

融合知识图谱与大语言模型的地学知识抽取与信息挖掘——以卡林型金矿为例

刘国庆^{1,2}, 陈国雄^{1*}

1. 中国地质大学（武汉），地质过程与成矿预测全国重点实验室，湖北武汉 430074
2. 中国地质大学（武汉）地球与行星科学学院，湖北武汉 430074

摘要: 针对地质勘查领域海量非结构化数据难以被有效利用以及通用大模型存在“事实幻觉”与专业逻辑匮乏等问题，本文提出一种融合知识图谱（KG）与检索增强生成（RAG）的垂直领域智能知识挖掘框架，并以中国黔西南与美国内华达地区的卡林型金矿成矿规律总结和对比研究为例进行了验证。首先，构建了基于本地化部署 DeepSeek-32B 的 RAG 智能问答系统，通过向量检索与生成式阅读理解，实现了专业知识的精准溯源与高可信问答。其次，利用大模型监督微调（SFT）技术，从数百份多源异构地质资料中高效构建了系统涵盖地层构造、蚀变矿物及控矿要素的跨区域成矿知识图谱。实验结果表明，该系统在客观准确性上显著优于 GPT-4o，在主观生成上具备高忠实度与完全可溯源性，有效解决了幻觉问题。基于图谱拓扑学的分析不仅定量揭示了两地成矿的宏观异同，还量化了从矿石实体、蚀变组合到地球化学元素异常的级联指示路径，证实了其发现隐性找矿线索的能力。该研究实现了从非结构化文本到结构化知识的智能转化与深度挖掘，为解决地学领域“海量数据、知识饥饿”困境提供了新的技术路径。

关键词: 卡林型金矿；大语言模型；知识图谱；检索增强生成；知识抽取

中图分类号: P628; P618.51

收稿日期: 2025-12-30

基金项目: 国家深地重大专项青年科学家课题（2024ZD10019007），贵州省地质矿产局地质科研项目(黔地质科合〔2025〕01号)，中央高校基本科研业务费专项资金资助项目（GUG-DMX2025-01），国家级大学生创新训练计划资助项目(202510491034)

作者简介 刘国庆（2004-），男，本科生，主要从事大数据找矿研究工作，ORCID:0009-0003-1518-6258，E-mail:liuguoping@cug.edu.cn.

***通讯作者** 陈国雄(1988-),男,研究员,主要从事数学地球科学领域教学科研,ORCID: 0000-0002-6785-9675, E-mail:gxchen@cug.edu.cn

Geological Knowledge Extraction and Information Mining via the Fusion of Knowledge Graphs and LLMs: A Case Study of Carlin-type Gold Deposits

Guoqing Liu^{1,2}, Guoxiong Chen^{1*}

1 State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China

2. School of Earth and Planetary Sciences, China University of Geosciences, Wuhan 430074, China

Abstract: To address the challenges in effectively utilizing massive unstructured data within geological exploration and the issues of hallucination and lack of specialized logic in general large language models (LLMs), we propose an intelligent knowledge mining framework for vertical domains that integrates Knowledge Graph (KG) and Retrieval-Augmented Generation (RAG). This framework is validated through a case study of Carlin-type gold deposits in the Southwest Guizhou, China, and Nevada, USA. Firstly, a RAG-based intelligent question-answering system was constructed using a locally deployed DeepSeek-32B model. Through vector retrieval and generative reading comprehension, the system achieves precise traceability of professional knowledge and highly reliable Q&A. Secondly, leveraging Supervised Fine-Tuning (SFT) techniques on the LLM, a cross-regional metallogenic knowledge graph systematically covering stratigraphy, structure, alteration minerals, and ore-controlling factors was efficiently built from hundreds of multi-sources, heterogeneous geological documents. Experimental results demonstrate that the proposed system significantly outperforms GPT-4o in terms of objective accuracy. For subjective content generation, it exhibits high faithfulness and full traceability, effectively mitigating the hallucination problem. Analyses based on graph topology not only quantitatively reveal the macroscopic similarities and differences in mineralization between the two regions but also quantify the cascading indicative pathways—from orebody entities and alteration assemblages to geochemical element anomalies. This confirms the system's capability to discover implicit clues for mineral exploration. This study realizes the intelligent transformation and in-depth mining of knowledge from unstructured text to structured representations. It offers a novel technical pathway to address the dilemma of "data-rich yet knowledge-poor" prevalent in the geoscience domain.

Key words: Carlin-type Gold Deposit; Large Language Model (LLM); Knowledge Graph; Retrieval-Augmented Generation (RAG); Knowledge Extraction

0 引言

以大数据与人工智能为代表的信息技术迅速迭代,正驱动当前科学研究范式发生深刻变革。地质学作为典型的数据密集型科学,地学数据的收集、更新、维护以及有效挖掘与分析已成为该领域关注和研究热点(赵鹏大, 2019; 周永章等, 2021a)。然而,地质学家长期积累的宝贵经验与智慧,大量以非结构化形态沉淀于文献、报告和图件之中,导致了“海量数据、知识饥饿”等困境(成秋明, 2025)。面对这一困境,目前获取知识的主要途径仍高度依赖于对海量文献的手动梳理与查阅。传统信息检索方法高度依赖关键词匹配,其在语义覆盖、推理能力及结果可解释性方面的固有缺陷,使其难以从复杂语料中自动提炼关键知识、识别隐含关系并进行逻辑推演,无法满足当前对精准知识挖掘与智能推理日益增长的需要(周永章等, 2021b; 董宇浩等, 2025; 冯婷婷等, 2025)。

与此同时,大语言模型(Large Language Models, LLMs)在语义理解、文本生成与智能问答等方面展现出了卓越能力,并正加速渗透至金融、医疗、法律等专业领域(Raiaan et al., 2024)。相关研究已开始探索 LLMs 在地学知识理解与问题解决中的应用潜力(Hu et al., 2023; Zhang et al., 2024)。尽管如此,在面对地质勘查与矿产资源研究等高度专业化的应用场景时,通用大语言模型仍面临严峻挑战。一方面,地质领域知识体系结构复杂、术语系统庞大且高度异构,通用模型难以准确解析专业概念及其层级关系;另一方面,由于缺乏面向特定任务的深层地质逻辑推理能力,模型生成内容常出现“事实幻觉”、语义漂移及逻辑跳跃等问题,难以满足精细化知识获取与深层次智能问答的实际需求(师路易和左仁广, 2025; 张宝一等, 2025)。

为弥补通用模型在专业性、时效性与可靠性方面的局限,检索增强生成(Retrieval-Augmented Generation, RAG)技术为构建高效、可信的领域知识服务系统提供了新途径(Lewis et al., 2020)。RAG 框架通过将外部知识库与语言生成模块动态融合,显著提升了模型对上下文信息的整合与利用能力,有效缓解了“模型幻觉”现象。然而,RAG 框架主要依赖向量空间的隐式语义匹配,通常局限于碎片化的文本片段获取,难以捕捉地质实体间跨文档的复杂关联。知识图谱(Knowledge Graph, KG)作为一种结构化的语义网络,能够以“实体-关系-实体”的形式清晰表达领域概念及其复杂的内在联系,是实现海量数据深度挖掘与宏观

成矿规律总结的关键工具(Ma, 2022; 张宝一等, 2025)。在地学领域, 融合知识图谱的智能系统已取得初步进展。例如, Fu et al. (2025) 开发的 GeoMinLM 系统, 通过整合专家知识图谱减少了模型幻觉, 提升了矿产勘探问答的专业性; 李博文等 (2025) 围绕地质领域复杂语义知识的组织与检索需求, 基于 GraphRAG 和 LLMs, 构建了一个面向胶东金矿的智能搜索系统。

然而, 构建高质量的知识图谱依赖于高效的知识抽取(Knowledge Extraction, KE)技术(邱芹军等, 2024)。KE 技术已从早期的规则与统计方法, 发展到以 BERT-BiLSTM-CRF 为代表的深度学习模型; 但传统深度学习对高质量标注数据依赖性强, 泛化能力弱, 且在处理复杂关系时能力受限(彭晶晶和林锴, 2024; 王成彬等, 2024; 冯婷婷等, 2025)。近年来, 大语言模型(LLMs)在处理海量非结构化文本及地学知识抽取方面引起了广泛关注, 展现出显著的应用潜力(Sutanto et al., 2024), 但通用模型在零样本或少样本条件下, 往往难以严格遵循复杂的地质图谱模式定义, 常出现实体类型混淆或输出格式不规范等问题(Li et al., 2023; Zhang et al., 2025)。鉴于此, 监督微调(Supervised Fine-Tuning, SFT)技术成为了破解这一瓶颈的关键。SFT 能够通过少量高质量的指令数据, 将地质专家的抽取逻辑“注入”模型参数中, 使其在保持通用语义理解能力的同时, 具备了严格的指令遵循与结构化输出能力, 从而为快速、低成本地构建高精度领域知识图谱提供了新的范式(Xin et al., 2025; Zhao et al., 2025)。

卡林型金矿是全球最重要的金矿类型之一, 尤以美国内华达地区和中国右江盆地最为典型(谭仕敏等, 2007; 谢卓君等, 2019)。这两大矿集区在地质特征上表现出显著的相似性, 例如金均以晶格金形式赋存于含砷黄铁矿中, 且矿床定位于大陆地壳边缘的造山后伸展环境, 受控于背斜、单斜及伸展断层构造(Wang and Groves, 2018)。然而, 针对中国右江盆地卡林型金矿的分类与成因至今仍存在广泛争议: 部分学者认为其属于典型的“卡林型(Carlin-type)”, 而另一部分观点则将其归为“类卡林型(Carlin-like)”甚至“造山型(Orogenic)”金矿(Deng and Wang, 2016; Pi et al., 2017)。这些分类上的争议, 本质上源于对成矿流体来源(大气降水、岩浆水或变质水)及成矿物理化学环境认识的显著分歧(Hu et al., 2017; Goldfarb et al., 2019)。要厘清这些争议, 必须采用多尺度的对比分析方法, 从矿集区尺度到矿床尺度对中美两大矿集区进行系统对比。然而, 针对跨尺度对比这

一挑战, 现有的解决方案往往局限于定性描述, 难以从海量混杂信息中系统识别出能够有效区分不同矿床亚类的关键判别指标。此外, 支撑这种对比的海量地质数据, 目前正面临着严重的知识壁垒, 现有的研究体系仍呈现出高度的碎片化特征。一方面, 核心研究成果被割裂在中英文不同的语言体系与数据库中, 缺乏统一的语义标准与共享平台; 另一方面, 目前的研究正处于从现象描述向系统建模与智能预测跨越的转型期(Cao et al., 2025)。传统的人工文献梳理已无法满足对海量、多源、跨语言数据的深度挖掘需求, 因此亟需构建一种结合大语言模型与知识图谱的新型智能系统, 在统一的语义框架下对中美两大矿集区的信息进行结构化重组与推理。

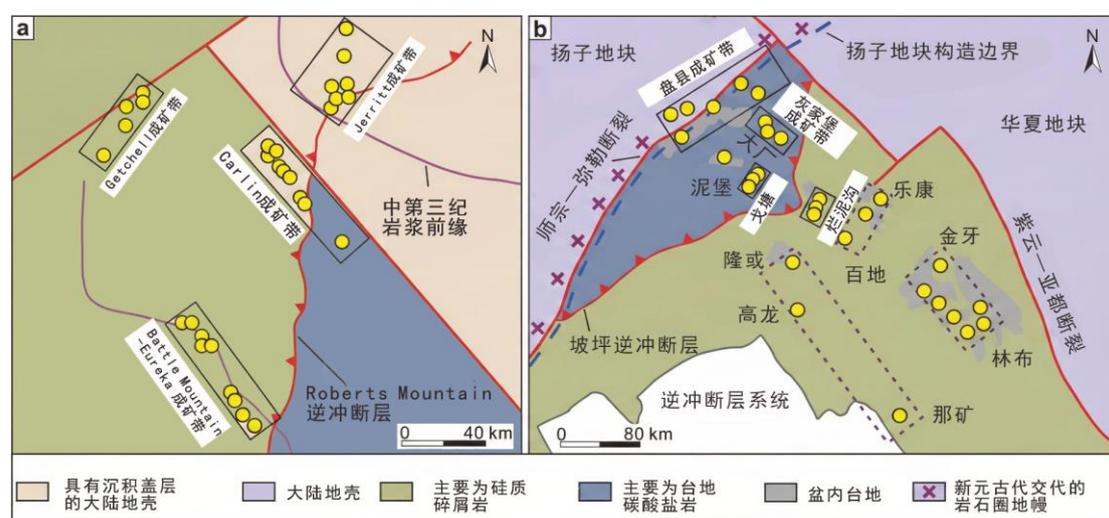


图 1 内华达地区 (a) 与右江盆地 (b) 卡林型金矿床分布(据 Wang et al.,2018 修改)

Fig. 1 Distribution of Carlin-type gold deposits in (a) Nevada and (b) the Youjiang Basin
(modified from Wang et al., 2018)

鉴于此, 本文拟结合大语言模型与知识图谱技术, 旨在构建一个面向跨区域卡林型金矿领域的智能知识挖掘系统, 为矿产勘查领域的信息智能化提供新的理论基础和技术实现路径。本文的主要研究内容包括: (1) 构建一个高性能、高安全性的本地化垂直领域大语言模型, 以满足专业地学问答的需求; (2) 论证并采用基于大语言模型微调(SFT)的知识抽取方法, 构建首个覆盖黔西南-内华达两大矿集区的卡林型金矿知识图谱; (3) 依托该知识图谱开展多维度的智能挖掘, 对比分析两大矿集区的成矿规律异同, 并探索图谱中隐藏的知识关联。

1 研究方法

由于地质勘查数据中文本描述多、实体关系复杂，通用大模型容易产生幻觉且缺乏专业逻辑，而知识图谱虽然结构清晰，但难以覆盖非结构化的语义细节；为此，本文提出了一种融合大语言模型与知识图谱的“双驱动”知识抽取与信息挖掘架构。

该架构主要通过以下两种方式实现技术融合：（1）构建阶段，利用经指令微调(SFT)的垂直领域大模型作为实体和关系的抽取工具，来解决地质非结构化文本向知识图谱转化的难题。（2）应用阶段，建立基于 RAG 和 KG“显隐互补”的协同挖掘模式；其中，RAG 模块主要负责处理“隐性知识”，通过向量检索保留文献中的具体描述和上下文（如具体的蚀变特征描述），解决精准问答与溯源的问题；KG 模块主要负责处理“显性知识”，通过结构化网络揭示地质实体间的关联路径（如成矿要素的共生组合），解决规律分析的问题。两者结合能够实现从单一知识点的查询到复杂成矿规律的挖掘。

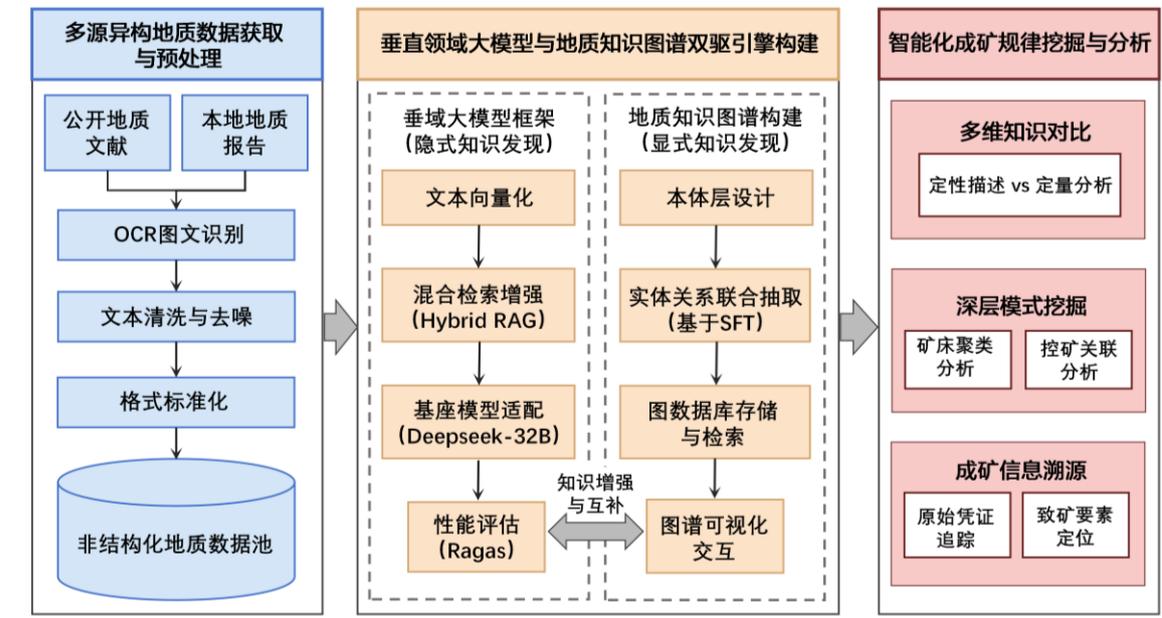


图 2 大语言模型与知识图谱双驱动的知识抽取与信息挖掘架构

Fig. 2 A dual-driven architecture for knowledge extraction and information mining based on large language models (LLMs) and knowledge graphs (KGs)

1.1 垂直领域大语言模型构建

为充分利用非结构化文献中的专家知识，并构建针对特定地质场景的专业问答系统，本项目选择本地化部署大语言模型方案。其必要性主要体现在：(1) 领域适配与系统可控性：本地化部署允许针对地质领域特有的多源异构数据（如各类公开地质报告、图件描述等）进行深度的参数微调与系统集成，避免了通用云端模型在处理特定专业语境时的局限性；(2) 成本优化：高频专业使用场景下，本地部署的边际成本低于按次计费的云 API。具体流程如下：

首先，针对黔西南卡林型金矿这一核心研究对象，从非结构化语料库中（详见 2.2.1 节）优选了近 200 余篇高质量核心学术文献，并利用 PDFMiner 等工具进行格式转换，构建了一个高质量、领域专注的本地专业知识库；之后在 RTX 4090D *2 的算力支持下，采用 RAGFlow 框架，该框架整合了文本解析、知识库构建、向量嵌入、大模型生成等多个环节；选用 Qwen3-Embedding-8B 作为高性能向量化模型，负责将知识库文本精准地转换为向量表示；采用 DeepSeek-R1-Distill-Qwen-32B 作为生成式大模型，通过 Ollama 框架进行本地化部署。构建了离线问答系统，整个系统在完全离线的环境中运行，用户输入专业问题后，系统首先将问题向量化，并在向量知识库中检索最相关的文本片段，然后将这些片段作为上下文与原始问题一同送入生成式大模型，最终生成精准、可靠且附带参考文献来源的回答，确保了知识的准确性和可追溯性。

1.2 跨区域知识图谱构建

为系统性地整合与对比全球两大代表性卡林型金矿集区的成矿规律，本文设计并实现了一套完整的跨区域知识图谱构建 workflow。该流程包括语料库构建、本体设计、知识抽取与图数据库存储四个核心阶段。

1.2.1 非结构化语料库构建

知识图谱的数据基础源于海量的非结构化文献。本文首先进行了系统性的文献搜集，共搜集整理了涵盖中国黔西南与美国内华达两个地区的学术论文、可公开获取的地质报告及专著等 405 份关键资料。为构建标准化的非结构化语料库，

本文对这批多源异构的资料进行了严格的预处理。首先，利用光学字符识别(OCR)技术处理扫描版报告与图件，将其转换为可编辑的文本格式；随后，通过文本提取与格式转换脚本，对所有文档进行统一的格式化清洗，去除无关信息（如目录索引、页眉、页脚、参考文献列表及乱码符号等），并规范化文本结构。这一流程确保了后续知识抽取所需语料的纯净度与规范性。

1.2.2 地学本体层设计

本体层是知识图谱的模式层，它定义了知识的组织结构和语义规则。为确保图谱的专业性和可扩展性，本文基于《矿床学》等教科书，并结合卡林型金矿的领域特点，自顶向下地设计了地学本体层架构（图3）。该架构共定义了9个核心实体类型，包括宏观的“构造单元”、“赋矿地层”，以及微观的“蚀变矿物”、“矿石矿物”等。同时，为描述实体间的复杂地质联系，设计了12个关键关系类型，如“位于”（空间关系）、“产于”（赋存关系）、“受...控制”（因果关系）和“包含”（组成关系）等。这套本体架构为后续的结构化知识抽取提供了清晰的框架和语义约束。

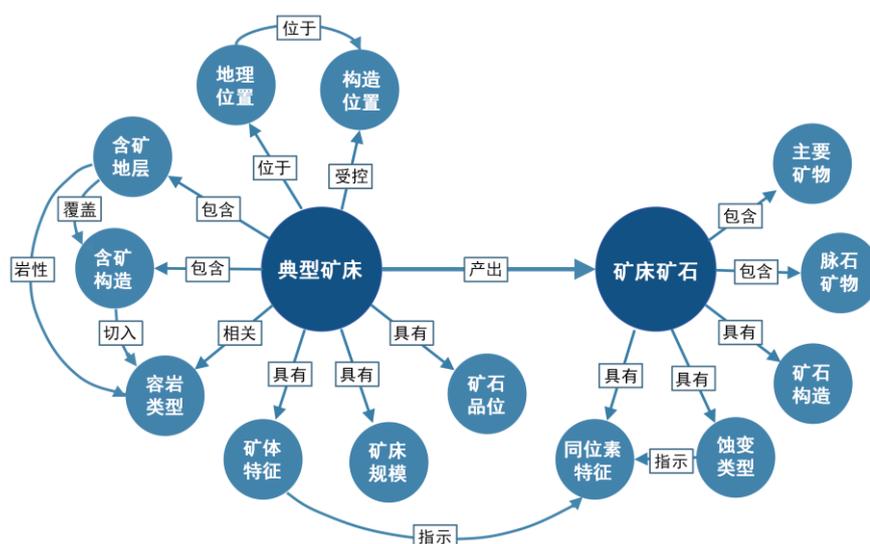


图 3 黔西南-内华达卡林型金矿知识图谱本体层架构. 注：图示中为视觉简洁，将连接不同属性实体（如规模、品位、特征等）的多种具体关系类型统一展示为通用标签“具有”。

Fig. 3 Ontology layer architecture of knowledge graph for Carlin-type gold deposits in SW Guizhou and Nevada

1.2.3 基于模型微调的知识抽取

知识抽取是图谱构建的核心技术环节，其技术路线的选择直接影响数据质量与构建效率。本文在模型选择阶段，对主流方案进行了系统性评估。以 BERT-BiLSTM-CRF 为代表的传统深度学习方案，虽在命名实体识别(NER)任务上表现成熟，但存在两个难以克服的局限；其一，标注成本高昂，模型严重依赖海量的字词级(Token-level)精细标注（如 BIO 体系），耗时费力；其二，跨语言迁移与对齐能力不足，传统模型缺乏统一的语义表征能力，不仅要求针对中英文文献分别构建训练集进行独立标注，且不同语言模型抽取出的实体与关系难以在同一语义空间下实现高效的对齐与融合，限制了图谱构建的整体连贯性(Xin et al., 2025; Zhao et al., 2025)。

鉴于传统方案的局限性，本文转向基于大语言模型的抽取路径。本文首先评估了未微调的基座大模型（如 Qwen2.5-7B (Base)），如后续性能评估（图 4）所示，基座模型在零样本条件下无法准确理解地质领域的特定抽取指令，表现为输出格式混乱、关键信息遗漏，F1 值极低。与此相对，基于大模型指令微调(SFT)的方案展现出显著的综合优势(Zhang and Soh, 2024)。该方案的数据构建成本相对可控，仅需构建“指令-答案”(Q&A)样本，远低于全量 token 级标注的成本(Jiang et al., 2024; Tian et al., 2024)。更重要的是，SFT 赋予了模型卓越的任务灵活性，使其能够精确理解指令并严格按照预定义的 JSON 格式输出结构化结果，高度契合了知识图谱构建的技术需求。因此，综合考量数据成本、模型灵活性与抽取性能，本文最终确定采用指令微调(SFT)作为知识抽取的技术路线。

首先，基于设计的本体层，利用 Easydataset 工具从语料库中人工构建了一个包含 1126 对高质量“指令-问题-答案”(Q&A)的微调数据集。该数据集（示例如图 5 所示）旨在覆盖所有预定义的实体与关系类型。接着，选用 Qwen2.5-7B 为基座模型，并利用 LLaMA-Factory 框架与上述 1126 条 Q&A 数据对模型进行指令微调。此过程旨在使模型能精准理解地质专业术语并严格执行结构化抽取任务。为定量评估 SFT 的有效性，本文从语料库中预留 10% 的文献作为测试集并进行人工标注。如图 4 所示，微调后的 Qwen2.5-7B 模型的抽取性能相较于基座模型取得了显著提升。最后，在性能验证后的 SFT 模型基础上，结合提示词工程，对全部 405 份预处理文献进行逐段自动化抽取，最终获取了“实体-关系-实

体”三元组知识。

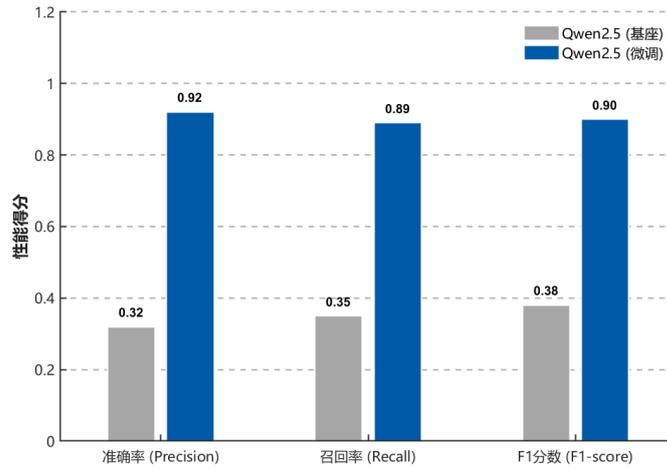


图 4 知识抽取模型微调前后性能对比

Fig. 4 Performance comparison of the knowledge extraction model before and after fine-tuning

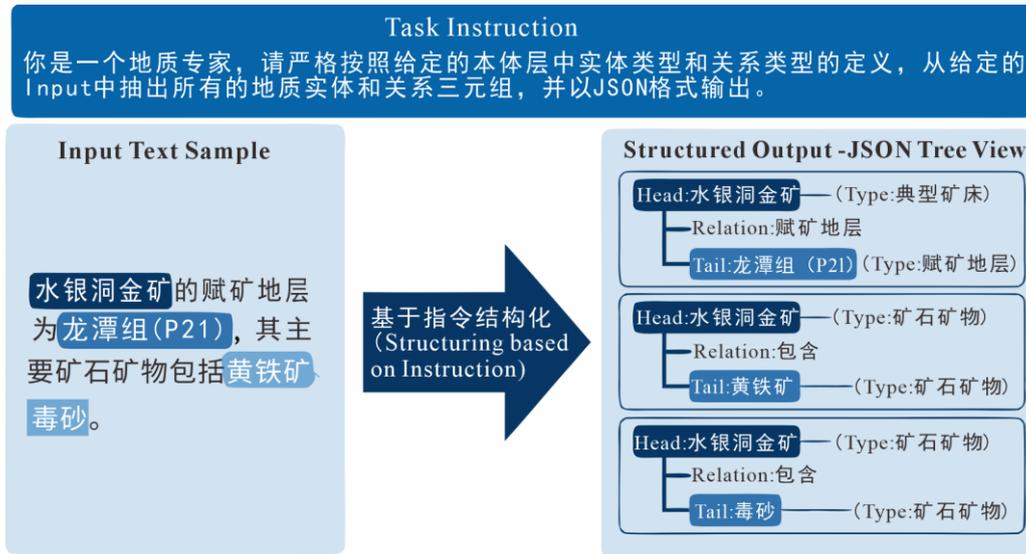


图 5 微调数据集 (Q&A 对) 示例

Fig. 5 Examples of fine-tuning dataset (Q&A pairs)

1.2.4 图数据库存储与可视化

自动化抽取的原始三元组数据（超过一万条）不可避免地存在冗余、歧义和不一致问题。为此，本文设计了一套清洗、消歧和合并的后处理程序，例如将“普安-马场断裂”与“普马断裂”等不同表述合并为同一规范化实体。还特别处理了中英双语信息；例如，对于地层、断层、元素等核心地质术语，系统以中文为基准进行归一化；对于同时包含英文名称的实体（如来自内华达地区的文献），

则将英文名称作为实体的别名或属性一并保留，以确保跨语言检索的准确性。经过处理后，最终成功构建了包含 6221 个核心实体和 11779 个有效关系的黔西南-内华达卡林型金矿知识图谱（图 6）。为实现图谱的有效管理和应用，本文将所有结构化数据存入的 Neo4j 图数据库中，并利用自研改进的 SmartKG 知识图谱管理平台，实现了图谱的可视化管理、多跳查询和智能检索。用户可通过该平台，直观地浏览两大矿集区的知识全貌，并进行深度的挖掘分析。

表 1 黔西南-内华达卡林型金矿知识图谱实体关系统计

Table 1 Statistics of entity relations in the knowledge graph for Carlin-type gold deposits in SW Guizhou and Nevada

区域	实体类型	实体个数	实体个数 (合并消歧后)	关系类型	关系个数 (合并消歧后)
黔西南	9	10212	5234	12	10256
内华达	8	2312	987	11	1523
合计	9	12524	6221	12	11779

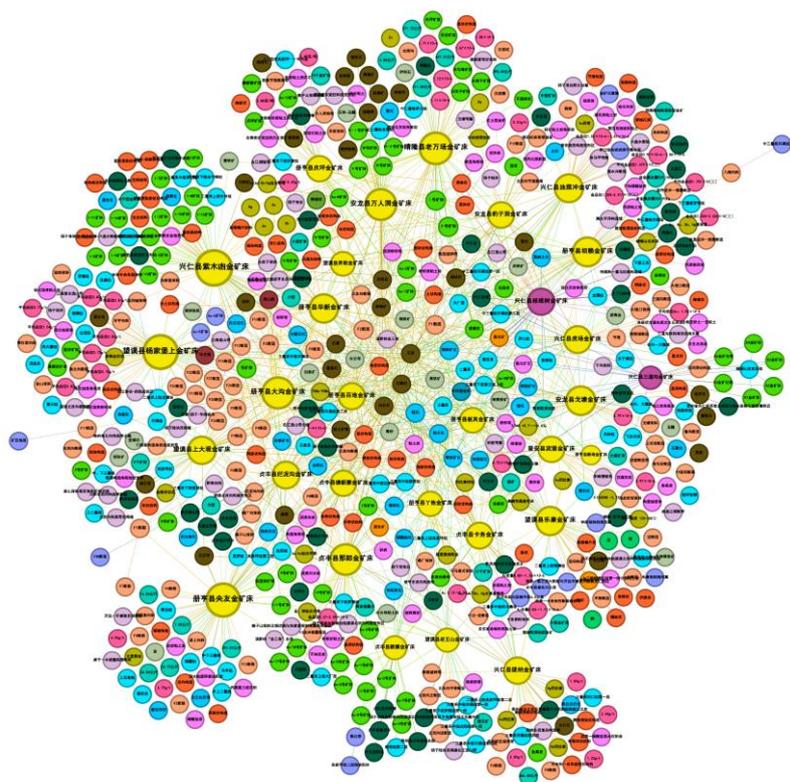


图 6 黔西南-内华达卡林型金矿知识图谱（部分展示）

Fig. 6 Knowledge graph of Carlin-type gold deposits in Southwest Guizhou and Nevada (partial)

1.3 知识图谱挖掘与分析算法

1.3.1 矿床社区发现聚类算法

为了定量揭示不同矿床在成矿要素上的内在关联与聚类模式,本文基于构建的知识图谱拓扑结构,采用层次聚类算法对典型矿床进行了社区发现分析。具体的研究方法如下:

本文将知识图谱中的每一个实体视为独立节点,利用 Jaccard 相似系数作为衡量矿床间地质特征相似性的核心指标。Jaccard 系数通过计算两个矿床节点在图谱中共享邻居节点(如共有的赋矿地层、控矿构造、围岩蚀变或矿石矿物等)的比例,客观量化了它们在地质成因与特征上的重叠程度。计算公式为:

$$J(A, B) = \frac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|} \quad (1)$$

其中, $N(A)$ 和 $N(B)$ 分别代表矿床 A 和矿床 B 在图谱中的邻居节点集合。

在此基础上,构建距离矩阵 ($\text{Distance} = 1 - J$), 并采用平均连接法进行自底向上的聚合聚类,最终生成矿床相似性谱系图。这种基于知识和数据驱动的方法能够自动识别出具有高度相似成矿特征的矿床组合(亚类),从而在宏观上揭示潜在的成矿系列划分与成因联系。

1.3.2 成矿关联指示强度计算

为了定量刻画不同地质要素间的指示强弱,本文基于图谱拓扑统计学定义了“关联指示强度”(Association Indicator Strength, AIS)。假设知识图谱三元组集合为 T , 对于级联路径中的任意上游实体 e_i (如“烂泥沟金矿”)与下游实体 e_j (如“黄铁矿化”), 其关联强度 $W_{i,j}$ 定义为该关系在多源语料库中的共现频次:

$$W(e_i, e_j) = \sum_{t \in T} \delta(t = \langle e_i, r, e_j \rangle) \quad (2)$$

其中, δ 为判别函数,当三元组 t 匹配当前的头尾实体对时取值为 1, 否则为 0。值得注意的是,基于文献挖掘的共现频次虽然并不直接等同于成矿物理化学过程中的作用强度或元素的富集程度,但它客观量化了地质学家在长期勘查实践中形成的专家共识与经验权重。在地质记录中,高频出现的共生组合往往对应着

那些具有高辨识度且在区域上广泛发育的地质特征。因此，本文定义的“关联强度”实质上表征了特定地质要素在现有知识体系中作为找矿线索的统计显著性与可信度。在此基础上，为了量化特定地球化学元素异常 (e_{chem}) 在成矿系统中的显著性，本文进一步定义了该元素的“累积指示权重” (Cumulative Indicator Weight, CIW)，即所有指向该元素的上游蚀变组合路径权重的总和：

$$CIW(e_{chem}) = \sum_{m \in S_{alt}} W(e_m, e_{chem}) \quad (3)$$

其中， S_{alt} 为与该元素相连的蚀变矿物节点集合。

2 垂直领域大模型性能评估

为全面、客观地评估所构建的本地化垂直领域智能问答系统在矿产勘查领域的实际应用效能，本文设计了“判别式-生成式”双重评测基准。该基准包含基于标准化试题的客观知识评测和基于开放式问答的主观生成质量评测，旨在从知识准确性、逻辑连贯性及来源可靠性等多个维度，对比分析本系统与主流通用大模型 (LLMs) 及未增强基座模型的性能差异。

2.1 客观题问答效果评估

客观评测旨在量化模型对地质专业事实性知识的掌握程度。本文基于权威文献和专著，设计了包含 200 道关于卡林型金矿专业知识的单选题 (图 7)，涵盖了 5 个方面：区域成矿地质背景、矿床系列及其地质特征、区域成矿系统的基本要素、区域成矿系统的控矿因素、区域成矿系统的控矿规律。测试结果 (表 2) 显示，本地部署的 RAG-黔西南模型准确率高达 99%，其性能不仅显著优于本地直接部署的 DeepSeek-32B 模型 (88%)，也超越了包括 Gemini 2.5 Pro (97%)、GPT-4o (94%) 在内的多个通用旗舰大模型的联网搜索版本。这充分证明了 RAG 架构在专业领域的巨大潜力和应用价值。

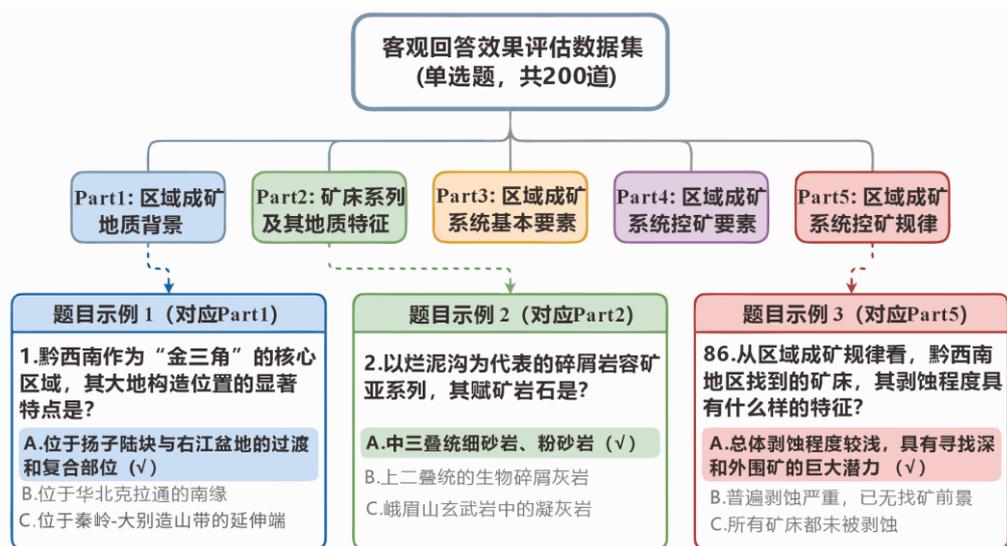


图 7 客观题回答效果评估数据集与示例试题

Fig. 7 Dataset and sample items for performance evaluation of objective question answering

表 2 垂直领域大模型客观题评估

Table 2 Objective evaluation of domain-specific large language models

模型名称	Part1	Part2	Part3	Part4	Part5	总得分
Gemini2.5-Flash	0.85	0.9	1	1	1	0.95
Gemini2.5-Pro	0.95	0.95	1	0.95	1	0.97
Qwen3-Max	0.92	0.95	0.93	1	1	0.95
Qwen2.5	0.65	0.75	0.84	0.95	0.96	0.81
Deepseek-R1	0.85	0.95	0.92	1	1	0.94
Deepseek-V3	0.80	0.80	1	1	0.95	0.91
豆包—1.6	0.91	0.95	1	1	1	0.97
GPT4.1 mini	0.85	0.95	0.90	1	1	0.94
GPT4o	0.91	0.80	1	1	1	0.94
GeoGPT-R1-Preview	0.85	0.95	1	1	0.90	0.94
Deepseek-32B-base	0.85	0.91	0.85	0.92	0.90	0.88
RAG-黔西南	0.95	1	0.98	0.99	1	0.99

2.2 主观问答性能评测

相较于客观题对单一知识点的考察, 主观问答主要评估模型在地质成因分析、成矿规律对比等复杂场景下的逻辑推理与长文本生成能力。鉴于生成式任务评价

的开放性与复杂性，本文构建了基于“专家-自动化”协同的混合评测体系，并选取了通用大语言模型（GPT-4o）、地学领域大模型（GeoGPT-R1-Preview）/未微调基座模型（DeepSeek-32B Base）与本文提出的检索增强模型（RAG-黔西南）进行横向对比实验。

本文设计了包含 30 道复杂地质问题的主观评测集（如“黔西南卡林型金矿床的类型有哪些”）。针对生成式任务难以量化的问题，本文引入 Ragas (Retrieval Augmented Generation Assessment) 自动化评测框架，结合专家人工复核，重点评估以下核心指标：（1）忠实度 (Faithfulness)：评估生成的答案是否严格遵循检索到的上下文信息，以量化模型对“事实幻觉”的抑制能力；（2）答案相关性 (Answer Relevance)：衡量生成内容与用户提问意图的契合程度，反映模型对地学专业术语的理解能力；（3）上下文精确度 (Context Precision)：评价检索模块返回的文献片段是否包含解决问题所需的关键证据，直接反映系统对私有数据的获取能力；（4）知识溯源性 (Knowledge Provenance)：通过人工核验，评估模型引用的参考文献是否真实存在且支持其论点(0 为不可溯源,1 为完全准确)。不同模型在上述指标上的性能对比结果如表 3 所示。

表 3 垂直领域大模型主观性能评估

Table 3 Subjective performance evaluation of domain-specific large language models

模型名称	忠实度	答案相关性	上下文精确度	知识溯源性	综合评价
GPT-4o	0.72	0.92	-	0.00	语言组织流畅，但在处理特定矿区数据时存在明显事实幻觉，且无法溯源。
GeoGPT-R1 Preview	0.78	0.88	0.65	0.40	具备地学常识与推理能力，但因缺乏本地私有文献支持，细节准确度不足。
DeepSeek-R1-32B-Base	0.65	0.80	-	0.00	缺乏领域知识注入，逻辑推理空泛，多为通用性描述。
RAG-黔西南	0.89	0.81	0.90	0.95	论据充分，能够实现从结论到原始本地文献的精准溯源。

注：通用模型与基座模型未挂载本地知识库，故“上下文精确度”不适用；GeoGPT-R1-Preview 测试时挂载通用地学库但未包含本文构建的知识库。

3 知识图谱应用与智能挖掘

3.1 定性对比分析

利用图谱的查询与可视化功能，本文得以在统一的本体框架下，对黔西南和内华达的成矿要素进行系统性的定性对比。对比结果（表 4）直观地揭示了二者的共性与特性。

在共性方面，两个卡林型金矿省表现出高度一致的矿石矿物组合与成矿地球化学亲和性。知识图谱的关联分析显示，两大矿集区均以含砷黄铁矿、毒砂为主要载金矿物，并普遍发育雄黄、雌黄等低温热液矿物。这一特征在图谱中表现为两个区域的矿床实体均与“As-Au-Hg-Sb”元素节点群存在高频共现关系，表明两者在成矿流体的低温、硫化物亲和性等本质特征上具有统一性；同时，知识图谱的拓扑结构也清晰地刻画了两大矿集区在宏观构造演化与微观赋矿环境上的显著差异（特性）。首先，在构造控矿样式上，内华达矿集区的矿床实体在图谱中高频关联“伸展构造”与“滑脱断层”（如 Roberts Mountain 推覆体）等节点，反映了其受控于活动大陆边缘的弧后伸展机制；相比之下，黔西南矿集区则呈现出独特的“褶皱-断裂”复合控矿特征，图谱中大量矿床节点不仅连接高角度断裂，更密集指向“背斜”（如灰家堡背斜）及“层间滑脱带”实体，量化了陆内造山与构造转换对成矿的控制作用；其次，在赋矿围岩与蚀变组合上，不同于内华达地区“碳酸盐岩”与“重晶石化”节点的紧密共现，黔西南矿床节点更倾向于与“泥质岩”、“粉砂岩”等细碎屑岩实体建立强连接，并因此表现出显著发育的“黏土化”特征。

表 4 基于知识图谱的黔西南-内华达卡林型金矿关键成矿要素对比

Table 4 Knowledge graph-based comparison of key metallogenic factors of Carlin-type gold deposits in Southwest Guizhou and Nevada

对比维度	黔西南矿集区	内华达矿集区	对比分析
构造背景	华南板块（被动大陆边缘）	北美板块（活动大陆边缘）	二者均形成于伸展背景，但黔西南的板内裂谷环境更为突出。

控矿构造	以 NE 向、NNE 向高角度断裂为主（如普安-马场断裂），褶皱（如清水河背斜）控矿作用显著。	以 NNW 向高角度断裂为主（如 Getchell 断裂），低角度拆离断层（如 Roberts 山拆离断层）作用关键。	黔西南呈现“断-褶”双重控制，而内华达的“断-断”（高角度与低角度）组合更典型。
赋矿地层	泥盆系（D）至二叠系（P），以泥质岩、粉砂岩（如龙潭组）为绝对主导。	寒武系（C）至泥盆系（D），以碳酸盐岩（如 Popovich 组）为主。	赋矿围岩的岩性差异（硅质-泥质 vs 碳酸盐岩）是导致二者地球化学特征差异的根本原因。
蚀变矿物	硅化、黄铁矿化、雄黄-雌黄矿化、黏土化（高岭石、伊利石）。	硅化、黄铁矿化、雄黄-雌黄矿化、重晶石化、碳酸盐化。	黔西南的“黏土化”更为发育，可能与含泥质围岩有关；内华达的“重晶石化”独有。
矿石矿物	毒砂、黄铁矿、雄黄、雌黄	毒砂、黄铁矿、雄黄、雌黄	组合高度相似，均以含砷矿物为主要载金矿物。

上述基于知识图谱自动聚合的对比结果，与前人开展的系统性人工对比研究结论高度吻合(Hofstra and Cline, 2000; Hofstra and Christensen, 2002; Xie et al., 2018; 谢卓君等, 2019)。本文构建的知识图谱的显著优势在于，将分散于海量文献中的多源异构对比要素在统一的语义框架下进行了结构化整合与可视化展示，从而直观地佐证了卡林型金矿成矿系统的“非唯一性”与“多样性”特征。

3.2 定量知识挖掘

为了定量揭示不同矿床在成矿要素上的内在关联与聚类模式，本文基于构建的知识图谱拓扑结构，采用 Jaccard 相似系数与层次聚类算法，对典型矿床进行了社区发现分析。从层次聚类谱系图（图 8）可见，右江盆地内的卡林型金矿床依据其关键控矿要素的拓扑相似性，被有效划分为边界清晰的不同亚类。基于社区发现分析结果与矿床地质资料挖掘，本文将这些亚类进一步归纳为三大成矿系列。表 5 系统总结了这三大成矿系列的典型聚类簇及其对应的成因机制。

例如，泥堡金矿与戈塘金矿被紧密聚为一类。通过图谱追溯二者的关联路径发现，这两处矿床在赋矿地层、围岩蚀变、矿石构造及矿体形态等关键成矿要素节点上拥有大量共有的邻居节点，导致其 Jaccard 相似系数较高。具体而言，在知识图谱的拓扑结构中，泥堡与戈塘金矿床均与“扬子准地台”这一构造位置节点相连，且在赋矿层位上表现出高度的共现关系，均包含“二叠系”、“茅口组”以及“龙潭组”节点，这反映了两者均受到相似的地层层位控制。此外，在矿体特征与物质组成维度，两个矿床节点不仅同时指向“似层状”、“透镜状”以

及“大型矿床”实体，还共享了“硅化”、“黄铁矿化”作为主要蚀变类型节点，以及“黄铁矿”、“石英”、“方解石”等矿物节点和“浸染状构造”、“角砾状构造”等结构构造节点。这种多维度的特征重叠客观验证了层次聚类结果的合理性，表明知识图谱通过将非结构化的地质描述转化为高维空间中的结构化关联，成功识别出泥堡与戈塘虽地理位置不同，但在成矿系统上属于同一具有显著层控特征的成因亚类(陶平等, 2002; 毛彬吉等, 2018)。

与此同时，知识图谱中非共有的特异性节点进一步揭示了两者虽同属层控型矿床，但在具体容矿岩石（容岩）性质上存在显著差异。在图谱的容岩节点连接上，泥堡金矿床特异性地指向“凝灰岩”与“沉凝灰岩”节点，体现了其矿体赋存与峨眉山玄武岩组火山碎屑岩密切相关的特征；而戈塘金矿床则显著关联于“硅化灰岩角砾岩”与“粘土岩角砾岩”节点，这反映了该矿床矿体主要受控于茅口组古岩溶侵蚀面上的角砾岩带（图9）。这种“同层异岩”的差异也映射在共生矿物节点的分布上，泥堡金矿床表现出与“毒砂”及“毒砂化”节点的强连接，而戈塘金矿床则独有“萤石”、“石膏”及“辉锑矿化”等低温矿物节点关联。这些特异性节点的存在使得知识图谱在通过聚类算法识别宏观层控共性的同时，能够保留并区分矿床个体的微观岩性指纹，精准刻画出不同矿床在成矿微环境上的区别(冉瑞德, 2005; 陈文斌等, 2009)

表 5 右江盆地卡林型金矿的图谱聚类特征与成矿系列划分

Table 5 Clustering characteristics and metallogenic system division of Carlin-type gold deposits in Youjiang Basin based on knowledge graph

成矿系列划分	图谱聚类簇组合	关键控矿要素	地质成因机制
I 类 断裂-构造主控系列	烂泥沟-板其-丫他-华新- 大沟-百地-塘新寨-那郎	高角度断裂 (F3)、毒砂/黄铁 矿化	属原生深部断控系统。受区域性深 大断裂及其次级破碎带严格控制，矿体 呈陡倾状延伸，切穿地层能力强，完整 保留了深部含矿流体沿构造通道垂向运 移与沉淀的记录。
	坝赖-弄朗	高角度逆冲断层 (F1, F2)、 断裂破碎带、许满组复理石	属碎屑岩容矿的断控型亚类。受 F1、F2 高角度断层破碎带控制，具有 显著的切层特征，属于深部流体在碎屑 岩区沿断裂通道贯入成矿。

成矿系列划分	图谱聚类簇组合	关键控矿要素	地质成因机制
II 类 层位-岩性主控系列	泥堡-戈塘	不整合面、硅化顶盖、层间滑脱带	属不整合面-层控型亚类。矿体受上覆不透水层（如粘土岩）的岩性封闭制约，成矿流体沿不整合面侧向运移特征显著，发育大规模的层状硅化体。
	紫木凶	背斜构造、Hg-Tl（铊）异常、雄黄矿化	属背斜-层控型亚类。位于灰家堡背斜核部，受层间虚脱空间与张性裂隙控制，矿体多呈顺层或鞍状产出，且表现出对 Hg-Tl 元素的特殊地球化学亲和性。
III 类 表生风化-堆积系列	万人洞-上大观	第四系粘土、茅口组古侵蚀面、褐铁矿化	属红土-岩溶堆积型成矿系统。严格受控于茅口组古岩溶侵蚀面与第四系风化壳，矿体赋存于岩溶洼地或剥夷面上的红土粘土及角砾岩中，表征为原生矿体经强烈的表生氧化、崩塌堆积与次生富集。
	油菜冲-虎场	残坡积物、褐铁矿化、粘土化	
	老万场-豹子洞	岩溶塌陷堆积、红土化、游离金	

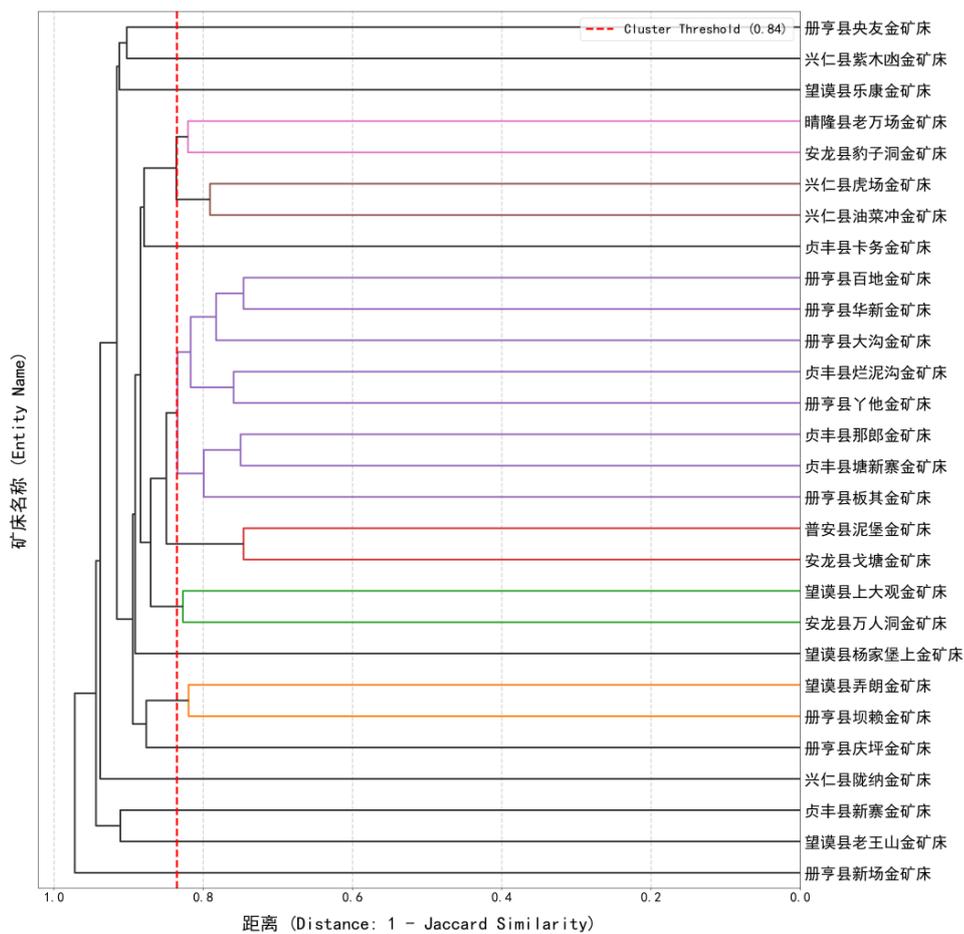


图 8 基于知识图谱的矿床社区发现聚类分析结果

Fig. 8 Cluster analysis results of ore deposit community detection based on knowledge graphs

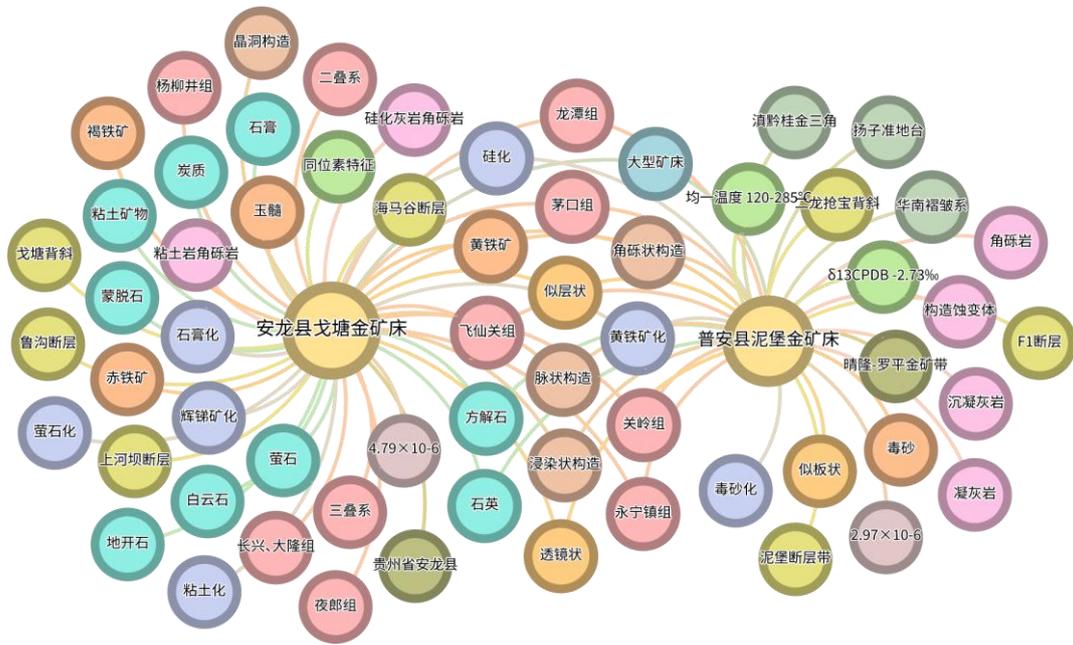


图 9 戈塘金矿床与泥堡金矿床图谱追溯

Fig. 9 Knowledge graph tracing of Getang and Nibao gold deposits

3.3 成矿信息追溯

基于构建的卡林型金矿知识图谱，本文对成矿信息的级联关系进行了深度挖掘与可视化表征（图 10）。该视图突破了传统静态节点的局限，利用 1.3.2 节定义的“关联指示强度” (AIS)与“累积指示权重” (CIW)，将烂泥沟等典型矿床的地质要素重构为“典型矿床—矿石特征—蚀变类型—地球化学异常”的多级流向结构图 10。即基于上述 W 值与 CIW 值生成的桑基能量分流图，图中连线宽度直观反映了实体间的 $W_{i,j}$ 值，而元素节点的排列顺序则依据 CIW 值由高到低呈现。

路径追踪定量显示：以“烂泥沟金矿石”为源点，通过共生关系向下游发散，揭示了矿石与硅化、黄铁矿化、毒砂化及辉锑矿化等围岩蚀变的紧密关联，进而利用指示关系将蚀变组合映射至地球化学域。结果表明，“黄铁矿化/毒砂化—指示—Au/As”路径的 W 值最高，定量验证了其作为主控成矿路径的主导作用；同时，图谱拓扑结构清晰解析了低温成矿元素的指示路径；例如，“辉锑矿化”不仅与 S 元素强关联，更作为指向“Sb”异常的特征性指示节点；而“辰砂矿

化”则构成了指示“Hg”异常的关键路径。这种全链条的信息追溯不仅验证了“Au-As-Sb-Hg”作为黔西南卡林型金矿典型地球化学组合的有效性，也为依据地表地球化学异常反演深部矿化蚀变特征提供了可视化的逻辑支撑。

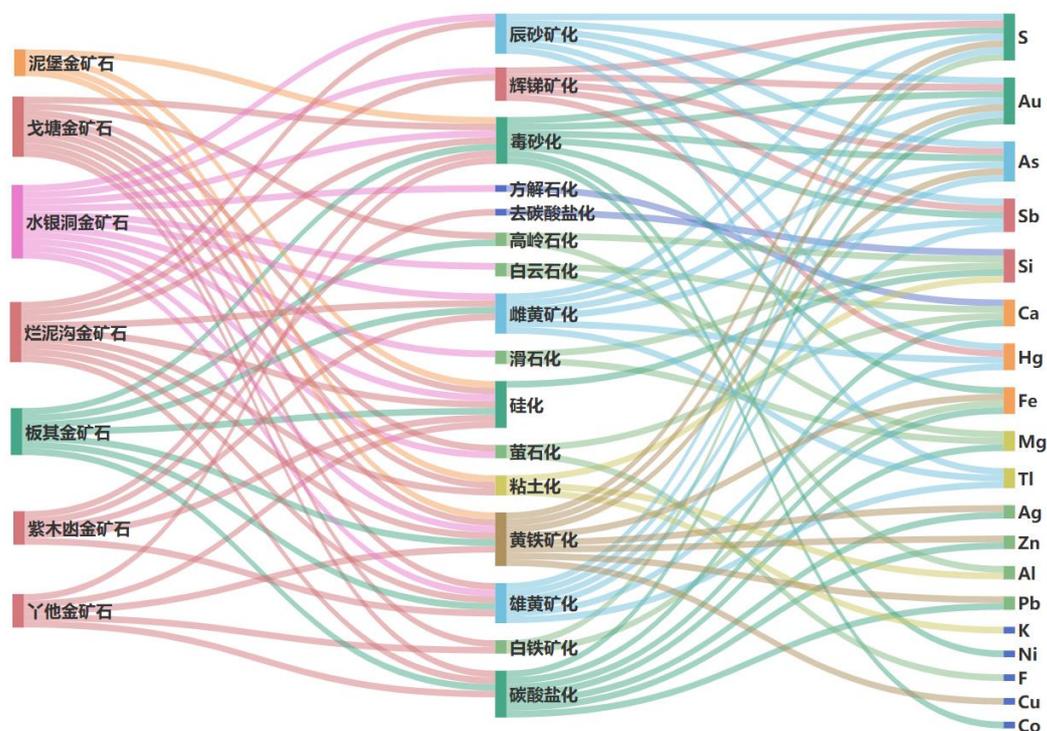


图 10 黔西南典型矿床-围岩蚀变-地球化学异常指示关系图

Fig. 10 Indicative relationships between typical deposits, wall-rock alterations, and geochemical anomalies in Southwest Guizhou.

4. 结论与展望

4.1 主要结论

本文针对地质勘查领域面临的海量非结构化数据难以有效利用与通用大模型专业性不足的双重挑战，创新性地提出了融合垂直领域大语言模型(LLMs)与知识图谱(KG)的智能知识挖掘框架，并以中国黔西南与美国内华达两大卡林型金矿集区为例进行了系统验证。主要研究结论如下：

(1) 本地化垂直领域 RAG 模型有效解决了专业问答的“幻觉”与安全问题。

本文构建的基于 DeepSeek-R1-Distill-Qwen-32B 的本地化 RAG 系统，在基于

本地化知识库架构的前提下,实现了对专业知识的精准检索与生成。实验表明,该模型在卡林型金矿专业客观题测试中达到了 99% 的准确率。在包含主观逻辑推理的评测中,该系统不仅优于 GPT-4o 等通用模型,相比于通用地学大模型(如 GeoGPT),在知识忠实度与可溯源性方面也展现出显著优势,证明了检索增强生成技术在提升地学知识服务可信度方面的巨大潜力;

(2) 基于监督微调(SFT)的知识抽取范式实现了低成本、高精度的成矿知识图谱构建。针对传统深度学习方法对标注数据依赖度高、灵活性差的痛点,本文验证了利用少量高质量指令微调大模型的优越性。SFT 后的 Qwen2.5 模型能够精确理解复杂的地质实体与关系定义,成功从 405 份多源异构文献中构建了包含 6221 个实体和 11779 个关系的跨区域卡林型金矿知识图谱。该方法不仅大幅降低了人工标注成本,也为快速构建其他矿种的专业图谱提供了可复制的技术路径;

(3) 知识图谱驱动的智能挖掘揭示了跨区域成矿规律的深层联系与指示路径。基于构建的知识图谱,本文实现了多维度的定量挖掘:一方面,通过图谱拓扑结构成功识别出“泥堡-戈塘”等具有高相似度的层控型矿床亚类,并精准捕捉了它们在容矿岩石(如凝灰岩与角砾岩)上的微观差异;另一方面,系统量化了从“典型矿床—矿石矿物—围岩蚀变”到“地球化学元素异常”的级联指示路径,证实了“黄铁矿化/毒砂化”对金矿化的主控权重,并从数据驱动角度验证了“Au-As-Sb-Hg”作为卡林型金矿典型地球化学组合的有效性,从而为利用地表地球化学异常反演深部矿化特征提供了可视化的逻辑支撑。

4.2 局限性与未来展望

尽管本文在构建地质领域智能知识系统方面取得了阶段性进展,但受限于数据模态和技术发展,仍存在一定局限性,未来的研究可从以下几个维度深入展开:

(1) 从纯文本挖掘向多模态融合拓展。目前本文的知识抽取主要聚焦于文本语料。然而,地质报告中蕴含大量关键信息的图件(如地质剖面图、镜下鉴定图)和表格数据尚未得到充分利用。未来工作应引入多模态大模型,实现对“图-文-表”异构数据的联合理解与语义对齐,构建更加全息、立体的地质知识图谱,以捕捉矿体空间形态等更直观的地质特征;

(2) 从知识检索向智能预测与辅助决策升级。当前的系统主要实现了知识的组织、检索与关联分析，属于认知智能的初级阶段。下一步将探索引入图神经网络（GNN）与因果推理机制，利用图谱中的隐含路径进行潜在找矿靶区的预测。例如，基于已知矿床的图谱子图特征，在未知区域搜索相似的成矿要素组合，从而将系统功能从“回答已知”拓展至“探索未知”；

(3) 构建基于 Agent 的地质智能体系统。随着 Agent（智能体）技术的发展，未来可将本文构建的 RAG 系统与知识图谱封装为核心工具，构建具备自主规划能力的“地质科研助理”。该智能体不仅能回答问题，还能自主分解复杂的成矿预测任务、调用外部分析工具、撰写地质报告，从而真正实现地学研究范式的智能化变革。

参考文献

- Cao, S., Hu, R., Zhou, Y., et al., 2025. Analysis of Trajectories and Developmental Prospects of Research on Carlin-Type Gold Deposits on the Basis of Big Data Community Detection Algorithms. *Ore Geology Reviews*, 176:106989. doi: 10.1016/j.oregeorev.2025.106989
- Chen, W.B., Wei, B.G., Yang, T.C., et al., 2009. Geological Character and Prospecting Potential of Nibao Gold Deposit in Pu'an County, Guizhou. *Guizhou Geology*, 26(3):170–176 (in Chinese with English abstract).
- Cheng, Q., 2025. A New Paradigm for Mineral Resource Prediction Based on Human Intelligence-Artificial Intelligence Integration. *Earth Science Frontiers*, 32(4):1–19 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2025.7.20
- Deng, J., and Wang, Q., 2016. Gold Mineralization in China: Metallogenic Provinces, Deposit Types and Tectonic Framework. *Gondwana Research*, 36:219–274. doi: 10.1016/j.gr.2015.10.003
- Dong, Y.H., Wang, Y.Z., Tian, J.T., et al., 2025. Research Progress on Porphyry Copper Deposit Prediction Based on Knowledge Graphs. *Earth Science Frontiers*, 32(4):280–290 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2024.1.58
- Feng, T.T., Cai, S.R., and Zhang, Z.J., 2025. Mining Elements of Carbonatite-Type Rare Earth Deposits Based on Knowledge Map. *Earth Science Frontiers*, 32(4):262–279 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2025.3.31
- Fu, Y., Wang, M., Wang, C., et al., 2025. GeoMinLM: A Large Language Model in Geology and Mineral Survey in Yunnan Province. *Ore Geology Reviews*, 178:106638. doi: 10.1016/j.oregeorev.2025.106638
- Goldfarb, R., Qiu, K., Deng, J., et al., 2019. Orogenic Gold Deposits of China. *Geological Society, London, Special Publications*, 480:263–288. doi: 10.1144/SP480-2018-175

- Hofstra, A.H., and Christensen, O.D., 2002. Comparison of Carlin-Type Au Deposits in the United States, China, and Indonesia—Implications for Genetic Models and Exploration. *US Geological Survey Open-File Report*, 02-131.
- Hofstra, A.H., and Cline, J.S., 2000. Characteristics and Models for Carlin-Type Gold Deposits. *Reviews in Economic Geology*, 13:163–220. doi: 10.5382/Rev.13.05
- Hu, R., Fu, S., Huang, Y., et al., 2017. The Giant South China Mesozoic Low-Temperature Metallogenic Domain: Reviews and a New Geodynamic Model. *Journal of Asian Earth Sciences*, 137:9–34. doi: 10.1016/j.jseas.2016.10.016
- Hu, Y., Mai, G., Cundy, C., et al., 2023. Geo-Knowledge-Guided GPT Models Improve the Extraction of Location Descriptions from Disaster-Related Social Media Messages. *International Journal of Geographical Information Science*, 37(11):2289–2318.
- Jiang, Z., Zhong, L., Sun, M., et al., 2024. Efficient Knowledge Infusion via KG-LLM Alignment. *arXiv preprint*, arXiv:2406.03746. doi: 10.48550/arXiv.2406.03746
- Lewis, P., Perez, E., Piktus, A., et al., 2020. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.
- Li, B.W., Wang, Y.Z., Ding, Z.J., et al., 2025. Intelligent Search Technology for Jiaodong Gold Deposits Based on Large Models and GraphRAG. *Earth Science Frontiers*, 32(4):155–164 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2025.4.77
- Li, G., Wang, P., and Ke, W., 2023. Revisiting Large Language Models as Zero-Shot Relation Extractors. *arXiv preprint*, arXiv:2310.05028. doi: 10.48550/arXiv.2310.05028
- Ma, X., 2022. Knowledge Graph Construction and Application in Geosciences: A Review. *Computers & Geosciences*, 161:105082. doi: 10.1016/j.cageo.2022.105082
- Mao, B.J., Ran, R.D., Kuang, S.D., et al., 2018. Genesis of the Getang Gold Deposit in the Southwest Guizhou. *Contributions to Geology and Mineral Resources Research*, 33(2):168–175 (in Chinese with English abstract).
- Peng, J.J., and Lin, K., 2024. Knowledge Graph Analysis of Mineralization Laws Research of Lithium Ore. *China Mining Magazine*, 33(9):228–235 (in Chinese with English abstract).
- Pi, Q., Hu, R., Xiong, B., et al., 2017. In Situ SIMS U-Pb Dating of Hydrothermal Rutile: Reliable Age for the Zhesang Carlin-Type Gold Deposit in the Golden Triangle Region, SW China. *Mineralium Deposita*, 52(8):1179–1190. doi: 10.1007/s00126-017-0715-y
- Qiu, Q.J., Tian, M., Wu, Q.R., et al., 2025. Construction and Application of Geological Knowledge Graph Based on Multi-Source Heterogeneous Data. *Earth Science Frontiers* (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2024.11.69
- Raiaan, M.A.K., Mukta, M.S.H., Fatema, K., et al., 2024. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access*, 12:26839–26874.
- Ran, R.D., 2005. Characteristic and Metallogenic Mechanism of the Gold Deposits with Karst Structure as Holding Ore Space in the Southwest of Guizhou: Taking the Getang Gold Deposit in Anlong as an Example. *Guizhou Geology*, 22(1):14–21 (in Chinese with English abstract).

- Shi, L.Y., and Zuo, R.G., 2025. Foundation Model for Mineral Prospectivity Mapping. *Earth Science* (in Chinese with English abstract). doi: 10.3799/dqkx.2025.190
- Sutanto, P., Santoso, J., Setiawan, E.I., et al., 2024. LLM Distillation for Efficient Few-Shot Multiple Choice Question Answering. *arXiv preprint*, arXiv:2412.09807.
- Tan, S.M., Shi, G.D., Lei, L.Q., et al., 2007. Carlin-Type Gold Deposits Distribution and Prospecting in China. *Geological Survey and Research*, 30(4):289–294 (in Chinese with English abstract).
- Tao, P., Li, P.G., and Li, K.Q., 2002. The Structure of the Deposits of the Nibao Goldfield and Its Relationship with Metallogenesis. *Guizhou Geology*, 19(4):221–227 (in Chinese with English abstract).
- Tian, S., Luo, Y., Xu, T., et al., 2024. KG-Adapter: Enabling Knowledge Graph Integration in Large Language Models Through Parameter-Efficient Fine-Tuning. *Findings of the Association for Computational Linguistics: ACL 2024*, 3813–3828. doi: 10.18653/v1/2024.findings-acl.229
- Wang, C.B., Wang, M.G., Wang, B., et al., 2024. Knowledge Graph-Infused Quantitative Mineral Resource Forecasting. *Earth Science Frontiers*, 31(4):26–36 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2024.5.3
- Wang, Q., and Groves, D., 2018. Carlin-Style Gold Deposits, Youjiang Basin, China: Tectono-Thermal and Structural Analogues of the Carlin-Type Gold Deposits, Nevada, USA. *Mineralium Deposita*, 53(7):909–918.
- Xie, Z.J., Xia, Y., Cline, J.S., et al., 2019. A Comparison Between Carlin-Type Au Deposits in Guizhou of China and Nevada of USA and Its Implications for Exploration. *Mineral Deposits*, 38(5):1077–1093 (in Chinese with English abstract).
- Xie, Z., Xia, Y., Cline, J.S., et al., 2018. Are There Carlin-Type Gold Deposits in China? A Comparison of the Guizhou, China, Deposits with Nevada, USA, Deposits. *Reviews in Economic Geology*, 20:187–233.
- Xin, Y., Li, S., Li, M., et al., 2025. Knowledge Graph Construction with BERT-BiLSTM-IDCNN-CRF and Graph Algorithms for Metallogenic Pattern Discovery: A Case Study of Pegmatite-Type Lithium Deposits in China. *Ore Geology Reviews*, 179:106514. doi: 10.1016/j.oregeorev.2025.106514
- Zhang, B., and Soh, H., 2024. Extract, Define, Canonicalize: An LLM-Based Framework for Knowledge Graph Construction. *arXiv preprint*, arXiv:2404.03868.
- Zhang, B.Y., Tang, J.C., Zhang, T.Y., et al., 2025. Knowledge Graph and Question-Answering Model for Geological Prospecting Empowered by Large Language Models. *Earth Science*, 50(8):1–24 (in Chinese with English abstract). doi: 10.3799/dqkx.2025.176
- Zhang, Y., Wei, C., He, Z., et al., 2024. GeoGPT: An Assistant for Understanding and Processing Geospatial Tasks. *International Journal of Applied Earth Observation and Geoinformation*, 131:103976. doi: 10.1016/j.jag.2024.103976
- Zhang, Z., Yang, Z., Jian, F., et al., 2025. Interpretability-Enhanced Mineral Prospectivity Models: A Synergistic Approach Using Large Language Models, Knowledge Graphs, and Machine Learning. *Mathematical Geosciences*. doi: 10.1007/s11004-025-10231-3
- Zhao, M., Zhang, Z., Yang, J., et al., 2025. Knowledge Graph Construction and Knowledge Discovery for Porphyry Copper Deposits. *Ore Geology Reviews*, 186:106875. doi: 10.1016/j.oregeorev.2025.106875

- Zhao, P.D., 2019. Characteristics and Rational Utilization of Geological Big Data. *Earth Science Frontiers*, 26(4):1–5 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2018.9.8
- Zhou, Y.Z., Zhang, Q.L., Huang, Y.J., et al., 2021a. Constructing Knowledge Graph for the Porphyry Copper Deposit in the Qingzhou-Hangzhou Bay Area: Insight into Knowledge Graph Based Mineral Resource Prediction and Evaluation. *Earth Science Frontiers*, 28(3):67–75 (in Chinese with English abstract). doi: 10.13745/j.esf.sf.2021.1.2
- Zhou, Y.Z., Zuo, R.G., Liu, G., et al., 2021b. The Great-Leap-Forward Development of Mathematical Geoscience During 2010-2019: Big Data and Artificial Intelligence Algorithm Are Changing Mathematical Geoscience. *Bulletin of Mineralogy, Petrology and Geochemistry*, 40(3):556–573 (in Chinese with English abstract). doi: 10.19658/j.issn.1007-2802.2021.40.038

附中文参考文献

- 陈文斌, 韦标根, 杨天才, 等, 2009. 贵州普安县泥堡金矿床地质特征与找矿潜力. *贵州地质*, 26 (3): 170 - 176.
- 成秋明, 2025. 面向人类智能与人工智能融合的矿产资源预测新范式. *地学前缘*, 32 (4): 1 - 19. doi: 10.13745/j.esf.sf.2025.7.20
- 董宇浩, 王永志, 田江涛, 等, 2025. 基于知识图谱的斑岩型铜矿预测研究进展. *地学前缘*, 32 (4): 280 - 290. doi: 10.13745/j.esf.sf.2024.1.58
- 冯婷婷, 蔡诗柔, 张振杰, 2025. 基于知识图谱的碳酸岩型稀土矿成矿要素挖掘. *地学前缘*, 32 (4): 262 - 279. doi: 10.13745/j.esf.sf.2025.3.31
- 李博文, 王永志, 丁正江, 等, 2025. 基于大模型与 GraphRAG 的胶东金矿智能搜索技术. *地学前缘*, 32 (4): 155 - 164. doi: 10.13745/j.esf.sf.2025.4.77
- 毛彬吉, 冉瑞德, 况顺达, 等, 2018. 黔西南戈塘金矿成因再认识. *地质找矿论丛*, 33 (2): 168 - 175.
- 彭晶晶, 林锴, 2024. 锂矿成矿规律研究的知识图谱分析. *中国矿业*, 33 (9): 228 - 235.
- 邱芹军, 田苗, 吴麒瑞, 等, 2025. 基于多源异构数据的地质知识图谱构建与应用. *地学前缘*. doi: 10.13745/j.esf.sf.2024.11.69
- 冉瑞德, 2005. 黔西南岩溶构造容矿金矿床特征及成矿机理——以安龙戈塘金矿床为例. *贵州地质*, 22 (1): 14 - 21.
- 师路易, 左仁广, 2025. 矿产预测大模型. *地球科学*. doi: 10.3799/dqkx.2025.190
- 谭仕敏, 施国栋, 雷良奇, 等, 2007. 中国卡林型金矿的分布规律及找矿前景. *地质调查与研究*, 30 (4): 289 - 294.
- 陶平, 李沛刚, 李克庆, 2002. 贵州泥堡金矿区矿床构造及其与成矿的关系. *贵州地质*, 19 (4): 221 - 227.
- 王成彬, 王明果, 王博, 等, 2024. 融合知识图谱的矿产资源定量预测. *地学前缘*, 31 (4): 26 - 36. doi: 10.13745/j.esf.sf.2024.5.3
- 谢卓君, 夏勇, Cline, J.S., 等, 2019. 中国贵州与美国内华达卡林型金矿对比及对找矿勘查的指示作用. *矿床地质*, 38 (5): 1077 - 1093.
- 张宝一, 唐嘉成, 张彤蕴, 等, 2025. 大语言模型赋能的地质找矿知识图谱与问答模型构建. *地球科学*, 50 (8): 1 - 24. doi:10.3799/dqkx.2025.176

- 赵鹏大, 2019. 地质大数据特点及其合理开发利用. 地学前缘, 26 (4): 1 - 5. doi: 10.13745/j.esf.sf.2018.9.8
- 周永章, 张前龙, 黄永健, 等, 2021. 钦杭成矿带斑岩铜矿知识图谱构建及应用展望. 地学前缘, 28 (3): 67 - 75. doi: 10.13745/j.esf.sf.2021.1.2
- 周永章, 左仁广, 刘刚, 等, 2021. 数学地球科学跨越发展的十年:大数据、人工智能算法正在改变地质学. 矿物岩石地球化学通报, 40 (3): 556 - 573.doi: 10.19658/j.issn.1007-2802.2021.40.038