·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

热震理论及其意义

张旗[1*]，原杰[2]，焦守涛[3, 4]

1. 中国科学院 地质与地球物理研究所 ，北京 100029；
2. 邢台学院 资源与环境学院，河北邢台 054001；
3. 中国地质调查局自然资源综合调查指挥中心 ，北京 100055；
4. 自然资源部 地质信息工程技术创新中心 ，北京 100055

热震理论是一个正在崛起的地震新理论，它认为地震不仅与构造有关，还与岩浆活动有关。一部地球演化史，即地球散热的历史。地震即是地球热能转变为动能的具体体现。

热震理论的具体表现是地下深处发生了爆炸，所以也可以称为爆炸地震理论。野外见到的隐爆角砾岩即是古爆炸地震的证据，也是地震的震源中心。据研究，中国近期发生的破坏性大的大地震，如 1966 年的邢台地震，1975 年的海城地震，1976 年的唐山地震、2008 年的汶川地震、2023 年的甘肃积石山地震，甚至今年（2025 年 1 月 7 日）刚刚发生的西藏定日 6.9 级大地震，都可能是爆炸地震所为。

热震理论意义巨大，表现在下述 4 个方面：

1. 解决了地震预测的难题。构造地震理论认为地震是不可能预测的，而爆炸地震理论认为地震是可以预测的。爆炸地震认为地震可以预测的根据是：地震发生是有前兆的，前兆包括地下深处有毒有害气体的泄漏，地温增加及其导致的地电异常以及动物异常反应等。

2. 提出了地震成矿新理论。大家都知道，隐爆角砾岩是有利于成矿的，隐爆角砾岩型矿床即爆炸成矿理论成因的。隐爆角砾岩下面有花岗岩（或斑岩）侵入体，斑岩型矿床也是爆炸造成的。此外，爆炸产生的强大冲击波会形成许多新的断裂，在这些断裂中沉淀形成的矿床也是爆炸成矿的产物。于是，根据爆炸出现的位置不同，产生不同的矿床，可以归纳为“三层楼”成矿模式：在爆炸中心之上为卡林型矿床，细脉浸染型矿床，石英脉型矿床等；在爆炸中心处为隐爆角砾岩型矿床；在爆炸中心之下为斑岩型矿床。

3. 减灾防震。地震是大自然现象，人类不可能防止地震的发生。但是，打深钻是提前释放出深部气体与流体，是有可能降低爆炸地震发生地震的几率，降低地震的震级。

4. 发现今天正在形成的花岗岩侵入及其伴生的矿床。从理论上，我们不能排除全球今天有正在形成的花岗岩侵入体，也不能排除全球今天正在形成的矿床。但是，它们在哪里呢？如何识别呢？全世界科学家一筹莫展。而爆炸地震理论则表明，爆炸地震的震源中心为隐爆角砾岩，隐爆角砾岩之下有正在侵入的花岗岩侵入体，而隐爆角砾岩本身即是一种矿床的载体。因此，爆炸地震解决了如何探明今天正在形成的花岗岩侵入体以及今天正在形成的与花岗岩有关的矿床的问题。而且，可以根据现在地壳的厚度知道正在形成的花岗岩的性质以及正在形成的矿床的类型。

关键词：热震；地震预测；地震成矿；减灾防震

深地国家科技重大专项：深地特深科学钻探选址研究（2024ZD1001000）；自然资源部深地科学与探测技术实验室开放课题（202211）；国家重点研发计划项目：基于地质云的地质灾害基础信息提取与大数据分析挖掘（2018YFC1505501）；中国地质调查局项目：地球科学大数据‘一张图’体系建设与知识服务（DD20230761）

[第一作者、通讯作者] 张旗（1937–），男，研究员，从事岩石学和地球化学相关的科研工作. E-mail：zq1937@126.com

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

基于大语言模型的地质文本中译英

于新慧[1,2,3]，周永章[1,2,3*]，朱彪彪[1,2,3]，王郑哲[1,2,3]

1. 中山大学 地球科学与工程学院，广东珠海 519000；
2. 中山大学地球环境与地球资源研究中心 ，广州 510275；
3. 广东省地质过程与矿产资源探查重点实验室 ，广东珠海 519000

大语言模型（LLM，Large Language Model）是一种基于深度学习的自然语言处理模型，具备处理复杂语言任务的能力，这些模型通过训练数亿至数千亿参数，能够执行包括但不限于文本分类、情感分析、机器翻译、问答系统和文本生成等任务。大语言模型凭借其卓越的语言理解和生成能力，已经成为推动人工智能在医疗和金融等多个领域研究和应用的前沿力量。目前人工智能技术在地球科学领域研究中已经得到了广泛的应用，并在多个方面取得了显著的进展。命名实体识别（NER，Named Entity Recognition），是指识别文本中具有特定意义的实体，主要包括人名、地名、机构名、专有名词等。命名实体识别不仅是文本处理的基础，还是信息提取、文档分类、知识图谱构建、机器翻译及问答系统等多种自然语言处理任务的重要基石。地质文本数据类型多样，文本的语言风格因数据来源和使用场景的不同而有所差异，但地质文本中均包含大量专业术语和领域特定词汇，如矿物学、岩石学、矿床学、构造地质学以及地质勘探技术等术语。这些术语在进行翻译时，常常会面临着专业术语的准确性与一致性弱、缺乏领域知识的深度以及文化和语言背景差异导致易产生误解等方面的问题。地质文本的翻译一般需要使用标准专业地质词典或数据库，并结合权威地质论文，特别是参考权威机构（如国际地层委员会（ICS））发布的术语标准，以及具备地质背景的专家或研究人员审校之后才能确保高质量的翻译结果。目前常用的翻译方式为学术翻译软件（如 Google Translate、DeepL 和知云文献翻译等）和大语言模型（如 Deepseek、ChatGPT、Kimi 和文心一言等）翻译。翻译软件更新术语快，可以自定义专业词库，但其缺乏深层次的语境分析。常常采用直译的方式，在处理复杂句式、专业术语时会出现逻辑混乱或翻译不准确的现象，特别是在进行涉密文本数据的翻译时，用户的数据隐私存在被泄露的风险。大语言模型有更好的上下文理解能力，具有灵活性与扩展性，可以通过微调适配垂直领域（如地质、医疗、法律等行业），提供个性化翻译服务，且支持实时对话交互。尤其是目前的开源大语言模型可以用于本地部署来规避用于数据隐私泄露的风险。但通用大语言模型会因原始训练数据缺乏专业领域知识而产生“幻觉”问题，且训练数据受限于历史文本，无法及时更新而导致大语言模型输出过时或错误信息。因此，本研究提出通过构建地质命名实体模型和地质专业词典的外部向量数据库与本地大语言模型结合，利用命名实体识别（NER）和检索增强生成（RAG，Retrieval-Augmented Generation）技术，基于大语言模型进行地质专业术语增强的中译英翻译方法。在充分发挥大语言模型的上下文理解和处理复杂语言现象的能力的同时，配备可随时更新地质术语的专业词库，进而实现地质文本的高质量翻译。

关键词：大语言模型；自然语言处理；地质翻译；NER；RAG

基金项目：National Key Research and Development Plan（2022YFF0801201）；国家重点研发计划（2022YFF0801201）；Supported by the National Natural Science Foundation of China（U1911202）；国家自然科学基金资助项目（U1911202）；Guangdong Key Areas Research and Development Project（2020B1111370001）；广东省重点领域研发计划项目（2020B1111370001）

第一作者简介：于新慧（1995–），博士研究生，研究方向：地质大数据. Email：yuxh29@mail2.sysu.edu.cn

*通信作者简介：周永章（1963–），教授，研究方向：大数据与数学地球科学. Email：zhouyz@mail.sysu.edu.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

基于计算机视觉的岩石薄片智能鉴定及系统

臧春艳[1*]

1. 中海油能源发展有限公司 中海油实验中心，天津 300457

岩石薄片是石油勘探开发中提供岩性和储集条件的重要基础实验。本次利用计算机视觉，基于深度学习模型建立了高效的岩石薄片智能鉴定系统，实现自动采集和交互式智能鉴定，从而代替传统手工鉴定。同时基于大数据智能学习，通过全自动高清、全息图像拍照系统进行鉴定矿物智能识别。

首先开发了不同光学系统下岩石薄片数字化信息采集系统，软件操作界面控制正交、 单偏、荧光、试板各状态自动定位图像采集，通过单偏正交多角度“原位叠加”采集技术，实现全薄片多光性多角度自动拼接。同时开发一套智能化信息管理系统，首先利用卷积神经网络模型对岩石进行识别，建立大规模的岩石矿物标准样本库，然后进行图像预处理，实现自动分割、自动识别、自动标注等功能。通过智能识别技术利用专家库进行训练，识别石英、不同长石类型和各种岩屑等几十种矿物，如长石细分为斜长石、微斜长石、条纹长石、正长石等，通过标准命名规则实现岩石薄片智能定名。该系统还可以定量分析孔隙含量、自动拼接全视域图像，更直观的展示孔缝分布等岩石结构特征。利用自动采集系统和智能识别系统来模拟人工鉴定流程，完成岩石薄片的智能鉴定，实现了不同长石矿物和不同类别的岩屑类型等几十种矿物智能鉴定的创新，准确率达到 95%以上，实现了由传统的人工鉴定走向智能鉴定。此项工作为勘探评价和储层研究提供更优更快的实验数据。

关键词：岩石薄片；智能鉴定；全自动采集；神经网络；岩屑

臧春艳（1976–），女，高级工程师，主要从事地质实验和储层评价研究.Email：zangchy@cnooc.com.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

矿物分布外检测

季晓慧[1*]

1. 中国地质大学（北京），北京 100083

深度学习已被越来越多地用于识别矿物。然而，深度学习只能用于识别训练集分布内的矿物，而训练集外的矿物不可避免地被错误分类为训练集中的某个类别。分布外检测可用于解决深度学习模型误将训练类别外的样本分类为训练集内类别的问题，已被应用于医学图像、基因测序、机械等领域，且表现出良好性能。将介绍常用的分布外检测方法，以及采用分布外检测来解决矿物识别时误将训练类别外的矿物识别为训练类别内某种矿物的方法，并介绍相关实验结果及未来进一步的工作。

关键词：矿物识别；深度学习；分布外检测

第一及通信作者简介：季晓慧（1977–），副教授，研究方向：人工智能，E-mail：xhji@cugb.edu.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

闪锌矿地球化学组成揭示铅锌矿床成因类型

曾庆文[1*]，牛旭东[1]，郁凡[1]

1. 中国地质大学（北京）地球科学与资源学院，北京 100083

长期以来根据特定矿物的成分特征确定矿床的成因类型，一直是经济地质学家和矿业公司关注的焦点。传统的二维图由于尺寸的限制，无法包含所有元素信息，可能会对判别结果产生偏差。在对铅锌矿床类型进行分类时尤其如此。本研究采用了四种广泛使用的机器学习算法（随机森林、极限梯度提升、支持向量机和多层感知器）来训练 4908 组闪锌矿元素数据，这些数据来自五种不同的铅锌矿床类型：VMS、SEDEX、MVT、矽卡岩和浅成低温热液矿床。然后通过主成分分析和 t 分布随机邻域嵌入对数据进行可视化解释，结果表明将闪锌矿元素数据简化为二维投影会导致重要特征信息的丢失，从而阻碍其有效区分矿床成因的能力。机器学习结果表明，所有四种模型在测试集上的宏观 F1 得分均高于 0.95，表现出优秀的鲁棒性和出色的泛化能力，证明了使用闪锌矿地球化学数据区分铅锌矿床类型的可靠性。SHAP 值分析强调了 Mn、Fe、Ge、Cd 和 Co 等元素在使用机器学习算法区分矿床类型方面的关键作用。这些模型预测外部独立数据集的综合结果准确率为 83%，同时也被应用于对三个未知类型的铅锌矿床进行分类，预测结果与地质观测结果一致。模型的参数已进一步导出并编程到 Excel 宏程序和交互友好的 EXE 应用程序中，可通过以下方式访问 https://sdeakii.github.io/machine-learning。

关键词：闪锌矿地球化学；机器学习；成因类型；铅锌矿床

国家自然基金委员会优秀青年基金项目（42322206）

第一作者及通信作者简介：曾庆文（1997–），博士研究生，矿物学、岩石学、矿床学专业. E-mail：qingwen@email.cugb.edu.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

基于 DeepSeek 大语言模型与 RAG 技术在滑坡领域垂直模型构建的初步探索

何陆灏[1]，周永章[1*]，马建华[1]，刘蕾[1]

1. 中山大学 地球科学于工程学院，广州 510000

随着大语言模型在自然语言处理领域的快速发展，其在地质灾害监测、预警与辅助决策等方面展现出越来越高的应用价值。然而，通用大语言模型在应对专业性极强的滑坡领域时，往往面临知识覆盖不足与专业术语理解不准确等问题。为此，本研究基于 DeepSeek 大语言模型，结合检索增强生成（RAG）技术，对滑坡领域垂直模型构建进行了初步探索。首先，对滑坡相关文献、案例与地理环境数据进行集中整理与高质量筛选，并利用向量化编码方式构建多模态知识库；随后，在 DeepSeek 模型推断环节运用 RAG 技术，动态检索并融合滑坡灾害风险评估、成因分析及监测预警等关键信息。初步实验结果表明，该垂直化方案在文本分析和知识推理方面更能适应滑坡领域的专业需求，为滑坡灾害预测与管理提供了新思路，同时也为其他专业领域构建垂直大模型提供了可行的技术参考。

关键词：DeepSeek；RAG；滑坡领域；垂直模型；地质灾害

基金项目：国家重点研发计划（2022YFF0801201）；国家自然科学基金资助项目（U1911202）；广东省重点领域研发计划项目（2020B1111370001）；
第一作者：何陆灏（1995–），在读博士研究生，研究方向：地质大数据. Email：hys_0438@qq.com
通讯作者：周永章（1963–），教授，研究方向：大数据与数学地球科学、资源与环境的科教工作. Email：zhouyz@mail.sysu.edu.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

战略性矿产找矿场景下多模态地质数据预训练模型构建与智能化应用

杜婷[1*]，王昭静[1]，周兆巍[1]

1. 包钢集团矿山研究院（有限责任公司）矿产资源战略研究所，内蒙古包头 014030

地质大数据是现代地球科学研究的重要基础，其中，岩石地球化学数据（例如元素丰度、矿物分布、矿物晶体化学参数及矿物标型学参数等）构成了地学研究的基本信息和关键维度。近些年，地学数据呈指数级增长，数据的多源性、高维性和非线性等特征显著，传统的统计分析方法在特征提取和模型识别方面存在较明显的局限性。基于深度神经网络架构，整合机器学习（随机森林、支持向量机等）、计算机视觉（卷积神经网络）和数据挖掘（关联规则分析）等技术，在地质大数据复杂特征提取，非线性关系提取，高效率模式识别，趋势预测分析等方面具有非常重要的意义。本研究旨在通过“数据驱动-模型构建-专家反演”的技术路径，针对稀土、铁、铌等关键战略矿产资源开展更精准、更高效的找矿辅助工作。首先，对多源地质数据集（如区域成矿背景、地质构造、矿物学解析数据、谱图解译数据、地质图件、地球物理及地球化学等）进行标准化处理，通过卷积核特征提取实现数据融合。其次，基于 Transformer 架构的多模态大模型，利用自注意力机制捕捉“元素组合-矿物相变-矿种分布-构造控矿”间的依赖关系，使用地质勘测数据进行预训练，并通过矿床尺度标注数据微调。最后，将预测结果、注意力权重图和特征热力图等呈现于专家评测系统进行解释性分析，矿床尺度标注数据通过对头数量、头维度、堆叠层数、隐藏层数、神经元数、激活函数阈值等超参数进行调整，并利用强化学习算法实现特征权重动态优化，反演成矿机理。通过跟踪 F1 值、置信度、位置精度和范围准等指标评估预测结果，达到工程应用标准，形成迭代化闭环。当前局限性主要体现在专家先验知识量化表征差异和不确定性，以及新的找矿理论和技术迭代速度失配等问题。后续将引入贝叶斯概率图模型，探索构建具有时序性、标型性的知识图谱，实现跨学科知识的深度融合和协同推理，进一步提升找矿算法的智能程度和泛化能力，从而为战略性矿产资源勘察找矿提供更有力的技术支撑。

关键词：地质大数据；大模型；人工智能；找矿预测；深度学习

杜婷（1988–），高级工程师，研究方向：数智化、信息化研究、科技情报分析 Email：361220391@qq.com
王昭静（1986–），高级工程师，研究方向：矿床地质、矿物学研究、人工智能大数据 Email：wzhjing_kevin@163.com
周兆巍（1992–），工程师，研究方向：采矿工程、战略分析 Email：775798850@qq.com

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

机器学习在岩矿地球化学研究中的可视化及应用实例

贺海洋[1*]，张焕宝[1]，李朋刚[1]，曾涛[1]

1. 南华大学 资源环境与安全工程学院，湖南衡阳 421001

人工智能的发展为解决地球科学问题提供了新思路。机器学习是实现人工智能的一种方法。地球化学数据分析为地球科学研究和相关应用提供了数据支持基础。为研究机器学习在地球化学研究中的应用，使用 CiteSpace 进行可视化分析，采用共现网络分析、突现性分析、中介中心性分析和聚类分析方法，并进行可视化图谱解读。机器学习在岩石构造背景和成矿潜力方面有广阔的应用前景。埃达克质岩具有重要的地球动力学和金属成矿意义，其构造背景的准确识别为探讨区域构造岩浆演化过程提供了重要依据将机器学习与地质大数据相结合，构建高精度埃达克质岩构造背景判别模型和可视化图解。本文收集了 1075 条全球埃达克质岩主、微量地球化学数据，使用主成分分析和 t 分布随机近邻嵌入等无监督学习方法进行高维数据降维，采用随机森林、支持向量机、人工神经网络和 K 近邻等机器学习方法进行数据训练，得出准确率为 98.5%的高斯核支持向量机埃达克质岩构造背景判别器，并提出 Ba-Sr/Nd 图解，为汇聚板块边缘、板内火山活动和太古宙克拉通（包括绿岩带）3 种构造背景判别提供依据。花岗岩型铀矿是我国主要的铀矿来源之一，因此，厘定经济、可靠的铀含矿潜力评价方法对于推动我国铀矿事业的发展具有重要意义。结合机器学习模型开展花岗岩型铀含矿潜力评价，在系统收集华南地区地球化学数据的基础上（1417 条，不包含预测数据），通过多层感知机、随机森林以及梯度提升决策树机器学习算法分别构建花岗岩型铀含矿评价模型，同时结合遗传算法和五折交叉优化模型；最后利用混淆矩阵、精确率、召回率和受试者工作特征曲线等开展评价精度验证。结果表明，梯度提升决策树在本数据集中的表现最佳，其在测试集中的准确率达 95.3%。综上，本文对机器学习在岩矿地球化学研究中的进展进行可视化分析，并探讨了机器学习在岩石构造背景和成矿潜力方面的应用。

关键词：机器学习；岩矿地球化学；知识图谱；埃达克质岩；花岗岩型铀矿

基金项目：湖南省自然科学基金项目（2023JJ30507）；湖南省研究生科研创新项目（CX20240824；CX20240825）

第一作者和通讯作者简介：贺海洋（1991–），讲师，研究方向：地学大数据. Email：hehy@usc.edu.cn

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

野外露头数字孪生系统呈现与应用——以陕西府谷县辫状河露头为例

印森林[1*]，李文军[2]

1. 长江大学 录井技术与工程研究院，湖北荆州 434023；
2. 武汉数智地质信息科技有限公司 ，武汉 430074

随着大数据、人工智能等新一代信息技术的快速发展，数字露头和数字地质迎来了新的发展契机。针对露头表征技术的不足，引入了无人机倾斜摄影技术，把野外露头地质研究、三维剖面精细扫描和数字信息化呈现结合起来，形成露头地质特征、地理地貌信息与数字特征的虚实孪生互动，精彩、精准、分层次的呈现了野外露头数字孪生模型。利用倾斜摄影技术对陕西府谷县天生桥露头进行了数据采集和处理，结合倾斜影像中的像素信息生成富有纹理的三维模型，建立了图片与坐标相对应的野外露头三维数字化模型，把生成的 OSGB 模型与地理遥感 DEM 数据在通过三维融合技术进行处理，形成大、小尺度数据的空间数据融合。随后，把野外露头研究的相关岩石体地质属性和工程参数属性数据进行输入。在无人机采集三维模型上形成一套覆盖物理模型的露头地质参数数据。地质属性包括：地层分层数据（海拔高差、颜色、岩性与厚度数据）、沉积构造数据（层理构造、层面构造、特殊构造）、储层砂体构型数据（几何形态、规模、方向与叠置样式）、储层质量数据（孔隙结构、孔隙度、渗透率、含油饱和度）。工程参数属性包括：工程力学参数（脆性、泊松比和杨氏模量）。构建倾斜摄影模型、露头数据模型和地质知识结构模型等，将地质知识结构模型映射到府谷倾斜摄影模型中，三种模型相互耦合演化构建野外露头场景数字孪生模型。该系统将大大拓展野外露头地质研究理论的内涵，对数字地质学的发展影响深远。

关键词：无人机；野外露头；数字孪生；辫状河；府谷县

印森林（1983–），副教授，博士生导师，研究方向：无人机露头表征与建模. Email：yinxiang_love@126.com

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

无人机倾斜摄影技术在致密砂岩露头储层构型解剖中的应用

印森林[1*]，李文军[2]

1. 长江大学 录井技术与工程研究院，湖北荆州 434023；
2. 武汉数智地质信息科技有限公司 ，武汉 430074

针对野外露头研究中存在互动性、定量化和可视化不足的问题，近年发展了无人机倾斜摄影技术可以有效解决上述问题。经无人机采集和处理后的高精度倾斜摄影三维模型，可以基于多视角开展储层构型互动解释，大大增强了储层构型解剖的可信度，较好解决了长期困扰地质研究人员的难题。以塔里木盆地库车河剖面侏罗系阿合组致密砂岩露头为例，经过无人机倾斜摄影采集处理后的模型，把 3-5 级储层构型定量化的方式呈现。研究发现：（1）提出了一套基于高精度倾斜摄影三维模型的储层构型表征技术。该模型可定量表达不同级次构型单元，也可以呈现构型要素的组合关系及其定量规模分布。把处理后的.OSGB 倾斜摄影三维数据模型输入到 EPS 软件中，完成了不同储层构型级别的互动解释工作；（2）建立了露头区中尺度河道叠加样式。通过典型露头区的解剖发现，主河道、次河道构型要素较为发育。砂体形态呈透镜状或半透镜状相互叠置，叠置样式主要有垂向深切叠加、侧向拼接等；（3）建立了定量知识库。剖面河道宽厚比呈较好的线性关系，相关系数 0.75。此外，结合密集采样分析测试工作，揭示了露头砂体内部非均质性特征，主河道砂体内部孔隙度向上渐变降低，呈正韵律分布样式，体现了河道垂向充填式沉积特征；而次河道砂体沿侧积方向叠置，孔隙度呈多期侧向渐变降低，体现了侧积式河沉积特征。该研究不仅较好的拓展了露头表征的技术体系，也为指导地下致密砂岩甜点分布提供了地质依据。

关键词：无人机；倾斜摄影；致密砂岩露头；储层构型；阿合组

第一作者：印森林（1983–），副教授，博士生导师，研究方向：无人机露头表征与建模研究. Email：yinxiang_love@126.com

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

城市湿地公园碳汇核算研究方法研究

冯晶晶[1, 3*]，高国辉[3, 2]，周世武[2, 3]，庄圣炜[3]，颜伟[3]，齐登位[3]，
周永章[2]，卢桂宁[1]

1. 华南理工大学 环境与能源学院，广州 510000；
2. 中山大学 地球环境与地球资源研究中心，广东珠海 519080；
3. 广东埃文低碳科技股份有限公司 ，广州 510000

在全球城市化加速和应对气候变化的大背景下，城市湿地在城市碳循环中的作用至关重要。本文系统评述了城市湿地公园碳汇计量方法，涵盖基于经验统计、生物量调查、碳通量测定以及碳循环模拟的多种方法。经验统计法采用 IPCC 国家温室气体清单法，虽应用广泛但存在局限性；生物量调查法通过样地调查和遥感技术估算碳汇，遥感技术在城市复杂环境下面临挑战；碳通量测定法中的微气象法和同化量法，分别存在下垫面要求高和受环境因素影响大的问题；碳循环模拟法利用多种模型估算碳汇，但模型普适性和参数获取存在难题。针对这些方法的不足，提出未来应从技术创新、多学科融合、标准化建设和跨区域协作等方向开展研究，以推动城市湿地公园碳汇计量研究的发展，助力城市实现“碳达峰、碳中和”目标，为城市生态保护和可持续发展提供科学依据。

关键词：城市湿地公园；碳汇计量；研究方法；碳循环

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

Developing a Smart Conversational Knowledge Base for the Qin-Hang Metallogenic Belt Using the ChatGLM3-6B Large Language Model

何陆灏[1*], 周永章[1]

1. 中山大学 地球环境与地球资源研究中心, 广东珠海 519080

This study leverages the ChatGLM3-6B large language model, which uses the Langchain framework, to develop a smart conversational knowledge base tailored for the Qinzhou - Hangzhou Bays metallogenic belt. The objective is to enhance the model's comprehension and response accuracy concerning specialized geological inquiries. To this end, retrieval-augmented generation(RAG)technology is applied, allowing for the integration of multisource knowledge and dynamic updates to the knowledge base, thus broadening the corpus, increasing knowledge coverage, and improving response quality. To evaluate the efficacy of this specialized knowledge base integration, the study employs BLEU scoring in conjunction with metrics such as precision, recall, and F1 score, comparing the knowledge generation across five language models: ChatGLM3-6B, ChatGLM3-6B RAG, ChatGPT-4, Bing, and Gemini. The evaluation results indicate that the ChatGLM3-6B model, when combined with the Qinzhou - Hangzhou Bays metallogenic knowledge base, demonstrates superior performance in all the metrics, particularly in terms of expertise and information coverage, achieving an F1 score of 0.8689, notably outperforming the other models. In summary, integrating domain-specific knowledge bases with RAG technology effectively enhances large language models for specialized domains. Future work should aim to further expand and refine the knowledge base and optimize its integration with the generation model to improve performance in complex fields, ultimately supporting the development and application of intelligent conversational systems in geology.

关键词: Domain-Specific Large Language Model; ChatGLM3-6B; Langchain; Qinzhou - Hangzhou Bays Metallogenic Belt; Retrieval-Augmented Generation (RAG)

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

基于 RAG 的智能找矿垂直领域大模型的构建

牛露佳[1*]，周永章[1]

1. 中山大学 地球环境与地球资源研究中心，广东珠海 519080

本研究提出了一种基于 RAG 和词典支持的智能找矿大语言模型-ProspectRAG，采用检索增强生成（RAG）的方法，结合领域专家知识和网络百科知识，实现对复杂地质术语的精准识别。通过关键词与保护词的双重机制，ProspectRAG 能够识别领域内的关系词和结构性术语，同时保护特定专业术语，避免其在复杂语境中的误分割与误解读，从而构建更完善的智能找矿术语词典。词典的引入显著提升了大模型的生成能力，且无需微调即可提升专业文本的理解和生成能力，有效推动了智能找矿领域的知识共享和自动化应用，为矿产勘查的智能化体系生成提供了坚实基础。

关键词：地质大数据；人工智能找矿；知识图谱；大语言模型；RAG

获国家重点研发计划项目“跨尺度多源异构地质资源大数据与知识融合理论和方法”（编号：2022YFF0800101）支持

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

Mini-CarbonGPT：碳中和垂直领域大语言模型

马建华[1*]，周永章[1]

1. 中山大学 地球环境与地球资源研究中心，广东珠海 519080

本研究构建的 Mini-CarbonGPT 是一个具有碳中和垂直领域知识的大型语言模型，它依托作者整理的一套全面资源，包括模型监督微调数据、检索数据库和评估数据训练而成。Mini-CarbonGPT 将开源的 GLM-4-9B 模型应用于监督微调数据。使用检索增强生成，在微调后的大模型上构建向量检索系统。本研究设计了专业的训练数据，包括模型预训练数据，并分析了用于构建这些数据集的方法。同时，从客观和主观两个角度创建评估数据，以评估碳中和领域及相关领域大型语言模型的专业理解能力。评估结果表明，在客观问题上 Mini-CarbonGPT 的正确率达到 80.57%，优于原始 GLM-4-9B 模型和 4 个商用 LLM 模型，验证了它的有效性。在主观题的自动化指标、GPT-o1 评分以及关键词覆盖率等方面亦显著提升。尽管语义表达尚需优化，Mini-CarbonGPT 获得的碳中和策略和知识可以作为推进碳中和领域研究者决策的基础。通过本研究的数据收集策略和积累的数据可以作为推进碳中和交叉领域的 LLM 研究，促进综合碳中和决策的进一步发展。

关键词：碳中和；大数据；垂直领域大模型；检索增强生成；碳中和语料库

本研究获国家重点研发计划（项目编号：2022YFF0800101）支持

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

新一代大数据与智能找矿系统研发

周永章[1*]，李文佳[1]，朱彪彪[1]，于新慧[1]，牛露佳[1]，徐述腾[1]

1. 中山大学 地球环境与地球资源研究中心，广东珠海 519080

本研发的智能找矿系统模型通过将区域地质与找矿垂直领域大模型与知识图谱结合，实现长程关联和隐式知识推理。通过关联规则等数据挖掘算法模块，挖掘出与矿致异常相关性最为密切的关联因子。通过深度学习算法高度融合知识推理结果和关联因子，实现区域知识、数据双驱动的智能找矿系统构建，使系统能够从区域地质与找矿垂直领域大数据中检测矿致异常。

关键词：地质大数据；智能找矿；垂直领域大模型；矿床知识图谱

本研究获国家重点研发计划项目“跨尺度多源异构地质资源大数据与知识融合理论和方法”（项目编号：2022YFF0800101）支持。

·专题 27：地学大数据挖掘、机器学习与人工智能算法应用·

Geochemistry π：无需编程基础即可开展机器学习的智能工具

张舟[1*]

1. 浙江大学 地球科学学院，杭州 310058

近年来，机器学习在地球化学领域的应用为相关科学问题带来了新认识。但是，地球化学科研人员需要花费大量时间精力实践机器学习方法。为了降低地球化学家应用数据挖掘方法研究的门槛，我们开发了 Geochemistry π，一款基于 Python 的高度自动化机器学习框架（Framework）。通过一键安装和运行 Geochemistry π，地球化学科研人员只需提供 Excel 表格数据，便能在终端命令行、Web Portal、Jupyter Notebook、Google Colab Notebooks 等多种模式下运行软件，通过问答式操作流程获得经过自动调参后的机器学习模型及其相关图表。

Geochemistry π以 Scikit-learn 库为基础，建立了自动化数据挖掘流程，能够训练各种分类、回归、降维和聚类算法。本工具通过构建数据传输和算法功能应用相分离的层级式架构，保证了框架的可扩展性和可移植性。同时，利用 FLAML 库的 CFO 和 BlendSearch 两种超参数搜索方法来实现 AutoML 模块，并结合 Ray 分布式计算框架来加速模型参数优化过程，集成 MLFlow 库用于机器学习生命周期管理与监测，方便用户在不同尺度下比较多组训练后的模型。此外，整个框架采用前后端分离模式构建 Web Portal 框架，通过友好的 Web 界面演示机器学习模型与数据科学工作流程。

关键词：机器学习；智能工具；地球化学

张舟（1986 年–），男，百人计划研究员，研究方向：地球化学大数据. E-mail：zhangzhou333@zju.edu.cn