

大模型驱动的东天山—北山找矿知识图谱构建及应用

田苗¹, 余吉远^{2*}, 田龙星³, 谢忠³, 陈建国⁴, 邱芹军³, 吴麒瑞⁵

1. 中国地质大学(武汉) 地质探测与评估教育部重点实验室, 武汉 430074; 2. 中国地质调查局 西安地质调查中心(西北地质科技创新中心), 西安 710119; 3. 中国地质大学(武汉) 计算机学院, 武汉 430074; 4. 中国地质大学(武汉) 资源学院, 武汉 430074; 5. 中国地质大学(武汉) 未来技术学院, 武汉 430074

摘要: 当前利用新一代智能信息找矿技术对于智能信息找矿具有重要意义。面对复杂成矿过程及多源异构地质数据, 已有专家经验规则及找矿方法无法满足矿产资源预测的智能化需求。为解决这一问题, 本文自上而下构建了面向智能找矿的地质本体, 并在此基础上聚焦于东天山—北山成矿带, 以东天山—北山期刊文献、会议论文及矿产资源地质报告为数据源, 提出了一种基于Deepseek-reasoner大语言模型和提示学习相结合的地质实体关系抽取方法, 对其中蕴含的39种关系进行了抽取, 构建形成东天山—北山找矿知识图谱。在此基础上, 本文利用智能问答、社区发现等方法开展东天山—北山找矿知识图谱的验证与应用, 为东天山—北山成矿带矿产资源预测提供有力的技术方法支撑。

关键词: 面向智能找矿的地质本体; 大语言模型; 提示模板; 地质实体关系抽取; 找矿知识图谱应用

中图分类号: P612 文章编号: 1007-2802(2026)03-0867-14 doi: 10.3724/j.issn.1007-2802.20260008

The construction and application of a Knowledge Graph for prospecting mineral resources in the East Tianshan—North Mountain region based on the large language model

TIAN Miao¹, YU Ji-yuan^{2*}, TIAN Long-xing³, XIE Zhong³, CHEN Jian-guo⁴, QIU Qin-jun³, WU Qi-rui⁵

1. Key Laboratory of Geological Survey and Evaluation of Ministry of Education, China University of Geosciences, Wuhan 430074, China;
2. Xi'an Center of Geological Survey, China Geological Survey (Geosciences Innovation Center of Northwest China), Xi'an 710119, China;
3. School of Computer and Science, China University of Geosciences, Wuhan 430074, China;
4. School of Earth Resources, China University of Geosciences, Wuhan 430074, China;
5. School of Future Technology, China University of Geosciences, Wuhan 430074, China

Abstract: The current utilization of the new generation intelligent information-based mineral exploration technology is of great significance for the intelligent information-based mineral exploration. Facing with complex mineralization processes and multi-source heterogeneous geoscientific data, the existing expert experience rules and mineral exploration methods are insufficient to meet the intelligent demands of mineral resource prediction. To address this issue, we have constructed a geological ontology for intelligent mineral exploration in a top-down manner in this paper. Based on this, by focusing on the East Tianshan—North Mountain mineralization belt, using journal articles, conference papers, and mineral resource geological reports related to the East Tianshan

收稿编号: 2025-0273, 2025-11-24 收到, 2026-01-17 改回

基金项目: 国家自然科学基金资助项目(42301492); 江苏省自然资源厅地质数据智能应用技术创新中心开放基金资助项目(GDIATIC-XM-202502); 中国地质调查局地质调查项目(DD20240029)

第一作者简介: 田苗(1995—), 女, 在读硕士研究生, 研究方向: 地质知识图谱构建与推理. email: tianmiao@cug.edu.cn

*通信作者简介: 余吉远(1978—), 男, 硕士, 教授级高级工程师, 研究方向: 岩石地球化学及综合研究. email: yujiyuan111@163.com

引用此文:

田苗, 余吉远, 田龙星, 谢忠, 陈建国, 邱芹军, 吴麒瑞. 2026. 大模型驱动的东天山—北山找矿知识图谱构建及应用. 矿物岩石地球化学通报, 45(3): 867–880
Tian M, Yu J Y, Tian L X, Xie Z, Chen J G, Qiu Q J, Wu Q R. 2026. The construction and application of a Knowledge Graph for prospecting mineral resources in the East Tianshan—North Mountain region based on the large language model. Bulletin of Mineralogy, Petrology and Geochemistry, 45(3): 867–880

-North Mountain region as data sources, we have proposed a method for extracting geological entity relationships based on the combination of the Deepseek-reasoner large language model and the prompt learning, then have extracted 39 types of relationships contained in the data sources and further have constructed a knowledge graph for the mineral exploration in the East Tianshan-North Mountain region. Building on this foundation, in this paper, we have used methods such as the intelligent question-answering and community detection to have validated and applied the knowledge graph for the mineral exploration in the East Tianshan-North Mountain region. These efforts provide robust technological and methodological support for predicting mineral resources within the East Tianshan-North Mountain metallogenic belt.

Key words: ore deposit ontology; large language models; prompt templates; extracting geological entity relationships; application of the knowledge graph for the exploration of mineral resources

0 引言

矿产地质勘探长期依赖地质物理、地球化学等数据和专业知识开展野外探测,取得了显著找矿成果(张春菊等,2023;周永章和肖凡,2024)。随着国家对矿产资源开发力度的加大,地表矿产资源日益减少,而深部矿产勘探面临实际地质数据严重匮乏的困境。因缺乏第一手地质工作数据,深部矿产勘探只能依赖间接手段推测,导致找矿难度不断增大(张春菊等,2023;周冠云等,2023;陈建平等,2024)。因此,对重点矿区进行的深部找矿,有效预测隐伏矿床已经成为矿产勘探行业重点关注的问题。大数据与人工智能时代为更加精准和高效地智能找矿提供了技术手段(张前龙等,2024a,2024b)。

多模态地质数据具有多源异构、时空跨度大、语义复杂等显著特征(王翔等,2015;邱芹军等,2024)。作为地质数据的基础载体,矿产资源地质文本数据中不仅记录了区域矿床、地层、地质构造、岩石、矿物、蚀变等关键地质信息,而且以文本、图件与数据表的形式完整保存了找矿线索与成矿规律,是后续开展矿产资源评价、成矿预测与勘查决策不可或缺的基础依据(Mantovani et al.,2020;Qiu et al.,2023a,2023b,2024)。然而,地质文本数据在数据提取、整理和分析过程中,由于缺乏统一的语义基础(魏东琦等,2021),导致其在地质找矿等应用场景中面临语义识别困难、语义整合与利用效率低下等突出问题(汪陈,2021;Qiu et al.,2023a,2023b,2024)。鉴于此,如何从数据中挖掘潜在的非显式表达,构建统一的语义表达体系,进而实现地质数据的整合共享、关联集成和深度挖掘分析,已成为当前地质科学研究领域的重大课题(张春菊等,2023;周仲礼等,2025)。以ChatGPT、DeepSeek等为代表的通用大语言模型(LLMs)在数据提取、分析和推理方面的突出能力为解决上述问题提供了新方向(周波和李珂,2025)。

找矿知识图谱以本体模型为核心,融合了多源地质数据与语义表示方法,通过规范化表达、智能化抽取、语义化关联等方式形成地质语义网络,通过关联矿床地质、矿物、矿体及勘探开发技术等要素及其关系,构建知识图谱结构,助力用户理解与利用知识(曹胜桃等,2024;张前龙等,2024a,2024b;何陆灏等,2025)。作为地球系统-成矿系统-找矿系统关联的知识图谱体系,为矿产资源预测与评估提供了新途径。知识图谱作为一种语义网络,可以将矿产地质资料中蕴含的各种矿床、岩石、地层、成矿地质体等地质对象实体、属性及地质语义关系进行表达,从而将大量非结构化文本数据转换为结构化的领域知识,对后续进行知识挖掘与推理、智能找矿预测具有重要意义(Pei et al.,2022;Qiu et al.,2024,2025)。

近年来,许多专家学者在本体构建(张春菊等,2023;Tian et al.,2025;吴麒瑞等,2025)、地质实体关系抽取研究方面取得了一定成果,但是面向下游找矿知识图谱的智能应用研究则严重偏弱,而矿床的成矿模式是构造运动、岩浆活动、地层等多维地质要素的复杂综合性结果,矿床的形成涉及构造演化、地质特征、地球化学及热液作用等多种因素,限制了找矿勘查的突破。

本研究以东天山-北山成矿带为研究对象,开展找矿知识图谱构建及智能化应用研究。以东天山-北山区域地质报告、矿产资源调查报告、期刊会议论文等文本为数据源,开展大模型驱动的东天山-北山找矿知识图谱构建及智能化应用。

1 找矿知识图谱总体架构设计

东天山-北山找矿知识图谱的构建主要以矿床本体设计为基础,以东天山-北山区域地质数据为数据源,在分析整理矿产资源领域实体及语义关系的基础上,通过自顶向下和自底向上相结合的方法构建了东天山-北山找矿知识图谱(图1)。首先,本体层的构建

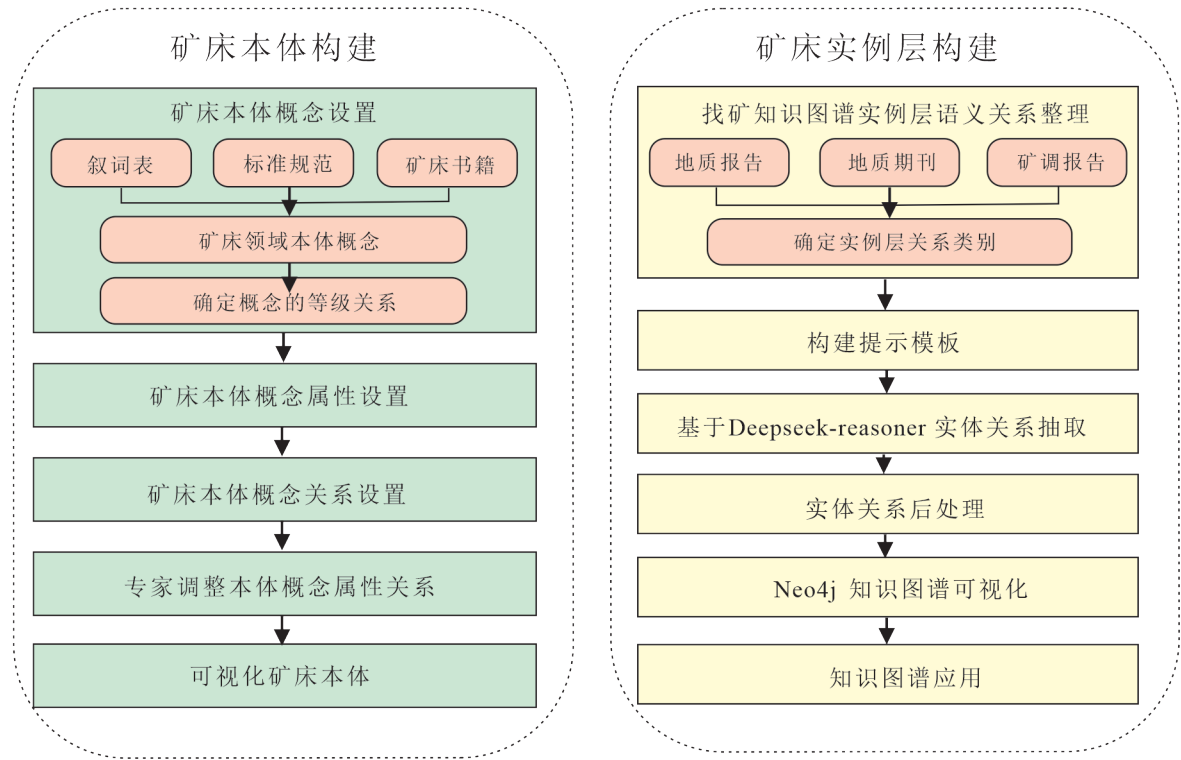


图1 东天山-北山找矿知识图谱构建流程

Fig.1 Construction process of the knowledge graph for the exploration of mineral resources in the Eastern Tianshan- North Mountain region

主要包含矿床本体概念的定义、矿床本体概念属性的设计以及矿床本体概念关系的设计。其次,针对实例层构建主要采用自底向上的方法进行,其通过搜集东天山-北山期刊文献、会议论文、地质报告以及矿产资源调查报告等相关资料,在分析整理找矿地质实体及语义关系的基础上,利用基于提示学习的大语言模型完成实体关系三元组的抽取,最终构建形成东天山-北山找矿知识图谱,并利用Neo4j对已构建的知识图谱进行储存。

2 矿床本体构建

矿床反映了成矿物质、来源、环境和成矿作用等多种地质过程,以矿床为单位有助于揭示“成矿地质体”与“成矿作用标志”的因果联系,从而展示成矿事件的时空整体性。本文主要参考现有标准规范、矿床书籍、叙词表以及专家知识等,在分析整理矿产资源领域矿床、矿段、成矿地质体、矿石、矿物等层次关系的基础上通过自顶向下方法构建矿床本体。

矿床概念模型图(图2)主要对矿床中矿区、矿段、矿体、成矿地质体、矿石、矿物等关键概念及这些概念之间的关系进行梳理,其中矿床的属性主要包含基本属性、空间属性、时间属性和非空间属性,基本属性包含概念名称、矿床类型,时间属性主要包含成矿年代和

发现时间,空间属性包括形态、走向、经纬度和所处地名,非空间属性包括成矿地质构造、围岩蚀变、地层、地球化学特征、控矿因素、产状、规模、地球物理特征、成矿作用标志和地质异常。在矿床中主要含有的地质构造、岩石、地形地貌、地名地址、矿物、地质年代、地层等地质实体和矿井、采矿场、公路、铁路等自然地理实体。矿段、矿体、成矿地质体、矿石均包含基本属性和地质属性,其中基本属性在“顶级概念(这里指矿床)”中定义,它可以被下一级概念(如矿区、矿段、矿体、矿石等)继承。

2.1 矿床本体概念定义

在矿产地质领域,矿床描述的概念主要包括矿区、矿体、岩石、矿石、矿物、地层、地质构造等。其中,矿区是指统一规划开发的矿产区域,其规模主要根据矿床的储量而定;矿体通常作为矿床的一部分,是指含有足够数量的矿石且具有开采价值的地质体,其具有一定的形状、产状及规模;而矿石是指可从中提取有用组分或其本身具有某种可被利用的性能的矿物集合体;此外,岩石作为矿床的载体以及矿床成因的基础,岩石为矿物的富集提供了物质基础和地质背景,矿床的位置也通常会受到岩石构造变化的影响。

2.2 矿床本体概念属性设计

矿床本体概念的属性主要是对概念本身进行描

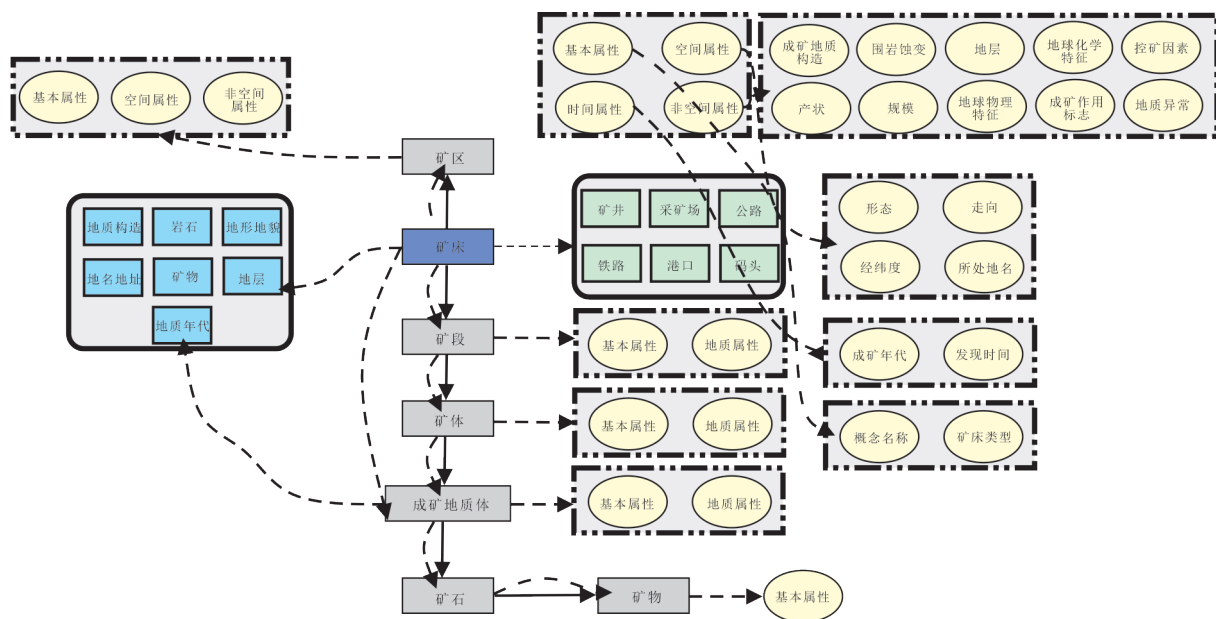


图2 矿床本体概念设计图

Fig.2 Conceptual design diagram of the deposit ontology

述和约束,因此本文主要对矿床、矿体、矿石、岩石、矿物这五大类概念进行了属性分类。由于矿床反映了成矿物质、来源、环境和成矿作用等多种地质过程,因此矿床主要包含空间特征和基本特征两大类属性;而矿体、岩石、矿石和矿物则主要分为基本特征和地质特征两大类特征,其具体属性定义如图3所示。

2.3 矿床本体概念关系设计

针对本体关系建模,本文主要定义了空间关系、构成关系、时间关系、语义关系/属种关系、关联关系和属性关系6种(表1)。可以看到,矿床、矿区、矿段、矿体、成矿地质体、矿石、矿物、岩石和地层这些概念之间均存在空间分布关系,矿区、矿床、矿段和矿体之间存在包含关系,矿床、矿区和矿段、矿体、成矿地质体、矿石从大到小,形成层次性结构。

3 东天山-北山找矿知识图谱实例层构建

3.1 数据源

东天山-北山成矿带位于中亚造山带南缘中段,主要由一系列微陆块、岛弧、海山大洋高原增生杂岩及蛇绿混杂岩构成,经历了新元古代-早中生代复杂的演化历程,形成了铜、镍、铁、钨、金等优势矿产(周军等, 2017; 贾晨辉, 2020; 张振亮等, 2022; 龙灵利等, 2025)。长期的地质调查形成了海量的东天山-北山成矿带地质资料,包括区域地质图、数据库及联测报告、项目成果报告,详细解剖调查区不同地质体的物质组成、结构构造、变质变形特征,系统梳理成矿带赋铜、镍、金矿等地质体的物质组成、改造过程、蚀变特征和

成矿时代等,完善成矿带构造框架,进而重建成矿带构造演化过程,分析不同构造单元在不同演化阶段的有利成矿地质条件,对于进一步探索中亚造山带南缘中段东天山-北山成矿带古生代俯冲-增生过程及造山后构造改造对铜、镍、金等元素富集的控制机理,综合评价其成矿潜力具有重要的研究意义及价值。

因此,本文主要聚焦于东天山-北山成矿带,搜集了与之相关的期刊文献、会议论文和区域地质调查报告共500多篇。首先对搜集的所有数据进行预处理,去除无关的章节内容,然后根据专家经验、地质词典和百度百科等制定相应的抽取及命名规则,为后续进行提示模板的设置提供依据。

3.2 基于找矿知识图谱的实体关系建模

针对上述搜集整理的数据进行分析,对东天山-北山找矿知识图谱实例层语义关系进行划分,其中将一级语义关系划分为包含关系、空间关系、时间关系、属性关系和功能关系等五大类,空间关系又分为拓扑关系、顺序关系和度量关系等三类,具体见表2。

3.3 基于大模型的东天山-北山实体关系抽取

本文主要以东天山-北山地区的期刊文献和地质报告为数据源,针对上述归纳的实体关系类别,利用基于提示模板和Deepseek-Reasoner(Deepseek-v3.2)大语言模型相结合的方法进行实体关系三元组抽取。为进一步提高东天山-北山实体关系抽取的结果,首先对搜集的数据进行整理和分析,其中重点对矿床、地层、岩石等地质实体的命名规则进行整理(表3)。由于所搜集的期刊文献和地质报告中包含大量的地质术语,

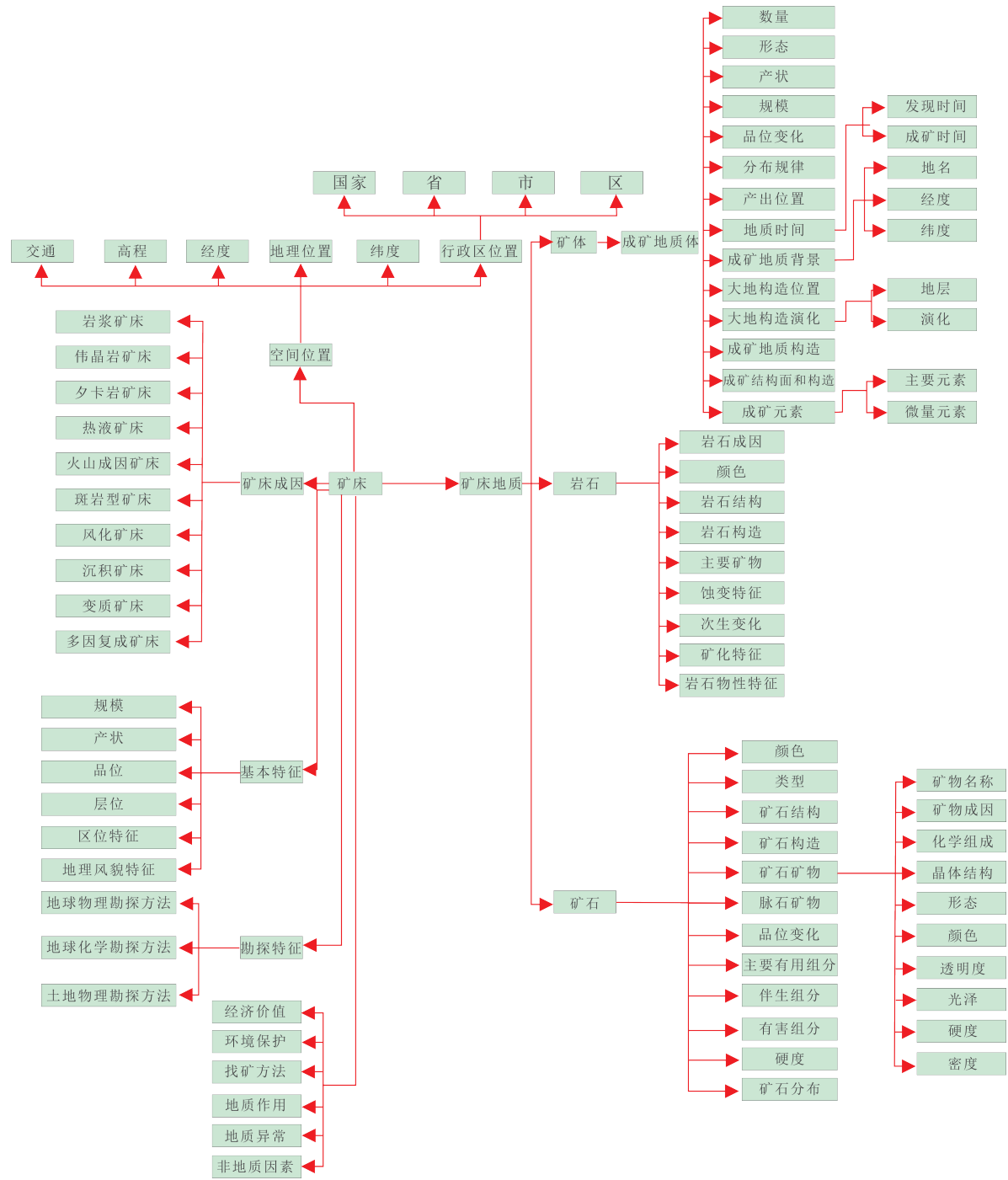


图3 矿床本体概念属性设计图

Fig.3 Conceptual attribute design diagram of the deposit ontology

表1 矿床本体概念关系设计

关系类别	关系名称
空间关系	空间分布关系
时间关系	发现时间、成矿时间
语义关系	包含关系
关联关系	层次关系
属性关系	属性关系

这些术语的识别相对复杂,且具有较高的领域特性,因此基于大语言模型的抽取效果与提示词的设计密切相关。通过整理这些术语的命名规则,能够提高抽取结果的准确性和效果。其次,在上述基础上对所搜集的期刊文献和地质报告中一些常见的关系对应的实体类别进行整理(表4)。第三,为更好提升东天山-北山实体关系抽取的精度,对若干特殊地质术语进行了系统整理构建了提示模板(表5),并通过列举具体的抽取示

表 2 找矿知识图谱实例层语义关系列表

Table 2 Schematic relationship list of the instance layer in the mineral exploration knowledge graphs		
一级语义关系	二级语义关系	举 例
包含关系	分别为	(大断裂南北侧,分别为,泥盆纪酸性侵入体) (大断裂南北侧,分别为,志留纪中酸性侵入体)
	拓扑关系(位于)	
空间关系	顺序关系	(闪长岩,顺序关系,粗面岩)
	度量关系	
时间关系	晚于、同生成矿、后生成矿、成矿期	(铁矿,同生找矿,镍矿)
属性关系	包含、属于、含有、为、成因类型、颜色、类型、矿石结构、矿石构造、长度、倾角、厚度、形态、展布、面积、粒度、形状、产出	(萤石矿,矿石结构,他形-半自形粒状结构)
		(萤石矿,矿石结构,碎裂结构)
		(萤石矿,矿石结构,自形结构)
		(萤石矿,矿石结构,碎斑胶结结构)
功能关系	发育、控矿构造、围岩蚀变、矿物组合、受控于、岩性、主要矿物、次要矿物、金属矿物、脉石矿物、矿石矿物、地层、地质年代、出露、赋存于、大地构造单元、大地构造位置、主要元素、微量元素	(矿区,围岩蚀变,硅化)
		(矿区,围岩蚀变,绿泥石化)
		(矿区,围岩蚀变,高岭土化)

表 3 常见地质实体命名规则

Table 3 Naming conventions for common geological entities		
地质实体	命名规则	举 例
岩石	一般以“岩石”或“岩”结尾	火成岩、沉积岩、变质岩、花岗岩、脉岩、橄榄岩等
地层	一般以“系”“组”“群”“宇”“界”“统”“阶”结尾	企鹅山群、底坎儿组、海相火山-沉积体系、小热泉子组、下石炭统干墩组、下-中奥陶统恰干布拉克组
地质年代	一般以“宙”“代”“纪”“世”“期”“时”结尾	中生代、元古宙
矿物	一般金属矿物以“矿”结尾;具有玻璃或金属光泽的矿物以“石”结尾;硫酸盐矿物以“矾”结尾;玉石类矿物以“玉”结尾;地表松散性矿物以“华”结尾	方铅矿、黄铜矿、赤铁矿、钾长石、绿泥石、孔雀石、闪镍矿、方解石、孔雀石、胆矾、铅矾、硬玉、软玉、神华、镍华、钨华
地质构造	一般以“构造”结尾	褶皱、断裂、劈理、节理、断层以及其他面状、线状构造
矿床	一般以“矿床”结尾	铜矿床、金矿床、铅矿床、铁矿床
矿石	一般以“矿石”结尾	贫矿石、普通矿石、富矿石
地名	一般以“山”和“区域”结尾	红石山、东天山、北山

例,进一步指导大语言模型在实体关系抽取任务中的有效学习和优化。

在上述地质实体关系命名规则定义及提示模板构建的基础上,本文针对特殊命名的地质实体也进行了相应的整理,并在提示模板中给出了相应的示例,如常见的矿物、矿石等(表6)。此外,针对地质实体关系抽取过程中存在的一些识别不精准等问题,采用了少样本示例的方式进行了总结归纳(表7),大模型通过学习这些少样本抽取示例,进而进一步提升地质实体关系的准确率。

3.4 知识图谱可视化

本文以东天山-北山的期刊文献数据和地质报告为数据源,利用基于提示学习的大语言模型对其中蕴含的39种实体关系进行抽取,共抽取得到三元组48 971条,并对每种关系的数量进行统计(图4)。此外,为了进

一步验证提示模板对于地质实体关系抽取精度的重要性,本文对比了使用提示模板前后使用Deepseek-reasoner大语言模型提取地质实体关系三元组的结果(表8),清晰展示了使用基于提示模板的Deepseek-reasoner大语言模型更能提高地质实体关系抽取的准确性。最后采用Neo4j数据库对提取的实体关系三元组进行存储,其可视化结果如图5所示。该知识图谱充分展示了东天山-北山矿床的多维度信息,为后续进行高效的数据管理与查询提供了进一步支撑。

4 东天山-北山找矿知识图谱的应用

4.1 基于知识图谱的智能查询

本研究利用已构建的东天山-北山找矿知识图谱,从简到难设计了一系列问答模板,然后利用Cypher查询语句对已构建的东天山-北山找矿知识图谱进行

表 4 常见地质关系命名规则

Table 4 Naming conventions for common geological relationships

地质关系	命名规则	举 例
矿石结构	subject一般是某一种矿石,object的命名一般以“结构”结尾	萤石矿矿石结构主要为他形-半自形粒状结构、碎裂结构,次为自形结构、碎斑胶结结构
矿石构造	subject一般是某一种矿石,object的命名一般以“构造”结尾	矿石具自形-半自形粒状结构,局部具陨铁结构。矿石构造以星散浸染状-稀疏浸染状为主,局部具稠密浸染状
岩性	object一般是某一种岩石	泥盆系下部岩性为一套深海的放射虫硅质岩、基性火山岩、碎屑岩夹碳酸盐岩
展布	object的命名一般以“向”结尾	黄山南矿床地表形态呈透镜状,近东西向展布,面积约为4 km ²
形态	object的命名一般以“状”结尾	黄山南矿床地表形态呈透镜状,近东西向展布,面积约为4 km ²
围岩蚀变	object的命名一般以“化”结尾	矿区围岩蚀变主要有硅化、绿泥石化、高岭土化等,为中低温热液蚀变组合
发育	object一般是岩石或地质构造,也有可能是地质年代	二叠纪时期北山地区广泛发育碱长花岗岩、黑云母二长花岗岩、角闪二长花岗岩
出露	object一般是岩石或者地层	张家口期火山岩系直接覆盖在古老基底或晚古生代花岗岩之上,火山岩呈大面积出露,火山喷发强度非常大,以强烈喷发的火山灰流相为主

表 5 提示模板示例

Table 5 Prompt template examples

"""你是专业知识三元组提取器,严格按以下规则输出:
1. 仅从文本提取(主体,关系,客体)三元组,忽略无关信息
2. 必须用JSON数组格式返回,每个元素含"subject""relation""object"字段
3. 抽取得到的三元组,relation的类型为[晚于、成矿期、包含、属于、含有、为、成因类型、颜色、类型、矿石结构、矿石构造、长度、倾角、厚度、形态、展布、面积、粒度、形状、产出、发育、控矿构造、受控于、岩性、主要矿物、次要矿物、金属矿物、脉石矿物、矿石矿物、地层、地质年代、出露、赋存于、大地构造单元、大地构造位置、主要元素、微量元素、围岩蚀变、矿物组合]
.....
"""

表 6 特殊地质实体举例

Table 6 Examples of special geological entities

实体类别	特殊命名实体
矿物	石英、云母、黑云母、磷、锆、铌、钽、铜蓝、石盐、钾盐、玉髓、绢云母等
大地构造位置	哈萨克斯坦-准噶尔板块、塔里木-华北板块等
大地构造单元	星星峡-早山微板块、敦煌微板块、大南湖-雀儿山活动陆缘带、觉罗塔格晚古生代岛弧、北山断褶带、库鲁克塔格-星星峡断隆
地质构造	褶皱、断裂、劈理等

表 7 少样本示例

Table 7 The few-shot examples

句 子	抽取结果示例
张家口期火山岩系直接覆盖在古老基底或晚古生代花岗岩之上,火山岩呈大面积出露,火山喷发强度非常大,以强烈喷发的火山灰流相为主	(张家口期火山岩系,出露,火山岩)
新疆黄山东铜镍矿床成矿类型属于岩浆熔离-贯入型硫化物铜镍矿床,矿床赋存于镁铁-超镁铁质岩体中	(新疆黄山东铜镍矿床,成矿类型,岩浆熔离-贯入型硫化物铜镍矿床); (新疆黄山东铜镍矿床,赋存于,镁铁-超镁铁质岩体)
黄山东基性-超基性岩体的主要岩石类型为闪长岩、角闪辉长岩、辉长岩、辉长苏长岩、苏长岩、角闪橄榄岩和二辉橄榄岩	(黄山东基性-超基性岩体,包含,闪长岩);(黄山东基性-超基性岩体,包含,角闪辉长岩);(黄山东基性-超基性岩体,包含,辉长苏长岩);(黄山东基性-超基性岩体,包含,辉长岩);(黄山东基性-超基性岩体,包含,苏长岩); (黄山东基性-超基性岩体,包含,角闪橄榄岩);(黄山东基性-超基性岩体,包含,二辉橄榄岩)
萤石矿矿石矿物单一,为萤石,主要呈浅绿色、白色,少量为绿色、紫色;脉石矿物因矿石类型不同而不同,对于条纹状、角砾状矿石,脉石矿物主要为石英、蛋白石、玉髓,次要矿物为绢云母、绿泥石及花岗质矿物	(萤石矿,矿石矿物,萤石);(萤石,颜色,浅绿色);(萤石,颜色,白色);(萤石,颜色,绿色);(萤石,颜色,紫色);(萤石矿,脉石矿物,石英);(萤石矿,脉石矿物,蛋白石);(萤石矿,脉石矿物,玉髓);(萤石矿,次要矿物,绢云母); (萤石矿,次要矿物,绿泥石);(萤石矿,次要矿物,花岗质矿物)

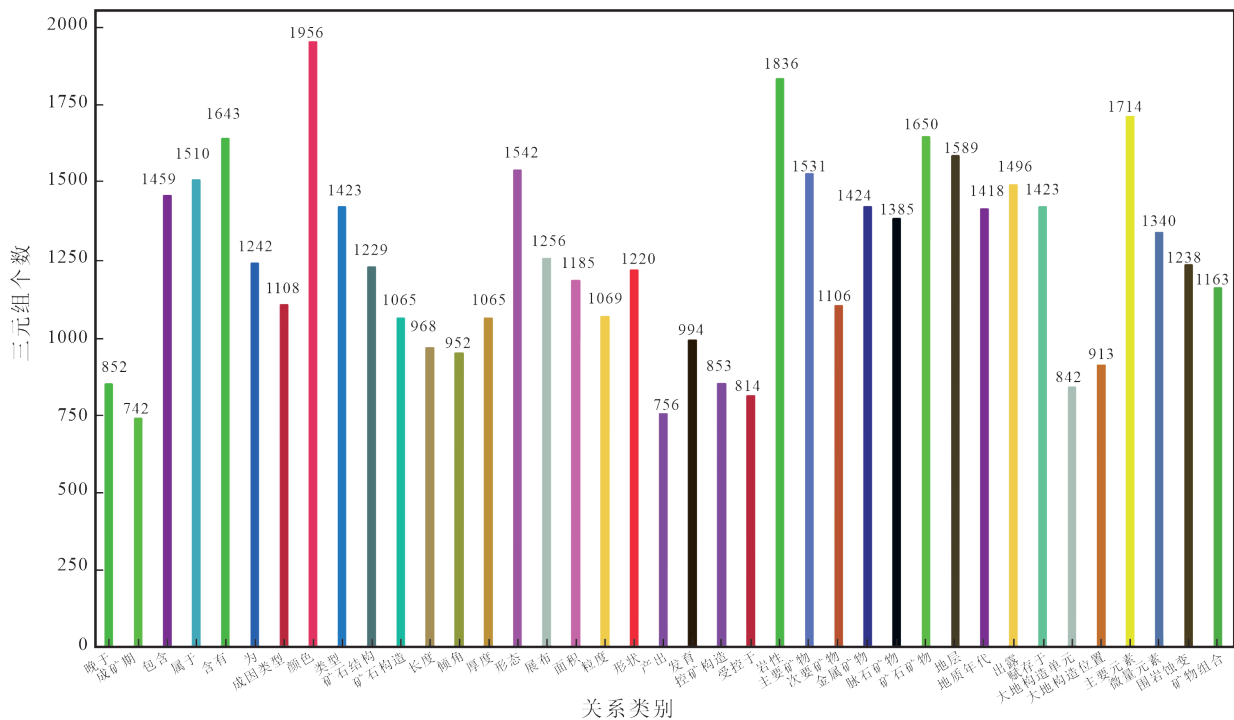


图4 实体关系三元组数量统计

Fig.4 Statistics for the number of entity relationship triples

表 8 提示模板使用前后结果对比

句 子	Deepseek-reasoner模型抽取结果	提示模板+ Deepseek-reasoner模型抽取结果
大哈拉军山组火山岩广泛分布于别珍套山没斯吾达坂南侧、阿吾拉勒山南侧	(大哈拉军山组火山岩,分布为,别珍套山没斯吾达坂南侧、阿吾拉勒山南侧)	(大哈拉军山组火山岩,位于,别珍套山没斯吾达坂南侧) (大哈拉军山组火山岩,位于,别珍套山没阿吾拉勒山南侧)
小热泉子组位于含化石(化石组合的时代为早石炭世)的下石炭统雅满苏组之下,二者为整合接触	(小热泉子,位于,含化石)	(下石炭统雅满苏组,地质年代,早石炭世)
黄山东基性-超基性岩体的主要岩石类型为闪长岩、角闪辉长岩、辉长岩、辉长苏长岩、苏长岩、角闪橄榄岩和二辉橄榄岩	(黄山东基性-超基性岩体,包含,闪长岩、角闪辉长岩、辉长岩、辉长苏长岩、苏长岩、角闪橄榄岩、二辉橄榄岩)	(黄山东基性-超基性岩体,包含,闪长岩); (黄山东基性-超基性岩体,包含,角闪辉长岩); (黄山东基性-超基性岩体,包含,辉长苏长岩); (黄山东基性-超基性岩体,包含,辉长岩); (黄山东基性-超基性岩体,包含,苏长岩); (黄山东基性-超基性岩体,包含,角闪橄榄岩); (黄山东基性-超基性岩体,包含,二辉橄榄岩)
张家口期火山岩系直接覆盖在古老基底或晚古生代花岗岩之上,火山岩呈大面积出露,火山喷发强度非常大,以强烈喷发的火山灰流相为主	(火山岩,出露,火山灰流相)	(张家口期火山岩系,出露,火山岩)

查询,以公婆泉铜矿床和黄山铜镍矿床进行展示(表9、10)。其中Cypher句法主要由4个不同的部分组成:Start表示查找图形中的起始节点;Match表示匹配图形模式,可以定位到感兴趣数据的子图形;Where表示过滤条件,可以基于某些标准过滤数据;Return表示返回的结果。

如表8和表9所示,通过基于知识图谱的智能问答可以快速了解矿床的矿物元素、岩石、地层、大地构造

位置、围岩蚀变、金属矿物、脉石矿物、地质构造、地质年代和矿床形状等。其中,矿物元素分布作为矿床组成的基本成分,不同的矿床具有独特的元素分布。金属矿物和脉石矿物则是矿床识别的重要指标,通过识别金属矿物的分布和富集情况可以初步评估矿床的经济潜力,而脉石矿物的分布情况则有助于判断矿床的成因、结构和演化过程,为矿床的进一步勘探提供线索。这些地质特征有助于分析矿床的分布规律,更精

表 10 基于东天山-北山找矿知识图谱的黄山铜镍矿床智能查询结果

Table 10 Intelligent question answering results for the Huangshan copper-nickel deposit based on the knowledge graph for the exploration of mineral resources in the East Tianshan–North Mountain region

问题	可视化结果	问题	可视化结果
黄铜镍矿床的矿物元素有哪些		黄铜镍矿床的地层有哪些	
黄铜镍矿床的大地构造位置是什么		黄铜镍矿床的金属矿物有哪些	
黄铜镍矿床的地质年代是什么		黄铜镍矿床的脉石矿物有哪些	
黄铜镍矿床的地质构造有哪些		黄铜镍矿床的围岩蚀变有哪些	
黄铜镍矿床的岩性有哪些		黄铜镍矿床的矿床形状有哪些	

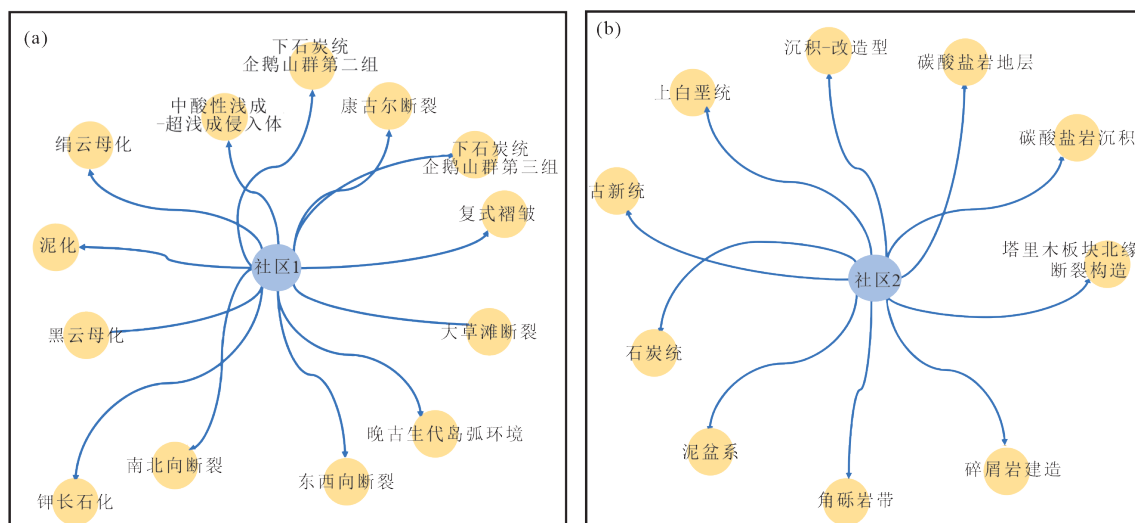


图6 基于Louvain社区发现算法可视化结果(部分)

Fig.6 Visualization results based on the Louvain community detection algorithm (Part)

空间中,然后利用DBSCAN聚类算法对嵌入表示后的向量进行聚类,结果见图7。

如图7所示,上述聚类较为集中,同簇节点之间的聚合性较强,不同簇之间的边界较为明显。其中对于粉色社区主要为矿床类型与地质构造单元,如多金属矿床、金矿床、钼矿床、康古尔-黄山韧性剪切带等。其中剪切带作为一种重要的构造单元,其在成矿过程中具有重要的作用,如构造控制、热液活动和蚀变作用等,通过分析矿床类型与剪切带的关联性,可以快速识别得到具有成矿潜力的剪切带区域,从而缩小范围,提高找矿效率。绿色社区主要指向“构造-岩浆-成矿”三位一体的成矿机制,形成了东天山成矿系统的核心控制因素。褐色社区主要包含了中-新元古界、中生界、古生界等地质年代,岩石组合主要包含细碧岩、火山碎屑岩、阳起石绿帘片岩、花岗岩等,该聚类反映了整个区域存在“早期洋壳-中期岩浆-晚期热液”的完整的演化序列,揭示该区域具备多类型成矿系统叠加发育的潜力。青色社区主要揭示了白山矿区成矿花岗岩的深部成因机制,从区域背景、矿化类型等角度进一步分析了在此构造位置形成大型斑岩钼矿的成因机制。紫色社区主要揭示了印支期造山后伸展环境下的壳幔相互作用的成矿系统,是对白山钼矿成因的构造背景的补

充。蓝色社区中主要包含了“镁铁质矿物”“橄榄岩相”“辉石岩”“岩浆熔离作用”等关键信息,而“黄铜矿+磁黄铁矿+辉石矿”是铜镍矿床的标志性矿物组合,其主要揭示了岩浆铜镍硫化物“构造-岩浆-成矿系统”的完整记录。红色社区为岩石地层,岩石地层为一个社区表明这些地层在成因上具有一定的相似性,且处于相似的构造背景中,该聚类结果表明这些地层具有相似的成矿环境。黄色社区中主要包含了I型花岗岩、花岗斑岩等,主要的矿物组合有辉钼矿和磁铁矿组合,且显示为白山地区,半自形-他形结构,其重点显示了斑岩-热液铜钼成矿系列。

5 结论

本文聚焦于东天山-北山成矿带这一研究区域,系统性地提出了一条基于“矿床本体构建-地质实体关系抽取-知识图谱构建-知识图谱应用”完整的技术路线。首先,基于现有的标准规范与专家经验,自上而下开展了矿床本体概念定义、属性及关系设计。其次,本文以东天山-北山区域地质文本数据为数据源,提出了一种基于大语言模型和提示模板相结合的地质实体关系抽取方法,构建形成东天山-北山的找矿知识图谱。第三,在上述知识图谱的基础上,本文从智能问答、社

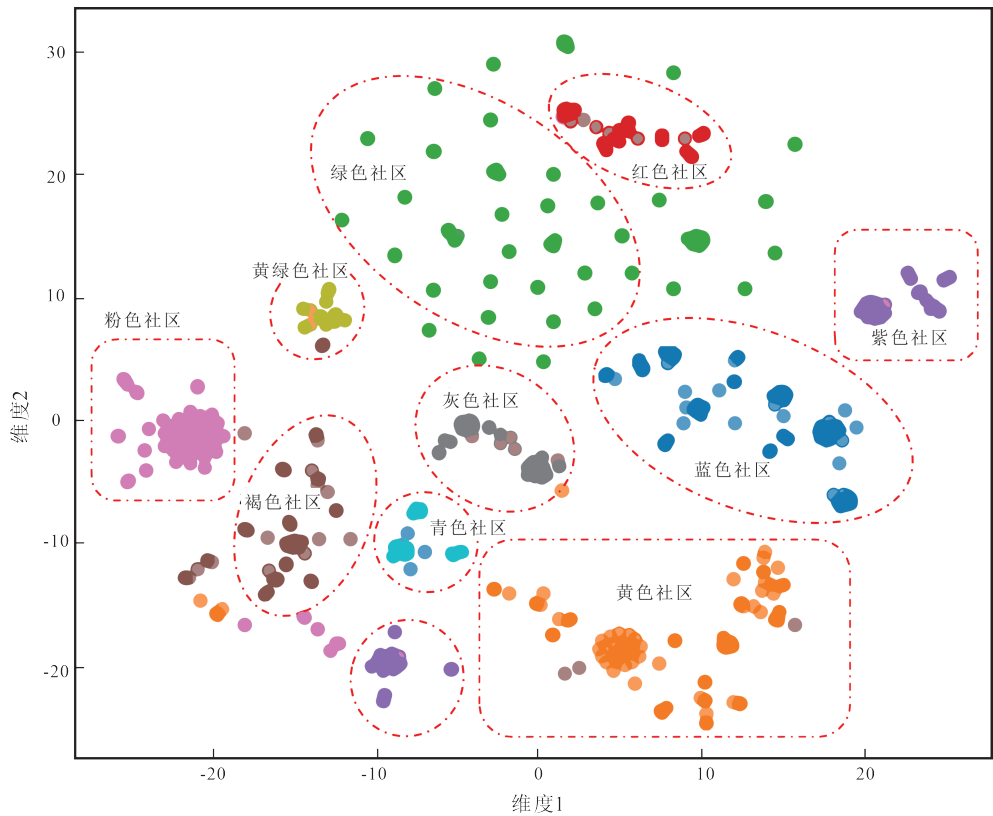


图7 找矿知识图谱聚类分析

Fig.7 Cluster analysis of the mineral exploration knowledge graph

区发现,聚类分析三个维度深度挖掘地质成矿规律,进一步探索东天山-北山成矿带的潜在成矿机制,从而进一步为区域地质勘探提供可解释、可扩展的知识支撑。

未来将进一步在现有找矿知识图谱构建的基础上,面向智能矿产资源预测应用场景,深入挖掘矿床演化的时空分布规则。通过总结这些规律进一步揭示矿床形成与分布的内在逻辑,同时将进一步利用显式的知识路径来解释大模型的推理过程,使模型的预测过程更加透明与可理解。通过领域知识与大模型推理的深入结合,显著提升矿产资源预测的精度与可靠性,为智能矿产资源预测服务提供更强的理论支持和技术保障,推动矿产资源勘查效率和可持续发展水平的提升。

作者贡献声明: 田苗,论文构思与撰写、数据收集、图件绘制、实验测试;余吉远,论文构思、数据收集、实验测试;邱芹军,论文构思、修改、数据收集;田龙星、谢忠、陈建国,论文构思、修改;吴麒瑞,数据收集和处理。

利益冲突声明: 作者保证本文无利益冲突。

致谢: 感谢周永章老师组织专题,以及审稿人提出的建设性意见。

参考文献 (References):

- 曹胜桃, 胡瑞忠, 周永章, 刘建中, 谭亲平, 高伟, 郑禄林, 郑禄璟, 宋威方. 2024. 基于大数据关联规则算法的卡林型金矿床元素富集规律及找矿方法研究. 地学前缘, (4): 58–72 [Cao S T, Hu R Z, Zhou Y Z, Liu J Z, Tan Q P, Gao W, Zheng L L, Zheng L J, Song W F. 2024. Element enrichment pattern and prospecting method for Carlin-type gold deposits based on big data association rule algorithm. Earth Science Frontiers, (4): 58–72 (in Chinese with English abstract)]
- 陈建平, 周冠云, 褚志远, 赵婕. 2024. 山东蒙阴金刚石矿床地质大数据三维建模与成矿预测. 矿床地质, 43(4): 802–820 [Chen J P, Zhou G Y, Chu Z Y, Zhao J. 2024. Geological big data three-dimensional modelling and mineralization prediction of diamond deposit in Mengyin, Shandong, China. Mineral Deposits, 43(4): 802–820 (in Chinese with English abstract)]
- 何陆灏, 周永章, 张灿. 2025. 基于深度学习的目标分割在岩石智能识别上的应用. 矿物岩石地球化学通报, 44(3): 525–541 [He L H, Zhou Y Z, Zhang C. 2025. Application of deep learning-based object segmentation in intelligent rock recognition. Bulletin of Mineralogy, Petrology and Geochemistry, 44(3): 525–541 (in Chinese with English abstract)]
- 贾晨辉. 2020. 新疆东天山典型铜镍硫化物矿床地质特征及成矿条件分析. (硕士学位论文) 西安: 长安大学 [Jia C H. 2020. Geological characteristics and metallogenic conditions of typical Cu-Ni sulfide deposits in East Tianshan, Xinjiang (Dissertation). Xi'an: Chang'an University, Xi'an, China (in Chinese with English abstract)]
- 龙灵利, 张志超, 王书来, 阴元军, 王新雨, 何西恒, 张延东, 卢正华, 肖国雄,

- 黄波. 2025. 新疆东天山土屋斑岩型铜矿床矿脉分带规律及找矿标志. 矿产勘查, 16(5): 979–989 [Long L L, Zhang Z C, Wang S L, Yin Y J, Wang X Y, He X H, Zhang Y D, Lu Z H, Xiao G X, Huang B. 2025. Ore vein zoning patterns and prospecting indicators of the Tuwu porphyry copper deposit, East Tianshan, Xinjiang. Mineral Exploration, 16(5): 979–989 (in Chinese with English abstract)]
- Mantovani A, Piana F, Lombardo V. 2020. Ontology-driven representation of knowledge for geological maps. Computers & Geosciences, 139: 104446
- Pei Y, Chai S L, Li X L, Samuel J C, Ma C Y, Chen H N, Lou R X, Gao Y. 2022. Construction and application of a knowledge graph for gold deposits in the Jiapigou gold metallogenic belt, Jilin Province, China. Minerals, 12(9): 1173
- Qiu Q J, Jin X G, Tian M, Wu Q R, Tao L F, Chen J G, Xie Z. 2025. Joint extraction method for entity relations from mineral resources reports integrating dependency parsing and improved graph convolutional networks. Ore Geology Reviews, 182: 106640
- Qiu Q J, Ma K, Lv H R, Tao L F, Xie Z. 2023a. Construction and application of a knowledge graph for iron deposits using text mining analytics and a deep learning algorithm. Mathematical Geosciences, 55(3): 423–456
- Qiu Q J, Tian M, Tao L F, Xie Z, Ma K. 2024. Semantic information extraction and search of mineral exploration data using text mining and deep learning methods. Ore Geology Reviews, 165: 105863
- Qiu Q J, Tian M, Xie Z, Tan Y J, Ma K, Wang Q F, Pan S Y, Tao L F. 2023b. Extracting named entity using entity labeling in geological text using deep learning approach. Journal of Earth Science, 34(5): 1406–1417
- Tian M, Xie Z, Qiu Q J, Wu Q R, Chen J G, Duan Y X, Tao L F. 2025. PMinrKG: Polymetallic mineral resources knowledge graph construction and its applications integrated with multimodal data International Journal of Digital Earth, 18(1): 2494285
- 邱芹军, 田苗, 吴麒瑞, 陈建国, 诸云强, 陈占龙, 谢忠. 2024. 基于多源异构数据的地质知识图谱构建与应用. 地学前缘, 1–17. <https://link.cnki.net/doi/10.13745/j.esf.sf.2024.11.69> [Qiu Q J, Tian M, Wu Q R, Chen J G, Zhu Y Q, Chen Z L, Xie Z. 2024. Construction and application of geological knowledge graph based on multi-source heterogeneous data. Earth Science Frontiers, 1–17. <https://link.cnki.net/doi/10.13745/j.esf.sf.2024.11.69> (in Chinese with English abstract)]
- 汪陈. 2021. 面向中文文献的金矿时空属性信息抽取及知识图谱可视化表达(硕士学位论文). 合肥: 合肥工业大学 [Wang C. 2021. Extraction of spatiotemporal attributes information of gold mines and visual expression of knowledge graphs for Chinese literature (Dissertation). Hefei: Hefei University of Technology (in Chinese with English abstract)]
- 王翔, 李景朝, 陈辉, 茹湘兰, 范海明, 郑啸, 梁婉娟. 2015. 大数据与地质资料信息服务: 需求、产品、技术、共享. 地质通报, 34(7): 1309–1315 [Wang X, Li J C, Chen H, Ru X L, Fan H M, Zheng X, Liang W J. 2015. Big and geological data information services. Geological Bulletin of China, 34(7): 1309–1315 (in Chinese with English abstract)]
- 魏东琦, 江宝得, 张静雅. 2021. 非结构化地质数据内容存储方法研究. 西北地质, 54(4): 266–273 [Wei D Q, Jiang B D, Zhang J Y. 2021. Research on content storage method of unstructured geological data. Northwestern Geology, 54(4): 266–273 (in Chinese with English abstract)]
- 吴麒瑞, 田苗, 谢忠, 邱芹军, 陈占龙, 陶留锋. 2025. 融合多模态数据的地震灾害知识图谱构建及应用. 地质科技通报, 44(4): 90–106 [Wu Q R,

- Tian M, Xie Z, Qiu Q J, Chen Z L, Tao L F. 2025. Construction and application of earthquake disaster knowledge graph fusing with multi-modal data. *Bulletin of Geological Science and Technology*, 44(4): 90–106 (in Chinese with English abstract)]
- 张春菊, 刘文聪, 张雪英, 叶鹏, 汪陈, 朱少楠, 张达玉. 2023. 基于本体的金矿知识图谱构建方法. *地球信息科学学报*, 25(7): 1269–1281 [Zhang C J, Liu W C, Zhang X Y, Ye P, Wang C, Zhu S N, Zhang D Y. 2023. Knowledge graph construction method of gold mine based on ontology. *Journal of Geo-Information Science*, 25(7): 1269–1281 (in Chinese with English abstract)]
- 张前龙, 周永章, 郭兰萱, 原桂强, 虞鹏鹏, 王汉雨, 朱彪彪, 韩枫, 龙师尧. 2024a. 找矿知识图谱的智能化应用: 以钦杭成矿带斑岩铜矿为例. *地学前缘*, 31(4): 7–15 [Zhang Q L, Zhou Y Z, Guo L X, Yuan G Q, Yu P P, Wang H Y, Zhu B B, Han F, Long S Y. 2024. Intelligent application of knowledge graphs in mineral prospecting: A case study of porphyry copper deposits in the Qin-Hang metallogenic belt. *Earth Science Frontiers*, 31(4): 7–15 (in Chinese with English abstract)]
- 张前龙, 周永章, 虞鹏鹏, 王汉雨, 韩枫, 贺炬翔. 多层次矿床本体的构建及在知识图谱中的应用. *矿物岩石地球化学通报*, 2024b, 43 (01): 211–217 [Zhang Q L, Zhou Y Z, Yu P P, Wang H Y, Han F, He J X. 2024b. Construction of a multi-level deposit ontology and its application in knowledge graphs[J]. *Bulletin of Mineralogy, Petrology and Geochemistry*, 43(1): 211–217 (in Chinese with English abstract)]
- 张振亮, 冯选洁, 赵国斌, 谭文娟. 2022. 东天山—北山地区中生代内生矿床成矿规律. *西北地质*, 55(4): 280–299 [Zhang Z L, Feng X J, Zhao G B, Tan W J. 2022. The metallogenic rule of Mesozoic hypogene deposits in the eastern Tianshan-Beishan area. *Northwestern Geology*, 55(4): 280–299 (in Chinese with English abstract)]
- 周波, 李珂. 2025. 基于GeoGPT与LightRAG的地质找矿知识图谱及知识问答模型构建方法研究. *地质通报*, [https://kns.cnki.net/KCMS/detail/de-](https://kns.cnki.net/KCMS/detail/detail.aspx?filename=ZQYD20251009002&dbname=CJFD&dbcode=CJFQ)tail.aspx?filename=ZQYD20251009002&dbname=CJFD&dbcode=CJFQ [Zhou B, Li K. 2025. A methodology for constructing a geological prospecting knowledge graph and question answering model based on GeoGPT and LightRAG. *Geological Bulletin of China*, <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=ZQYD20251009002&dbname=CJFD&dbcode=CJFQ> (in Chinese with English abstract)]
- 周冠云, 朱成河, 高存山, 褚志远, 肖克炎, 王江霞, 陈建平. 2023. 大数据框架下金刚石地质成矿信息挖掘及应用. *地球学报*, 44(5): 909–915 [Zhou G Y, Zhu C H, Gao C S, Chu Z Y, Xiao K Y, Wang J X, Chen J P. 2023. Mining and application of diamond geological metallogenic information under the framework of big data. *Acta Geoscientica Sinica*, 44(5): 909–915 (in Chinese with English abstract)]
- 周军, 杨万志, 彭峰, 刘亚楠, 魏宇鹏, 梁乐. 2017. 新疆东天山雅满苏北山地区铜镍找矿潜力分析. *新疆地质*, 35(2): 157–162 [Zhou J, Yang W Z, Peng F, Liu Y N, Wei Y P, Liang L. 2017. Prospecting potential Analysis of YaManSuBeishan area copper-nickel deposit in Eastern Tianshan, Xinjiang. *Xinjiang Geology*, 35(2): 157–162 (in Chinese with English abstract)]
- 周永章, 肖凡. 2024. 管窥人工智能与大数据地球科学研究新进展. *地学前缘*, 31(4): 1–6 [Zhou Y Z, Xiao F. 2024. Overview: A glimpse of the latest advances in artificial intelligence and big data geoscience research. *Earth Science Frontiers*, 31(4): 1–6 (in Chinese with English abstract)]
- 周仲礼, 柳炳利, 龚成杰, 曹昌杰, 孔韞辉, 李程, 但诗瑶, 王政尧. 2025. 矿产勘查知识图谱: 研究进展、关键问题与未来展望. *成都理工大学学报(自然科学版)*, 52(5): 827–843 [Zhou Z L, Liu B L, Gong C J, Cao C J, Kong Y H, Li C, Dan S Y, Wang Z Y. 2025. Knowledge graphs for mineral exploration: Research progress, key challenges, and future perspectives. *Journal of Chengdu University of Technology (Science & Technology Edition)*, 52(5): 827–843 (in Chinese with English abstract)]