引文格式：徐坚,李国忠,杨恒玲,等. 基于领域本体的数字孪生流域知识图谱构建与应用[J]. 地理空间信息, 2024, 22(9): 37-41. doi:10. 3969/j. issn. 1672-4623.2024.09.008

基于领域本体的数字孪生流域知识图谱构建与应用

徐　坚[1,2]，李国忠[1,2*]，杨恒玲[3]，徐　健[1,2]，赵保成[1,2]，张双印[1,2]，付珺琳[1,2]，王丽华[1,2]
（1. 长江科学院空间信息技术应用研究所，湖北 武汉 430010；2. 武汉市智慧流域工程技术研究中心，湖北 武汉 430010；
3. 湖南省澧水流域水利水电综合开发公司，湖南 长沙 410000）

摘　要：为支撑数字孪生流域知识图谱的快速高效构建，以澧水流域为例，参照水利对象分类总则，通过梳理澧水流域内与江垭皂市工程相关的知识对象，选取关键属性，运用本体语义关系表达物理流域中的江河湖泊、水利工程和水利对象治理管理活动等实体、概念及其关系，构建了面向防洪兴利的数字孪生流域知识本体。基于水利领域本体模型和设计图数据模式进行结构化分类和关联，快速构建了数字孪生流域知识图谱；再利用 MySQL 和 Neo4j 图数据库进行知识混合存储，利用 Cypher 检索语句查询和应用对象关系，以支撑水利知识的快速检索和定位。

关键词：领域本体；数字孪生；知识图谱；本体语义；图数据库
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2024）09-0037-05

Construction and Application of Digital Twin Basin Knowledge Graph Based on Domain Ontology

XU Jian[1,2], LI Guozhong[1,2], YANG Hengling[3], XU Jian[1,2], ZHAO Baocheng[1,2], ZHANG Shuangyin[1,2],
FU Junlin[1,2], WANG Lihua[1,2]
(1. Spatial Information Technology Application Department, Changjiang River Scientific Research Institute, Wuhan 430010, China;
2. Wuhan Center for Intelligent Drainage Engineering Technology Research, Wuhan 430010, China;
3. Hunan Lishui Hydro & Power Co., Ltd., Changsha 410000, China)

Abstract: In order to support the rapid and efficient construction of digital twin basin knowledge graph, taking the Lishui basin for example, referring to the general classification rules of water conservancy objects, we sorted out the knowledge objects related to Jiangya and Zaoshi reservoir in the Lishui basin, and selected key attributes. Then, we used the ontology semantic relations to express the entities, concepts and relationships of rivers and lakes, water conservancy projects and water conservancy object governance and management activities in physical basins, and constructed digital twin basin knowledge ontology for flood control and benefit. We carried out structured classification and correlation based on the ontology model of water conservancy domain and the design diagram data model, thus realizing the rapid construction of digital twin basin knowledge graph. We used MySQL and Neo4j graph database to mixed store knowledge. Finally, we used Cypher retrieval statement to implement object relational query and application, which could support the rapid retrieval and positioning of water conservancy knowledge.

Key words: domain ontology, digital twin, knowledge graph, ontological semantics, graph database

近年来，知识图谱在互联网[1]、教育[2]、金融[3]、医疗[4]、自然资源[5]等行业都发挥了重要作用，尤其在水利领域，成为数字孪生建设的关键内核。水利知识作为数字孪生平台的智能驱动内核之一，其作用日益突出[6]。在数字孪生小浪底建设中，采用知识引擎、知识库和知识应用组成的知识库架构体系实现了基于水利知识的检索、问答、推荐和推理决策等多样化服务[7]。在数字孪生永定河建设中，基于大语言模型、知识图谱和向量数据库，实现了知识检索和智能问答，并开发水利对象关联查询、业务规则匹配、历史场景复演、实时调度方案编制等水利业务智能应用[8]。对于洪泽湖天然水湿地的提取，在光谱特征分析的基础上，构建了天然淡水湿地信息提取特征集，并利用决策树算法开展数据挖掘和生成知识规则[9]。

收稿日期：2024-05-07。
项目来源：武汉市知识创新专项基础研究资助项目（2022010801010238）；湖南省水利科技资助项目（XSKJ2023059-59）；武汉市重点研发计划资助项目（2023010402010586）；水利部重大科技资助项目（SKS-2022161）；中央级公益性科研院所基本科研业务费专项资助项目（CKSF2023313/KJ）；长江科学院开放基金资助项目（CKWV20231160/KY）。
第一作者简介：徐坚（1988—），硕士，高级工程师，研究方向为水利数字孪生、遥感水文，E-mail：xujianwhu@gmail.com。
通信作者：李国忠（1990—），硕士，工程师，研究方向为水利信息化、地理信息本体，E-mail： 373866923@qq.com。

水利知识提供了描述原理、规律、规则、经验、技能、方法等信息[10]，主要实现对水利对象关联关系和水利规律等知识的抽取、管理和组合应用，支撑正向智能推理和反向溯因分析[11]。此外，结合区块链[12]、场景推理[13]、预报模型[14]等方法，知识图谱在水利行业还有诸多应用。

知识图谱作为水利知识表达和展示的核心技术之一，是数字孪生流域知识建设的基础支撑，尤其是在水利对象关系表达方面。水利对象关联关系用于描述物理流域中的江河湖泊、水利工程和水利对象治理管理活动等实体、概念及其关系[15]，是其他水利知识融合的基础。利用本体理论，面向周边地缘环境主题，可基于主题本体的语义构建搜索引擎基本框架[16]。水利领域知识图谱的构建目前主要有自顶向下和自底向上两种方式，前者是针对特定的行业，由该行业专家定义好顶层本体与数据模式，再将抽取到的实体加入到知识库中，有利于抽取新的实例，保证抽取质量；后者主要依赖开放链接数据集和百科网站，在这些结构化知识中进行自动学习，直接将抽取数据中发现的实体、关系以及属性合并到知识图谱中[17]，能发现新的模式。

领域本体构建是一个复杂的过程，需要对所涉及的领域有深入的理解。构建本体可促进知识的共享和重用，是知识图谱构建和语义网技术的基础。领域知识构建的主要过程包括知识表示、知识存储等，尽管目前相关原理和应用都已取得了较好的成果，但仍在快速发展之中。近年来，领域知识图谱的发展趋势发生了一系列变化：在知识表示方面，通常采用三元组表达事实知识，但在决策、推理等相关应用中，需要依赖于大量专家知识、动态知识进行辅助判断，而专家知识的表示已超出了常规知识表示的范畴，在大数据的赋能下，知识表示的重心将逐步过渡到动态知识是必然趋势；在知识存储方面，现阶段一般利用RDF格式存储领域知识，但针对选择性低、复杂查询效率低的问题，知识存储的重心将逐步过渡到专业图数据库是必然趋势；在领域知识图谱应用方面，领域知识图谱通常用来辅助各种复杂的分析应用或决策支持，在互联网、电商等领域均存在广泛应用，但水利领域知识的专业性较强，主要在本行业内应用，知识构建基础相对薄弱，相关知识应用在水利领域仍处于探索试点阶段。水利知识图谱聚焦对象关联与规律抽取，支持正反向推理分析，与区块链等技术相结合拓展应用边界。领域本体构建是基础，通过自顶向下和自底向上两种方式，促进知识标准化与共享。当前知识图谱发展趋向动态知识表示与高效图数据库存储，水利领域的应用虽尚处初级，但在专业分析和决策支持上潜力巨大，本文将以澧水流域为例，积极探索其在数字孪生流域中的深度应用。

1 水利知识对象分类

1.1 研究对象

澧水流域隶属于长江流域，位于湖南省西北部，流经湘鄂两省，是湖南省4大河流之一；流域面积为1.53万km^2；地跨湖南省的永顺、桑植、慈利、张家界、石门县、临澧、澧县、津市和湖北省的鹤峰、五峰土家族自治州县10个县（市）；一级支流包括左岸的溇水、渫水以及右岸道水等，分别于慈利、三江口和道河口汇入干流，在溇水和渫水上分别建有江垭水库和皂市水库。本文以澧水流域为例，梳理水利知识对象。

1.2 知识分类

水利知识对象分类是对水利领域内的各种实体和概念进行系统化、标准化分类，以便于信息的组织、管理和检索，通常涉及水文、水资源、水利工程、水生态、水土保持、农村水利、水灾害防御、水利信息化等多个领域[18]。SL/T 213–2020《水利对象分类与编码总则》规范了水利对象的分类编码规则，便于水利信息的整合与共享，表明水利知识对象分类不仅是学术研究的内容，也是行业标准化的一部分。构建水利知识图谱时，这些分类可作为框架，将相关的数据、文献、规定、技术规范等信息组织起来，形成一个结构化的知识体系。这样的知识图谱有助于实现水利知识的跨域查询、检索和智能化决策支持。SL/T 213–2020《水利对象分类与编码总则》中的水利对象分类采用线分类法，将水利对象划分为大类、中类、小类和基类4级，大类为一级分类，按水利对象是否由人力干预形成划分为自然和非自然两类；中类为二级分类，自然大类划分为地表和地下两个中类；非自然大类划分为设施和非设施两个中类；小类为三级分类，地表和地下两个中类均按照水利对象的集输水特征，划分为集储水单元和输水通道两个小类，设施中类划分为独立工程和组合工程两个小类，非设施中类按管理行为的施受关系划分为行为主体和行为客体两个小类；基类是在小类的基础上划分的具体可实例化的四级对象类。大类、中类、小类为固定分类，基类为可扩展分类。

水利知识包括事实类知识和认知类知识，因此详细的掌握水利综合知识，对构建澧水流域水工程知识图谱具有关键作用。通过描述水利知识的组成单元和概念，体现不同类型水利知识的关联特征，以构建完善的水利综合知识图谱[19]，再通过前期资料分析和水

利对象分类梳理，得出澧水流域水利知识类别（图1），事实类水利知识一般分为自然、工程和社会对象，分别指在自然界天然存在的流域、湖泊等，人工修建的水库、灌溉渠、水电站、泵房等人为安全措施以及水利管理人员对水文、水域地理位置、职责等类的管理，其特点是具备抽象性、认同性和普适性，也是目前水利学科知识和预防措施的关键点；认知类知识主要是指在通过科学理论、原理、方法、数学模型、统计特征等对水利知识进行相关研究。

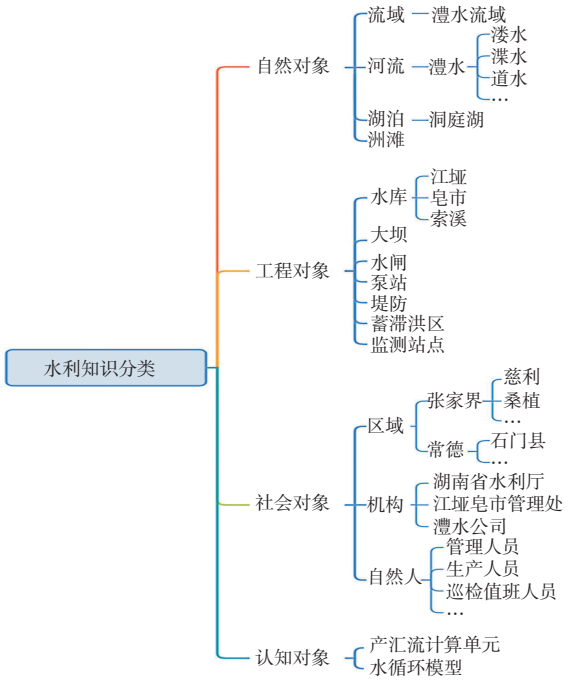


图1　水利对象分类

2　水利领域本体构建

本体是一个描述性的形式化概念体系，是知识图谱的模型层，包括点类型（实体分类或概念）、点类型属性（点类型特征）以及边类型、边类型连接点类型，边类型也包含边类型属性。领域本体是指领域概念体系（领域概念定义和概念间的关系）的形式化规范定义。本文关注水利领域知识，属于特定领域内概念与概念的关系，因此基于领域本体建立模型，不同领域的构建方法也不同。常用的方法主要有骨架法、七步法、IDEF5法等，具体流程见图2。

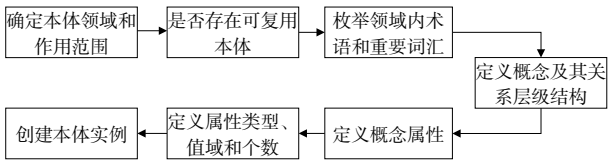


图2　本体构建流程图

目前关于领域本体构建没有统一的标准，本文采用七步法结合IDEF5法中本体进行验证与优化。国内外现有的本体建模工具以Protege为代表，领域知识图谱的构建大多采用自顶向下和自底向上相结合的方式，利用本体编辑器构建水利知识本体。水利领域本体的构建过程为：①定义点类型，实体分类按照水利对象的大类、中类、小类等进行细分；②定义点类型的分类关系（子类—父类的分类树），本体中的点类型可以有最高抽象点类型（自然对象、工程对象等），也可以有子点类型（工程对象如水库、堤防、水闸等），甚至可以更细分（水库如大I型水库、大II型水库、中型水库等）；③定义边类型、边类型的有向性，边类型即对象关系类型，如河流的上下游关系、干支流关系、左右岸关系等，也可以是社会对象的管理隶属关系和知识实体的空间位置关系，边类型的有向性包括出向和入向，表征关系作用的方向；④定义点、边类型属性以及属性值取值范围；⑤设定点、边类型属性取值的缺省值。

Protege为开源本体构建工具，具备易操作、储存格式多样、可编辑、可扩展、支持逻辑推理等特点，为国内外学者的首选。本文结合上述水利对象分类方式，利用Protege工具构建本体模型（图3）。将Protege构建的本体模型导出为OWL文件，OWL是W3C开发的一种网络本体语言，用于对本体进行语义描述；通过提供更多具有形式语义的词汇，使其在Web内容的机器可理解性方面优于XML、RDF和RDF Schema。OWL可作为RDFS的一个扩展，添加额外的预定义词汇，具有快速灵活的数据建模能力和高效的自动推理。

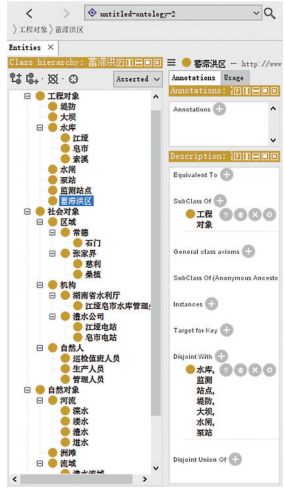


图3　水利对象分级层次

3　水利知识图谱构建与应用

本体和知识图谱均为知识表示的重要工具，本体侧重于概念的形式化描述和知识结构定义，知识图谱则侧重于知识的存储、查询和应用。以本体为模型

层，建立数据与本体的一一映射关系，再将映射结果导入到图数据库中，形成知识图谱，因此本体其实是知识图谱的构建范式，是数据的组织模板。本体转换为知识图谱是一个将本体论结构化的知识表示转换为图数据库的过程，其中本体中的类、属性、实例和关系被映射为图数据库中的节点和边。知识图谱是一种结构化的语义知识库，用于存储实体（节点）及其相互关系（边），便于进行复杂查询和推理。将本体转换为知识图谱的主要步骤包括：①概念映射，将本体中的类转换为知识图谱中的实体，每个类代表一个特定的概念或实体类型；②属性映射，将本体中的数据属性和对象属性转换为知识图谱中的属性或关系，数据属性通常表示实体属性，对象属性表示实体之间的关系；③实例转换，将本体中的实例转换为知识图谱中的实例节点，每个实例代表一个具体的实体；④关系构建，基于本体中定义的关系，构建知识图谱中的边，这些边连接不同的节点，表示实体之间的各种语义关系；⑤属性值填充，将实例的数据属性值填充到知识图谱中相应的节点属性；⑥推理和消歧，利用推理机制补全知识图谱中可能缺失的信息，并对实体进行消歧，确保每个节点代表唯一的实体；⑦图数据库存储，将转换后的结构存储到图数据库中，如Neo4j、OrientDB等；⑧查询和更新，通过图查询语言（如Cypher for Neo4j）对知识图谱进行查询和更新。

本体转换为知识图谱后，可利用图数据库的优势进行高效的数据存储和查询，同时知识图谱的结构化表示也便于进行复杂的数据分析和机器学习。本文选用Neo4j数据库作为图关系存储，MySQL数据作为实体属性存储，采用混合存储模型管理知识数据。Neo4j是一个高性能的图形数据库管理系统，允许存储和管理复杂的关系数据。Neo4j以其高效的图形数据模型、灵活的数据模式和强大的查询语言Cypher而闻名。Cypher是专为图形数据库设计的声明式查询语言，由Neo4j图形数据库引入，允许用户以直观和灵活的方式对图形数据进行查询和操作。Cypher语言的设计目标是简洁、易读、易写，同时提供高性能的查询能力。例如，要查找所有名为“Alice”的节点，并返回它们的名称，Cypher查询语句为：MATCH(person:Person{name:Alice}) RETURN person.name。基于Cypher语言建立澧水流域的知识图谱，并进行灵活的查询和推理，输入查询所有的语句为：MATCH(n) RETURN n，返回的Neo4j中存储的所有节点和关系见图4。基于该知识图谱查询水文站上下游关系，输入查询语句MATCH(n)-[r:下游]->(p) RETURN n,p，结果见图5a。知识推理“位于”关系的查询语句为：MATCH(n)-[r:位于]->(p) RETURN n,p，结果见图5b。空间关系“流经”的关系查询语句为：MATCH(n)-[r:流经]->(p) RETURN n,p，结果见图5c。

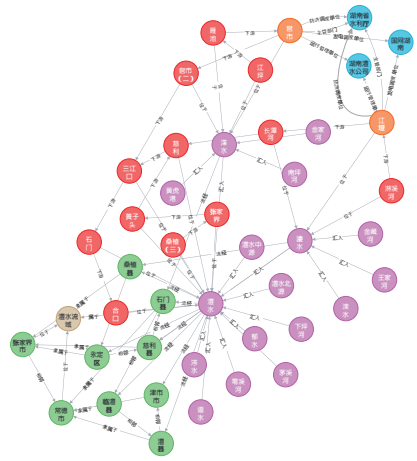
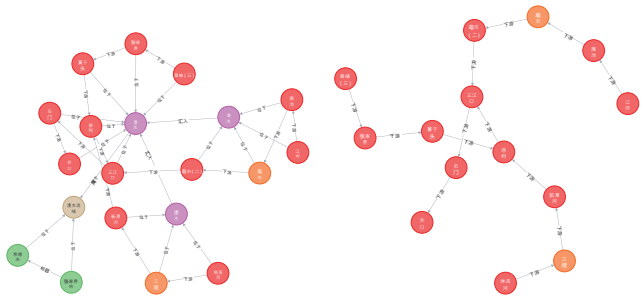
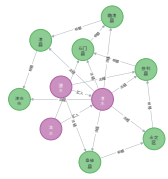


图4 澧水流域知识图谱



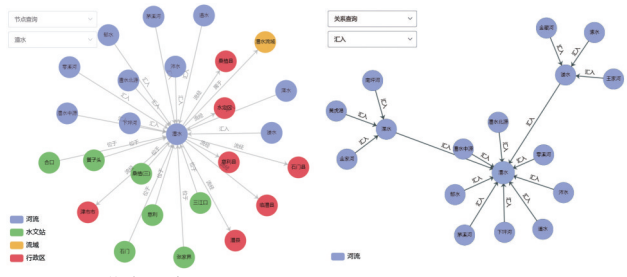
a 水文站上下游关系　　b “位于”关系



c “流经”关系

图5 空间关系查询

除了可使用专业的图数据库语言管理知识图谱外，本文基于Springboot+Vue开发框架实现了知识图谱的检索、查询和动态更新，便于知识的可视化管理（图6）。



a 节点查询界面　　b 关系查询界面

图6 流域知识可视化系统原型界面

4 结 语

本文通过梳理澧水流域范围内各类水利知识对象，基于领域本体建模方法，运用图关系数据库技术建立了江垭皂市工程相关对象的知识图谱，将水利专业知识和空间位置关系进行逻辑化、数字化表达，可支持按照实体类别、对象关系查询，如查询河流水系对象、河流水位站依附关系、上下游关系等，并在前端界面实现了可视化、查询、简单推理等功能。随着数字孪生流域技术的高速发展，未来知识图谱将结合新一代AI大语言模型，进一步探究大型语言模型在知识图谱构建和推理方面的应用，助力数字孪生流域建设的智能化水平再上新的台阶。

参考文献

[1] 魏明珠,郑荣,高志豪,等. 融合知识图谱和深度神经网络的产业新兴技术预测模型研究[J]. 情报学报,2022,41(11):1 134-1 148

[2] 杨玉基,许斌,胡家威,等. 一种准确而高效的领域知识图谱构建方法[J]. 软件学报,2018,29(10):2 931-2 947

[3] 林夏莹,连海峰,黄思洁. 基于知识图谱的银行客户数据管理研究[J]. 金融科技时代,2023,31(8):10-15

[4] 宋源,于彤,朱玲,等. 功能性胃肠病中医药知识图谱构建方法研究[J]. 中国数字医学,2022,17(10):48-53

[5] 陈虎,李宏伟,马雷雷. 本体在地理知识库构建中的应用[J]. 地理空间信息,2011,9(5):78-80

[6] 蔡阳. 智慧水利建设现状分析与发展思考[J]. 水利信息化,2018(4):1-6

[7] 张军珲,霍建伟,崔培,等. 数字孪生小浪底知识库建设研究与实践[J]. 水利信息化,2024(1):13-18

[8] 李巍,李树元. 基于大语言模型和知识图谱的数字孪生流域知识平台设计[J]. 海河水利,2023(12):75-78

[9] 高雷亭,阮仁宗,颜梅春,等. 基于知识规则的天然淡水湿地信息提取[J]. 地理空间信息,2014,12(3):33-35

[10] 蒋乐龙,吉斯琴图雅. 数字孪生技术在水利工程总承包项目管理应用探索[J]. 水利水电工程设计,2022,41(4):30-34

[11] 刘昌军,刘业森,武甲庆,等. 面向防洪“四预”的数字孪生流域知识平台建设探索[J]. 中国防汛抗旱,2023,33(3):34-41

[12] 徐健,徐坚,魏思奇,等. 区块链技术在智慧水利体系建设中的应用展望[J]. 长江科学院院报,2023,40(4):150-154

[13] 高英. 基于知识库和场景推理的乌江智能化防洪调度[EB/OL]. (2024-03-21)[2024-04-15]. https://kns. cnki. net/ kcms2/article/abstract?v=8dkf_uZKVx0vfSAJmQuqpj-FyzqT5Yh5c7hqbYXRN7NotxufETJj5xW4eUMnQtestHW_3PhBnqg23q4aaAE3-VlWsqr5iosuCC-W-fZmCSjpKOoprL 6Y 2Y5O-4qNohkhqYz-wVVAU_1k=&uniplatform=NZKPT&language=CHS

[14] 陈华,徐坚,盛晟,等. 面向流域防洪安全的态势图谱构建及可视化方法[J]. 河海大学学报(自然科学版),2020,48(6):498-505

[15] 冶运涛,蒋云钟,曹引,等. 以数字孪生水利为核心的智慧水利标准体系研究[J]. 华北水利水电大学学报(自然科学版),2023,44(4):1-16

[16] 马雷雷,梁汝鹏,连世伟,等. 一种主题本体驱动的语义搜索方法[J]. 地理空间信息,2013,11(4):46-48

[17] 冯钧,杭婷婷,陈菊,等. 领域知识图谱研究进展及其在水利领域的应用[J]. 河海大学学报(自然科学版),2021,49(1):26-34

[18] 疏正宏,季永兴,吴兆丹,等. 长三角示范区水利建设标准协同路径研究及实践[J]. 人民长江,2023,54(6):132-139

[19] 王亮. 知识图谱技术在北江流域水利工程联合调度中的应用[J]. 黑龙江水利科技,2021,49(12):187-190

（上接第28页）

[27] 陈忠媛,高永年. 长江水系湖泊型流域生态用地长时间序列时空变化分析[J]. 中国科学院大学学报,2022,39(2):172-184

[28] 黄安东,赵明松,郜敏,等. 1980—2020年安徽省土地利用时空演化特征[J]. 科学技术与工程,2022,22(11):4 627-4 653

[29] 左岍,周勇,李晴,等. 鄂西南地区土地利用格局时空变化及轨迹特征分析[J]. 水土保持学报,2022,36(1):161-169

[30] 龙花楼,王文杰,翟刚,等. 安徽省土地利用变化及其驱动力分析[J]. 长江流域资源与环境,2002,11(6):526-530

[31] 蒋婷婷,郭伟玲,管礼松,等. 1980—2018年安徽省土地利用时空变化分析[J]. 测绘,2021,44(4):166-171

[32] 林跃胜. 皖南茶园土壤重金属形态分布特征及其生物有效性研究[D]. 芜湖:安徽师范大学,2013

[33] 徐新良,刘纪元,张树文,等. 中国多时期土地利用土地覆被遥感监测数据集(CNLUCC)[Z]. 中国科学院资源环境科学数据中心数据注册与出版系统(http://www. resdc. cn/DOI),2018

[34] 闫淑君,洪伟. 闽江口琅岐岛土地利用的时空动态[J]. 应用生态学报,2009,20(5):1 243-1 247

[35] 朱会义,李秀彬. 关于区域土地利用变化指数模型方法的讨论[J]. 地理学报,2003,58(5):643-650

[36] 全斌. 土地利用与土地覆被变化学导论[M]. 北京:中国环境科学出版社,2010