

中国行政区划沿革知识图谱构建方法研究

刘纪猛¹, 王曙^{2,3*}, 代小亮^{2,3}, 王春玲^{2,3}, 葛双双⁴, 韩保民¹, 诸云强^{2,5*}

1. 山东理工大学建筑工程与空间信息学院, 淄博 255000;

2. 中国科学院地理科学与资源研究所地理信息科学与技术重点实验室, 北京 100101;

3. 中国科学院大学, 北京 100049;

4. 黑龙江工程学院测绘工程学院, 哈尔滨 150000;

5. 海南大学生态学院、空天智能学院, 海口 570228

摘要: 行政区划沿革是研究国家治理、区域发展和空间格局演变的重要基础。然而, 传统行政区划沿革相关数据集, 如政府公告、地方志、统计年鉴及地理空间数据, 通常以碎片化形式存储, 缺乏动态关联与语义整合能力, 难以支撑复杂的时空查询与演变分析。针对行政区划数据集来源多样、演变类型复杂、时空特征耦合紧密等特点, 本文提出了一套中国行政区划沿革知识图谱的构建方法。在知识建模层面, 通过构建行政区划沿革知识本体, 规范行政区划实体、属性及其时空演变关系的语义表达; 在知识抽取层面, 针对多源异构数据设计了差异化的知识抽取方法: 对结构化数据采用规则映射与 GIS 拓扑计算方法, 提取空间关系与演变信息。对非结构化公告文本, 结合领域本体与大语言模型进行联合抽取, 实现行政区划节点、属性及沿革关系的结构化表达; 在知识融合层面, 综合利用行政区划实体在时间维度上的重叠特征与语义相似性, 开展实体对齐与融合; 在应用层面, 将融合后的知识数据存储于 Neo4j 图数据库中, 构建中国行政区划沿革知识图谱。

关键词: 行政区划; 知识图谱; 区划沿革; 演变分析; 中国

DOI: <https://doi.org/10.3974/geodp.2026.01.08>

CSTR: <https://cstr.escience.org.cn/CSTR:20146.14.2026.01.08>

1 前言

行政区划沿革是国家或地区在不同历史时期, 根据政治、经济、民族状况以及历史地理条件等要素的变化, 对行政区域划分进行长时序动态调整的过程^[1,2]。通过行政层级调整、区域重组、驻地迁移等实施方法, 优化区域空间结构^[3,4], 提升国家治理水平^[5,6], 促进社会经济提升^[7], 进而推动区域和国家的高质量可持续发展^[8,9]。行政区划沿革的数据及其蕴含的规律, 对统筹和制定区域发展战略和国家现代化治理具有重要参考价值。

目前, 中国行政区划沿革数据, 主要以 4 种形式进行记录和使用: 第一, 公告文档, 以公开的文件及报告形式记录行政区划沿革信息, 如民政部逐年发布的行政区划调整公

收稿日期: 2025-11-07; 修订日期: 2026-01-20; 出版日期: 2026-02-25

基金项目: 国家自然科学基金(42471503); 中国科学院(XDB0740200-01)

*通讯作者: 诸云强, 海南大学生态学院, 997455@hainanu.edu.cn; 王曙, 中国科学院地理科学与资源研究所, wangshu@igsnrr.ac.cn

引用方式: 刘纪猛, 王曙, 代小亮等. 中国行政区划沿革知识图谱构建方法研究[J]. 全球变化数据学报, 2026, 10(1): 54-68. <https://doi.org/10.3974/geodp.2026.01.08>. <https://cstr.escience.org.cn/CSTR:20146.14.2026.01.08>.

告^[10]、地方年鉴及志书^[11]等；第二，地图图像，以地图或图像为载体，采用年度信息汇编形式记录行政区划边界及信息，如《中华人民共和国行政区划图集》^[12,13]、《中华人民共和国行政区划简册》^[14]、《中华人民共和国行政区划沿革地图集》^[15]和《中华人民共和国行政区划变迁地图集（1980–2017）》^[16]等；第三，地理空间数据，以半结构化的矢量数据，按年度逐层精确记录行政区划边界及各类属性信息，如 GlobalAdministrativeAreas(GADM)全球多级行政区划数据集^[17]、OpenStreetMap(OSM)数据^[18]、天地图行政区划底图服务^[19]等；第四，研究论文，不同研究目标驱动下的行政区划变迁记录、数据或系统，如事件视角下扬州行政区划沿革图谱^[20]、中国行政区划格局演变分析^[21]、行政区划沿革知识图谱^[22]等。

然而，现有行政区划沿革知识图谱构建方法在处理海量、多源、异构数据时仍存在一定局限。当前地理知识模型与行政区划知识模型对行政区划演变过程的表达能力有限，未能深入刻画行政区划相关的时空属性和关系，难以全面呈现行政区划变迁的空间复杂性与动态性。此外，针对上述多源异构数据的抽取、对齐与融合，目前仍缺乏成体系的方法，难以精准地挖掘和整合行政区划信息。

鉴于此，本文提出了一种面向中国行政区划沿革知识图谱的体系化构建方法，旨在解决现有数据难以支撑深层次、长时序时空演化分析的问题。本研究首先通过构建行政区划本体，建立行政区划演变的标准化表达框架；随后，针对多源异构数据的特征，提出一套集成大语言模型（Large Language Model, LLM）、规则映射与 GIS 拓扑计算等差异化知识抽取方法，实现从公告文档、地理空间数据等数据集到动态演化逻辑的知识转化。在此基础上，通过时空拓扑约束与语义对齐技术，完成多源实体的融合，并依托 Neo4j 图数据库，构建起边界精准、语义关联、沿革连续的中国行政区划沿革知识图谱。

2 中国行政区划沿革知识图谱构建方法

2.1 行政区划沿革知识图谱构建总体框架

针对中国行政区划沿革数据多源异构、时空演化复杂的问题，本文提出一套面向行政区划沿革知识图谱的总体构建框架（图 1）。本框架以行政区划沿革本体模型为核心语义约束核心，融合基于多源数据的知识抽取与实体对齐方法，实现行政区划沿革知识的统一。

在本体模型设计阶段，综合分析现有本体、政府公告、科技文献等相关数据，对行政区划沿革的过程中可能涉及的概念、属性与关系进行系统梳理，进而构建涵盖行政区划层级、属性体系及演化关系的本体模型。

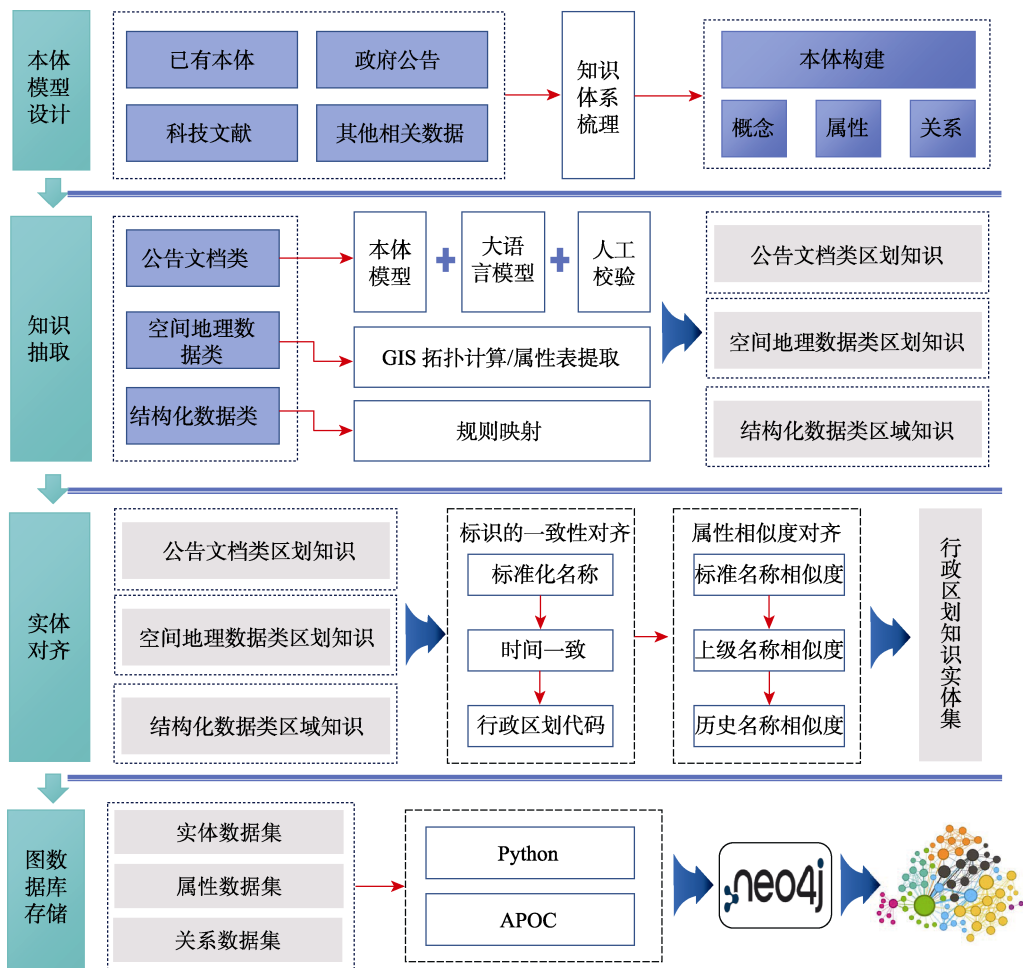


图 1 行政区划沿革知识图谱构建总体方法框架图

在知识抽取阶段，根据数据源的结构不同，采用不同的方法抽取行政区划实体：对公告文档等非结构化文本，结合本体约束与大语言模型进行语义抽取，辅以人工校验提高数据精度；对地理空间数据，利用 GIS 拓扑计算与属性表提取空间关系与属性信息；对结构化数据，通过规则映射得到本体约束下结构化知识。

在实体对齐阶段，由于数据来源不同，可能会存在相同属性描述存在差异的情况。比如“同名异地”，北京市和吉林省长春市都存在名称为朝阳区的下级行政区划，因此需要将不同数据源提取的实体进行对齐。本研究采用行政区划标识的一致性以及属性信息相似度相结合的方法进行实体对齐。

最终，使用 Python 代码和 APOC (Awesome Procedures on Cypher) 库将抽取与对齐后的知识按照本体模型设计的概念、属性、关系存储结构存入 Neo4j 图数据库实现可视化。

2.2 行政区划沿革本体建模方法

采用自顶向下的本体建模方法，从领域的最顶层概念出发，定义该领域内最核心、

最通用的分类和逻辑架构。具体来说，首先根据领域的现有本体、政府公告、科技文献等资料，确定本体的“骨架”，将行政区划实体作为本体的核心；其次，在顶层概念的基础上，逐步向下细分出子类 and 下属层级，将行政区划实体进一步细分，划分为“省级”“地级”“县级”行政区划单元；最后，在顶层架构搭建完成后，为每一层级定义具体的属性和相互关系。

2.3 知识抽取方法

在知识抽取实例阶段，针对不同数据源格式可能存在不一致的问题，比如地理空间数据多以矢量格式存储，公告文档以非结构化文本形式存在等，采取不同的知识抽取方法，包括：本体语义驱动的大语言模型（LLM）知识抽取方法、基于规则映射与字段解析的属性提取方法、基于空间拓扑算子的行政区划关系识别方法以及基于属性字段变化的行政区划演变关系判定方法。

针对公告文档等非结构化文本数据，采用本体语义驱动的大语言模型知识抽取方法，通过将本体模型中的“实体-属性-关系”结构转化为 LLM 识别逻辑，让大语言模型根据本体内容进行定向解析与匹配。在具体抽取过程中，使用提示工程技术，通过多轮对话与指令构建结构化抽取模板（Prompt），要求 LLM 识别“主语（S）、谓语关系（P）、宾语（O）及时间（T）”要素，并输出为规范的四元组。抽取结果再结合本体约束进行人工校验，剔除不符合本体语义规范的内容。

针对长时序、多源的 GIS 矢量数据，需规范化提取行政区划实体的属性信息与空间几何信息。具体采用基于规则映射与字段解析的属性提取方法：首先建立本体属性与 GIS 字段的映射规则库，对属性表字段进行标准化转换；随后利用开源地理信息工具 GeoPandas¹读取矢量数据，将属性字段与几何对象联合处理，计算并提取行政区划的空间范围；同时通过文件名解析获取对应年份信息，统一赋值至实体的时空属性中，最终形成行政区划实体属性数据集。

为识别行政区划间的空间关系及演变关系，采用基于空间拓扑算子的行政区划关系识别方法。该方法的核心原理是将行政区划矢量边界抽象为几何对象，通过拓扑关系判定与空间叠置计算，将行政区划间的相邻、包含等静态关系，以及合并、拆分、设立、撤销等动态演变关系识别出来。在同一年份内，采用 DE-9IM^[23]矩阵描述两个行政区划几何对象 A 与 B 在内部、边界和外部之间的 9 种交集关系，其拓扑关系可表示为公式 1：

$$M(A, B) = \begin{bmatrix} \dim(I(A) \cap I(B)) & \dim(I(A) \cap B(B)) & \dim(I(A) \cap E(B)) \\ \dim(B(A) \cap I(B)) & \dim(B(A) \cap B(B)) & \dim(B(A) \cap E(B)) \\ \dim(E(A) \cap I(B)) & \dim(E(A) \cap B(B)) & \dim(E(A) \cap E(B)) \end{bmatrix} \quad (1)$$

通过对矩阵中各元素取值模式的组合判断，可以精确刻画两个空间对象之间的拓扑关

¹ https://gitcode.com/gh_mirrors/ge/geopandas/.

系类型。比如, 当 $\dim(I(A) \cap I(B)) = \emptyset$ 且矩阵中至少有一个边界交集不为 \emptyset 时, 判定两个行政区划为相邻关系, 当 $A \cap B = A$ 且 $I(A) \cap I(B) = \emptyset$ 时, 判定两个行政区划为包含关系。

在演变关系抽取时, 基于 GIS 叠置分析方法^[24], 对相邻时间节点 t 与 $t+1$ 的行政区划空间几何数据集 $\{G_i^t\}$ 与 $\{G_j^{t+1}\}$ 进行空间交集计算, 判断行政区划是否在时间上延续或消失^[25]。针对存在空间交集的情形, 引入面积覆盖率模型^[26]进一步确定演变类型。行政区划的空间覆盖率计算公式如下:

$$C(G_i^t, G_j^{t+1}) = \frac{\text{Area}(G_i^t \cap G_j^{t+1})}{\text{Area}(G_i^t)} \times 100\% \quad (2)$$

首先设定覆盖率阈值, 当覆盖率接近 100%, 表示行政区划主体基本保持不变; 当多个 G_i^t 对同一个 G_j^{t+1} 的覆盖率超过设定的阈值, 判定为发生了合并; 当单个 G_i^t 对多个 G_j^{t+1} 的覆盖率超过设定的阈值时, 判定该行政区划发生了拆分。

行政区划演变关系除了表示空间范围变化的合并、拆分关系外, 还存在更名、行政级别调整、隶属调整等关系, 使用基于属性字段变化的行政区划演变关系判定方法来识别这类关系。首先对相邻年份已通过空间叠置与覆盖率判定为高度重合的行政区划进行筛选, 作为属性比对的候选集; 随后对候选集的中文名称、行政区划代码、行政级别、隶属行政区划、政府驻地等关键属性字段进行逐一比对。例如, 当空间范围保持恒定但中文名称字段发生变化时, 判定为“名称变更”关系; 当行政区划代码保持一致而隶属关系字段发生变化时, 判定为“隶属关系调整”等。

2.4 实体对齐方法

实体对齐是指在多源数据环境下, 识别并判定来自不同数据源, 但指向同一现实对象的实体实例, 并将这些实例进行异构消歧, 实现统一的步骤。在本研究中, 数据来自于公告文本、GIS 矢量数据及其他结构化或半结构化数据, 由于不同数据源命名规范、属性完整性等方面存在差异, 同一行政区划在不同数据源中可能存在表述不一致的情况。因此, 针对实体标识完备程度的不同, 设计了标识一致性对齐与属性相似度对齐两类方法。

针对具备标准标识或规范名称的实体实例, 本研究采用标识一致性对齐方法^[27], 利用行政区划实体中标准中文全称和唯一的行政区划代码, 在统一的时间基准 T 下进行精准锚定实体, 首先确定两个数据源中的实体在同一时间断面上, 如果实体具备相同的 6 位行政区划代码, 则判定为同一实体, 若区划代码缺失, 则对比标准中文全称及其隶属行政区划。公式表达如下:

$$S_{id}(e_1, e_2) = \begin{cases} 1, & \text{if Code}(e_1) = \text{Code}(e_2) \wedge T_1 = T_2 \\ 1, & \text{if } [\text{Name}(e_1) = \text{Name}(e_2) \wedge \text{Parent}(e_1) = \text{Parent}(e_2)] \wedge T_1 = T_2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

式中, (e_1, e_2) 为来自不同数据源 S_1 和 S_2 的实体; $\text{Code}(e)$ 为行政区划代码; $\text{Name}(e)$ 为行政区划中文名称全称; $\text{Parent}(e)$ 为行政区划上级行政区划中文名称全称; T 表示数据所属年份断面; S_{id} 表示一致性得分, 当 $S_{id}=1$ 认为是同一实体。

针对行政区划代码残缺且中文名称为非标准名称的情况, 本研究采用属性相似度对齐方法, 通过计算实体的多个属性之间的相似度来确定是否为同一实体^[27]。设行政区划实体 e_i 与 e_j 的属性集合为 $P = \{p_1, p_2, \dots, p_n\}$, 其综合属性相似度计算方法如下:

$$S(e_i, e_j) = \sum_{k=1}^n w_k \cdot \text{sim}_k(p_k(e_i), p_k(e_j)) \quad (4)$$

式中, w_k 为属性 p_k 的权重, 且 $\sum w_k = 1$; $\text{sim}_k()$ 为对应属性的相似度函数。在对行政区划实体进行对齐时, 首先设定决策阈值 θ , 选取中文名称与隶属行政区划属性作为核心评价指标。赋予中文名称较高的权重, 赋予隶属行政区划辅助判定的权重。通过上述公式加权求和得出综合相似度得分, 将其与预设的决策阈值 θ 进行比对, 若 $\text{Sim}(e_i, e_j) \geq \theta$, 则判定两实体表示同一行政区划。

3 1949–2023 年行政区划沿革知识图谱构建与验证

3.1 数据源

行政区划沿革知识图谱构建的数据资料主要来源于不同年份的中国行政区划矢量数据、中国省级统计年鉴以及县级以上行政区划变更情况公告。具体来源及说明参见表 1。

本文所使用的中国行政区划矢量数据数据由两部分组成, 其中 1949–2014 年的矢量数据来自于伯克利大学图书馆地理信息数据库, 2014–2023 年的矢量数据引用自 Cheng Rui 等 (2025) 发布的 China Temporal Sequence Administrative Map (CTAmap) 多时序行政区划数据库^[28]。CTAmap 以 2020 年中国行政区划格局作为基准, 融合了国家基础地理信息数据中心 (National Geomatics Center of China, NGCC) 基础地理信息数据与民政部行政区划代码等权威资料, 通过时间正向推演与反向回溯相结合的方法, 重建了 2009–2023 年间省级、市级与县级行政区划的空间边界、行政代码及层级隶属等属性。本研究开始之前对该数据库数据的空间边界与标准行政区划地图进行了对照检查, 并将关键属性信息与民政部历年公开发布的行政区划官方资料进行了核实验证, 发现该数据在空间形态、属性信息等方面均与权威资料高度一致, 能够作为中国行政区划沿革知识图谱数据集构建的数据基础。

表 1 中国行政区划沿革知识图谱数据来源情况

条目	中国行政区划矢量数据		中国省级统计年鉴	县级以上行政区划变更情况
数据年代	1949–2014	2014–2023	1981–2023	1978–2023
时间分辨率	年	年	年	年
空间分辨率	省、市、县	省、市、县	省	省、市、县
数据记录信息	行政区划基本信息	行政区划基本信息	经济数据、社会数据、资源数据等	批文名称、变更时间、县级以上行政区划变更情况
数据格式	Shpfile	Shpfile	PDF	.txt
数据量	1.08 GB	0.87 GB	472 MB	610 KB
数据来源	伯克利大学图书馆地理信息数据库 ²	以国家基础地理信息数据中心的县区划数据作为矢量基础 ³ ，辅以高德行政区划数据 ⁴ 、天地图行政区划数据 ⁵ ，参考历年来民政部公布的行政区划 ⁶ 为属性基础绘制 ^{128]}	国家统计局 ⁷	中华人民共和国民政部 ⁸ 、行政区划网 ⁹

注：中国行政区划矢量数据与中国省级统计年鉴数据范围为中国大陆省、市、县级行政区划以及香港、澳门，暂不包含台湾省。

3.2 中国行政区划沿革本体模型

本研究构建的行政区划沿革本体模型包含两大类核心概念及其属性和 4 大类关系。顶层核心概念包括行政区划（或行政区划单元）和行政中心。行政区划是指国家为行政管理而划分的区域，是地方行政机关所管辖的区域，包括：国家级行政区划、省级行政区划、地级行政区划以及县级行政区划四级；行政中心是指一个国家的中央政府或地方政府所在地，包含首都及地方政权驻地。行政区的属性主要包括用于描述和标识行政区的标识信息，规范管理层级的管理信息，限定地理范围的空间信息以及记录生命周期的时间信息，具体来看行政区划实体包含核心属性与拓展属性两大类属性。核心属性与行政区划演变具有强相关的关系，当行政区划发生变更时，必然涉及至少一种核心属性的变化；拓展属性是对行政区划的进一步描述，与行政区划变更没有直接的关联，但往往会因为行政区划的演变而发生相应的变化，例如，行政区划空间范围的调整，对应的 GDP 也会发生变化。行政中心的属性主要包括：明确行政中心身份信息的标识属性，精准定位行政中心位置和范围的空间属性，记录行政中心存续的时间属性以及明确行政中心权力与架构的管理属性，具体概念及属性分类参见表 2。

² <https://geodata.lib.berkeley.edu/>.

³ <https://www.webmap.cn/commres.do?method=dataDownload>.

⁴ https://datav.aliyun.com/portal/school/atlas/area_selector.

⁵ <https://www.mca.gov.cn/n156/n186/index.html>.

⁶ <https://cloudcenter.tianditu.gov.cn/administrativeDivision>.

⁷ <https://www.stats.gov.cn/sj/ndsjs/>.

⁸ <http://xzqh.mca.gov.cn/description?dcpid=1>.

⁹ <https://www.xzqh.org/show/china/>.

表 2 本体概念及属性分类表

序号	顶层概念	属性级别	属性类组	属性名称	
1	行政区划 (行政区划单元)	核心属性	标识属性	中文名称	
2				行政区划代码	
3				管理属性	隶属行政区划
4			被管理行政区划		
5			空间属性		空间范围
6				面积	
7				时间属性	成立时间
8			变更或撤销时间		
9			日期		
10		拓展属性	社会人文类属性	地区生产总值	
11				第一产业生产总值	
12				第二产业生产总值	
13				第三产业生产总值	
14				人口数量	
15				人口性别构成	
16				人均可支配收入	
17				统计时间	
18				自然地理类属性	地形类型
19					平均海拔高度
20		流经河流名称			
21		主要矿产储量			
22	平均气温				
23	平均降水量				
24	粮食作物播种面积				
26			统计时间		
27	行政中心	基本属性	标识属性	中文名称	
28				别名	
29				中心类型	
30				所属行政区划名称	
31		所属行政区划代码			
32		空间属性	地址		
33			空间坐标		
34			空间范围		
35			地理参考系		
36		时间属性	设立时间		
37			终止时间		
38		管理属性	管辖范围		
39			包含部门		
40			行政职能		

注：行政区划包含的核心属性和拓展属性为国家级行政区、省级行政区、地级行政区以及县级行政区的共有属性；行政中心包含的属性为首都及地方政权驻地的共有属性。

本研究构建的行政区划沿革本体模型包括空间关系、层级关系、演变关系和关联关系 4 大类关系。空间关系是用于刻画行政区划实体在地理空间上的相互位置，层级关系是用于描述不同级别行政区划实体之间的隶属情况，演变关系则作为特殊的时间关系用于记录不同时期行政区划的变迁情况，关联关系用于将行政区划实体与其经济、自然等指标关联起来，详细关系类型及示例参见表 3。

表 3 本体部分关系类型与实例数据表

关系类组	序号	关系名称	示例
空间关系	1	相邻	河南省与山东省相邻
	2	包含	北京市包含朝阳区
层级关系	3	省辖市级为	河南省辖市级为洛阳市
	4	市辖县级为	洛阳市市辖县级为洛龙区
	5	直辖县级为	新疆维吾尔自治区直辖县级为铁门关市
演变关系	6	设立	设立漳州市龙海区
	7	撤销行政区划是	撤销县级龙海市
	8	名称变更为	山南地区名称变更为山南市
	9	区域合并为	山南地区区域合并为山南市
	10	区域拆分为	塔城地区区域拆分为塔城地区、胡杨河市
	11	政府驻地迁移至	铜川市人民政府驻地由王益区红旗街迁至耀州区正阳路
关联关系	12	区划代码变更	沙县区划代码变更为 350427
	13	地区生产总值是	2022 年北京市地区生产总值是 41,610.9 亿元
	14	地区生产总值-第一产业是	2022 年北京市第一产业地区生产总值是 111.5 亿元
	15	地区生产总值-第二产业是	2022 年北京市第二产业地区生产总值是 6,605.1 亿元
	16	地区生产总值-第三产业是	2022 年北京市第三产业地区生产总值是 34,894.3 亿元
	17	人口数量是	2022 年北京市人口数量是 2,184 万人
	18	人均可支配收入是	2022 年北京市人均可支配收入是 77,414.548,84 元
	19	平均降水量是	2022 年北京市年平均降水量是 585.4 mm
	20	平均气温是	2022 年北京市年平均气温是 13.4 °C
	21	农作物总播种面积是	2022 年北京市农作物总播种面积是 143.8 千 hm ²
		

3.3 知识抽取与对齐

本实验以 3.1 中提及的多种数据源为对象，按照第 2 章节使用的知识抽取与对齐方法完成行政区划实体、属性及关系的结构化获取与整理。在知识抽取阶段，针对中

国行政区划矢量数据，本文采用规则映射与字段解析的方法，提取矢量数据属性表中的中文名称、行政代码及隶属关系等静态属性；同时，利用 GeoPandas 对空间几何对象进行解析，将政区边界序列化为空间范围属性。以此为基础，引入 DE-9IM 空间拓扑算子与叠置分析等方法，横向识别同一年份行政区划间的相邻关系，纵向通过相邻年份空间对象对比，判定行政区划的设立、撤销、合并与拆分等演变关系。针对县级以上行政区划变更情况公告，采用本体语义驱动的大语言模型抽取方法，识别行政区划变更事件中的主体、演变类型、时间及批文信息等信息，对应本体设计中的演变关系；最后，针对中国省级统计年鉴中的统计指标，抽取为拓展属性信息，与对应行政区划实体建立属性关联关系。

在知识对齐阶段，针对不同数据源中抽取的行政区划实体、属性及关系，依次采用标识一致性对齐与属性相似度对齐方法进行融合。首先，基于行政区划代码、中文名称及时间点信息，对来自 GIS 矢量数据、统计年鉴与文本公告中的行政区划实体进行精准对齐；对于代码缺失或名称存在差异的实体，使用属性相似度对齐方法，综合比较中文名称与隶属行政区划等关键属性，完成实体合并与消歧。在关系层面，通过对齐后的实体标识统一空间关系与演变关系的指向对象。

3.4 知识图谱存储

本文选择 Neo4j 图数据库进行行政区划沿革知识图谱的存储与管理。Neo4j 是一个原生的图数据库，数据以节点、属性、关系的形式直接存储，实体及其语义关系可以自然地映射为图结构，支持高效的语义推理和路径分析。使用 py2neo 建立 Python 与 Neo4j 的连接池，采用 Neo4j 的 APOC 批处理技术，将 3.3 得到的节点、属性、关系分批次存储。最终，构建的中国行政区划沿革知识图谱包括 233,243 个行政区划节点和 89,073 个属性节点，具体节点类型以及对应数量如表 4 所示；中国行政区划沿革知识图谱涵盖的关系主要包括 1,352,945 个空间关系、268,689 个层级关系、35,033 个演变关系和用于关联拓展属性的 344,275 个关联关系，关系类型及关系数量示例参见表 5。

表 4 中国行政区划沿革知识图谱节点类型及节点数量示例

序号	节点类型	节点数量
1	省级行政区划	2,550
2	地级行政区划	24,754
3	县级行政区划	205,939
4	自然地理信息	36,788
5	社会人文信息	50,447

表 5 中国行政区划沿革知识图谱关系类型及关系数量示例

序号	关系名称	关系数量	序号	关系名称	关系数量
1	相邻	1,352,945	11	地区生产总值是	2,065
2	省辖市级为	24,755	12	地区生产总值-第一产业是	2,065
3	市辖县级为	240,132	13	地区生产总值-第二产业是	2,065
4	直辖县级为	3,802	14	地区生产总值-第三产业是	2,065
5	设立	6,918	15	人口数量是	682
6	撤销行政区划是	6,186	16	人均可支配收入是	272
7	名称变更为	2,528	17	平均降水量是	442
8	区域合并为	11,443	18	平均气温是是	442
9	区域拆分为	11,928	19	农作物总播种面积是	2,189
10	区划代码变更	9,134		……	

图 2 展示了图谱节点和关系的示例图。可以看到，图谱中不仅包含当前行政区划空间范围，还包含多个历史时期的行政区划实体，每个行政区对应其有效时间区间，通过这些丰富的属性信息就可以体现出行政区划的演变。以成都市为例，图谱中存在不同时间段的“成都市”节点，如图 3 所示：成都市（2016）、成都市（2017）通过演变关系（区域合并为）将两个时间段的节点连接起来，由此形成沿革链条，同时也存在资阳市（2016）通过演变关系（区域拆分为）与成都市（2017–2023）连接起来，总的来看，就是在 2016 年时成都市发生了演变，由原来成都市的区域加上资阳市的部分区域形成了新的成都市。

3.5 知识图谱构建方法可行性检验与质量评估

本文将行政区划沿革知识图谱记录的演变数据与民政部公布的县级以上行政区划变更情况进行对比实验，验证中国行政区划沿革知识图谱的体系化构建方法可行性与质量。具体步骤为：首先，收集民政部发布的县级以上行政区划变更情况公告资料；其次，随机选取 2015–2023 年间不同年份的行政区划变更记录提取变更类型、变更前后行政区划、以及变更前后隶属信息作为验证集。最后，将本知识图谱与民政部发布的县级以上行政区划变更情况公告主要对比 4 个方面的信息：

一是行政区划的撤销与设立。在本图谱中，主要体现为撤销后下一年节点是否继续存在，以及设立后节点是否成功生成。二是行政区划的名称信息是否准确。在本图谱中体现为节点的属性信息是否与公告记录一致。三是公告中记录的行政区划隶属信息是否准确。在本图谱中，这体现为节点之间的隶属关系。最后，变更记录的准确性，在本图谱中表现为前后行政区划节点之间是否存在特定的演变关系。差异对比结果示例参见表 6。

表 6 中国行政区划沿革知识图谱记录与民政部公告差异对比示例

年份	官方公告内容	数据集记录	差异说明
2016	安徽省人民政府由合肥市庐阳区长江路 221 号迁移至合肥市包河区中山路 1 号	数据集中记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无政府驻地属性
2016	河北省人民政府驻地由石家庄市桥西区维明南大街 46 号迁至石家庄市长安区裕华东路 113 号	数据集中记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无政府驻地属性
2016	撤销邯郸县	数据集中邯郸县记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无邯郸县
2016	撤销宜州市，设立河池市宜州区	数据集中宜州市记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无宜州市，宜州区从 2015 年开始就已经存在
2018	泽州县人民政府由现驻地迁至金村镇府城街 001 号	数据集中记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无政府驻地属性
2018	撤销茫崖行政委员会和冷湖行政委员会，设立县级茫崖市	数据集中茫崖行政委员会和冷湖行政委员会记录为空	数据集设计省级、地级、县级三级行政区划，委员会不属于这三级
2018	北京市人民政府驻地由东城区正义路 2 号迁至通州区运河东大街 57 号	数据集中记录为空	中国行政区划矢量数据和中国省级统计年鉴数据源中无政府驻地属性
2020	海南省三沙市设立西沙区、南沙区	数据集西沙区、南沙区日期属性均为 2015–2023	中国行政区划矢量数据和中国省级统计年鉴数据源中西沙区、南沙区从 2015 年开始就已经存在

造成上述差异的主要有 3 种原因：其一，知识图谱中数据设计仅覆盖省级、地级与县级行政区划，不包括行政委员会等特殊机构，从而造成记录缺失；其二，个别信息（如政府驻地）在原始数据中缺乏系统性收录，难以在图谱中体现；其三，部分区划调整在数据源中存在时间提前或延后的现象，导致与公告发布时间不完全一致。

随后使用验证集在行政区划沿革知识图谱中进行查询，使用召回率、准确率和 F1 值验证知识图谱的查询性能。通过比对，发现知识图谱中记录的行政区划信息与民政部公布行政区划变更情况相比，平均召回率、平均准确率和平均 F1 值均超过 95%，整体记录与官方资料高度一致，证明了本文提出的中国行政区划沿革知识图谱的构建方法体系的可行性。具体性能的比对结果见表 7。

表 7 中国行政区划沿革知识图谱数据集查询性能

	2016 年	2018 年	2020 年	2021 年	平均值
召回率	0.918	0.947	0.944	1.000	0.952
准确率	0.949	0.950	0.944	1.000	0.961
F1 值	0.933	0.948	0.944	1.000	0.956

注：平均值为各指标的算术平均值。

4 总结与展望

针对当前传统行政区划数据集因来源多样、格式异构导致的数据分散存储、跨数据集关联困难以及复杂演变查询难以实现等问题，本文提出并实现了一套面向中国行政区划沿革知识图谱的系统化构建方法，涵盖“本体建模-多源知识抽取-实体对齐融合-图数据库组织”4个关键环节。首先，构建了以行政区划实体为核心，以空间关系、层级关系和演变关系为主要关系类型的行政区划沿革本体模型；其次，提出了针对矢量空间数据、统计年鉴和公告文本的差异化知识抽取方法；然后，设计了基于标识一致性与属性相似度的实体对齐方法，解决了多源数据中行政区划实体重复、异名和歧义等问题；最终利用 Neo4j 图数据库与 APOC 库的批处理机制，实现了中国行政区划沿革知识图谱的存储与可视化，同时对图谱质量进行了验证。

未来研究可从多个方向进一步拓展。一方面，将当前时间范围向更长历史时期延伸，构建覆盖更长周期的行政区划沿革知识图谱，以揭示更宏观的演变规律；另一方面，在现有基础上融入更细粒度的社会经济数据、政策法规文本、历史文献等资料，进一步丰富知识图谱的节点属性和关系类型，探究区划变迁的深层驱动机制；最后，与大模型相结合，借助大模型强大的语言理解和生成能力，实现更智能化的行政区划变迁查询与分析功能，降低用户使用门槛，使其能够更直观、便捷地探索和利用行政区划沿革知识。

作者分工：刘纪猛、王曙、诸云强和韩保民负责方法的总体设计；刘纪猛、王春玲和葛双双进行了方法的研究与实践、论文写作；代小亮采集了中国行政区划沿革知识图谱构建数据源数据；刘纪猛、葛双双、王曙和诸云强参与了论文的修改。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] 任晓茹, 任福, 陈慧萍等. 湖北省政区历史沿革展示系统的设计与实现[J]. 测绘地理信息, 2017, 42(3): 6.
- [2] 赵逸才, 王开泳, 赵彪等. 过去 2200 年中国县级政区设置的时空过程与格局[J]. 地理学报, 2024, 79(4): 890-908.
- [3] Feng, R. D., Wang, K. Y. The direct and lag effects of administrative division adjustment on urban expansion patterns in Chinese mega-urban agglomerations [J]. *Land Use Policy*, 2022, 112: 105805.
- [4] Feng, R. D., Wang, K. Y. Spatiotemporal effects of administrative division adjustment on urban expansion in China [J]. *Land Use Policy*, 2021, 101: 105143.
- [5] Wang, F. L., Liu, Y. G. China's urban planning and administrative urbanization: case of Ordos [J]. *Urban design and planning*, 2014, 167(5): 196-208.
- [6] Chen, Y. L., Yu, P. H., Wang, L., et al. Polycentric urban development with state-led administrative division adjustment: a policy insight for urban spatial transformation [J]. *Journal of Geographical Sciences*, 2023, 33(12): 2400-2424.
- [7] Feng, R. D., Wang, K. Y., Wang, F. Y. Quantifying influences of administrative division adjustment on PM_{2.5} pollution in China's mega-urban agglomerations [J]. *Journal of environmental management*, 2022,

- 302: 113993.
- [8] Wei, S., Zheng, W., Wang, L. Understanding the configuration of bus networks in urban China from the perspective of network types and administrative division effect [J]. *Transport Policy*, 2021, 104: 1–17.
- [9] Zhu, J. H., Chen, X., Chen, T. Spheres of urban influence and factors in Beijing-Tianjin-Hebei Metropolitan Region based on viewpoint of administrative division adjustment [J]. *Chinese Geographical Science*, 2017, 27(5): 709–721.
- [10] 中华人民共和国民政部. 中华人民共和国二〇二三年县级以上行政区划变更情况[EB/OL]. (2023-04-03) [2025-09-16]. <http://xzqh.mca.gov.cn/description?dcpid=2023>.
- [11] 段柄仁. 北京志·统计志[M]. 北京: 北京市地方志编纂委员会, 2016.
- [12] 戴均良, 白泊. 中华人民共和国行政区划图集[M]. 北京: 中国地图出版社, 2005.
- [13] 中华人民共和国民政部, 国家测绘局. 中华人民共和国行政区划图集[M]. 北京: 中国地图出版社, 2005.
- [14] 戴均良. 中华人民共和国行政区划简册[M]. 北京: 中国社会出版社, 2009.
- [15] 陈洪玲. 中华人民共和国行政区划沿革地图集(精)[M]. 北京: 中国地图出版社, 2003.
- [16] 杨雅萍. 中华人民共和国行政区划变迁地图集(1980–2017)[M]. 北京: 中国地图出版社, 2016.
- [17] 联合国粮食及农业组织(FAO). 行政边界(1级)- GADM 3.6[EB/OL]. (2024-7-30) [2025-09-16]. <https://data.apps.fao.org/catalog/dataset/aecbbc85-2a46-498b-83b4-beca24178f71>.
- [18] Geofabrik 数据平台. 中国地区 OpenStreetMap 数据集[EB/OL]. (2025-09-15) [2025-09-16]. <https://download.geofabrik.de/asia/china.html>.
- [19] 天地图服务中心. 中国行政区划数据集 [EB/OL]. (2024-05-01) [2025-09-16]. <https://cloudcenter.tianditu.gov.cn/administrativeDivision>.
- [20] 鲁艳旭, 张雪英, 张春菊. 事件视角下的行政区划沿革知识图谱构建方法[J]. 地球信息科学学报, 2025, 27(10): 2440–2452.
- [21] 朱建华, 陈田, 王开泳等. 改革开放以来中国行政区划格局演变与驱动力分析[J]. 地理研究, 2015, 34(2): 247–258.
- [22] 陈书慧. 建国以来行政区划沿革知识图谱构建方法研究[D]. 南京: 南京师范大学, 2022.
- [23] Clementini, E., Di Felice, P., van Oosterom, P. A small set of formal topological relationships suitable for end-user interaction [J]. *Springer Berlin Heidelberg*, 2005, 277–295.
- [24] 邵滢滢, 王庆, 孙琨. GIS 中基于拓扑信息的矢量图形叠置方法[J]. 东南大学学报(自然科学版), 2006, 36(3): 442–445.
- [25] 王春玲, 诸云强, 王曙等. 考虑时空特征及演化关系的行政区划知识图谱构建[J]. 地球信息科学学报, 2026, 28(1): 89–104.
- [26] 翁剑成, 葛昱, 王昌等. 面向公交服务评价的公交出行指数与分析模型[J]. 公路交通科技, 2016, 33(1): 130–134.
- [27] 庄严, 李国良, 冯建华. 知识库实体对齐技术综述[J]. 计算机研究与发展, 2016, 53(1): 165–192.
- [28] Rui, C., Zhang, H. F., Chen, B. Z. China temporal administrative map: a multitemporal database for Chinese historical administrative divisions (2009–2023) [C]. Third International Conference on Environmental Remote Sensing and Geographic Information Technology (ERSGIT 2024). SPIE, 2025, 13565: 497–507. <https://doi.org/10.1117/12.3059430>.