representation model with three levels of “geographical concepts-geographic entities-geographical relations” to describe semantic units of geographical knowledge and their logic relations. Finally, the spatiotemporal features are assumed as the basic conditions, and based on the statuses of geographic entities, we propose a formal method based on “process-relation” to represent the evolution process of geographic entities and complex geographic relations. In conclusion, our results can effectively solve the basic problems of GeoKG construction, which are crucial for the acquisition, fusion, reasoning, and application of geographic knowledge. Furthermore, it can be extended in geoscience fields such as geology, environmental science, and meteorology, and has the prospects for the promotion of geoscience knowledge services.

Keywords geographical entity, spatio-temporal feature, geographical knowledge representation model, geographical knowledge formalization, geographical knowledge graph

Xueying ZHANG was born in 1970. She is a professor at the School of Geography of Nanjing Normal University, and her research interest include geographic big data mining, smart cities, and humanity and social geographic information systems.

Chunju ZHANG was born in 1984 and received her Ph.D. degree from Nanjing Normal University in 2013. She is an associate professor at Hefei University of Technology, and her research interests include intelligent processing and remote sensing and geographic knowledge discovery.

Mingguang WU was born in 1979. He is a professor at the School of Geography of Nanjing Normal University. His research interests include geographic information visualization and geographic information system.

Guonian LV was born in 1961. He is a professor at the School of Geography of Nanjing Normal University. His research interests include virtual geographical environment, geographical knowledge discovery, and intelligent geographical information system.