

Exploration of AI-Based Geological Prospecting Information Mining in Xinjiang under the Big Data Context

Wenhuan Shen

China Chemical Geology and Mine Bureau, Beijing, 100013, China

Abstract

Xinjiang is in a critical position for ensuring national energy and resource security. However, geological exploration has long been constrained by the problems of complex structures and scattered data. Traditional exploration methods have shown significant technical limitations in predicting deep-seated hidden deposits and analyzing mineralization principles. However, the combination of big data and artificial intelligence technology has brought new ways to break through this situation, that is, by integrating various sources and types of data, using intelligent algorithms to explore the underlying patterns, and implementing quantitative estimates of mineralization possibilities, thus achieving a shift from relying on experience to relying on data.

Keywords

Big data; Artificial intelligence; Geological prospecting in Xinjiang; Mineralization prediction; Deep learning

大数据背景下基于 AI 人工智能的新疆地质找矿信息挖掘探讨

申文环

中化地质矿山总局, 中国·北京 100013

摘要

新疆处于国家能源资源安全保障的关键地位,但长期以来,地质找矿工作一直被构造复杂和数据分散这两个难题所束缚,传统的找矿手段在对深层隐伏矿的预测以及成矿原理的分析时,已经表现出突出的技术局限性,不过,大数据和人工智能技术相结合,给冲破这种局面带来了新途径,即经由整合多种来源不同且类型各异的数据,利用智能算法去探寻其中蕴含的规律,并实施成矿可能性的定量估计,从而做到从依靠经验到凭借数据的模式转变。

关键词

大数据; 人工智能; 新疆地质找矿; 成矿预测; 深度学习

1 引言

1.1 研究背景

新疆矿产资源的禀赋地位对于国家战略格局而言十分关键,这里占有了中国陆地面积的六分之一,承担着保障能源资源安全与‘一带一路’资源枢纽建设的双重使命,但是,当地的地质找矿工作一直被复杂构造和技术局限所困,艰难前行。准噶尔盆地,塔里木盆地,天山造山带,阿尔泰造山带这四个构造单元相互交织叠加,形成了极为复杂的成矿地

质环境,这样的复杂性是资源富集的原因,更是找矿工作遭遇的最大阻碍,传统的找矿方法按照野外考察,物化探异常划定范围,依靠专家经验作出判断这样一种线性路径,在浅表矿勘查时还算有效,可是到了深部隐伏矿预测的时候,它的不足就完全显现出来:数据分布在各个部门和项目当中,缺少一个统一整合的平台;异常解读很大程度上依赖个人经验,主观色彩浓而且难以复现;要想分析深部构造就得做很多人工反演,效率低而且精准度不高。

大数据时代促使信息处理逻辑产生根本变革,海量数据由负担变为财富,算法成为生产力而非仅仅作为工具,人工智能技术尤其是机器学习和深度学习,表现出从繁杂数据自动获取规律的强劲能力,这恰是地质找矿急需之处,把地质调查数据,物探数据,化探数据以及遥感数据整合到一个平台上之后,让神经网络去学习已知矿床的特性模式,利用加强学习算法来改善勘查决策,于是,一种崭新的找矿模式便近在眼前。

【基金项目】中国煤炭地质总局重大科技专项新一轮找矿突破专项研究项目资助成果(项目编号:ZMKJ-2025-ZX04-3)。

【作者简介】申文环(1981—),男,中国山东临朐人,硕士,高级工程师,从事固体矿产成矿规律研究。

1.2 研究目的和意义

传统找矿存在核心难点,即“经验”无法拓展,资深地质专家即便倾尽一生,所及范围也有限,其知识难以完好传承,亦难速成复本。大数据与 AI 相结合带来了突破口,可把专家经验转为可学习的特征模式,把主观评判转为概率预测,把单点勘查转为全域优选,这种转换意义不只是效率的优化,还开创了“数据促使察觉”的先河,算法也许会找出人类专家未曾留意到的成矿联系,进而做到真正的知识增量。

2 新疆地质找矿现状与挑战

2.1 新疆地质特征概述

新疆大地构造格局是由许多不同时代,不同性质的构造单元拼合起来形成的,从北到南依次是阿尔泰造山带,准噶尔盆地,天山造山带,塔里木盆地,昆仑-阿尔金造山带,这种“盆-山-盆-山”的格局是古亚洲洋和特提斯洋长期演化而产生的结果,主要的岩石类型包含前寒武纪变质岩系、古生代火山-沉积岩系以及中生代陆相碎屑岩系,地层分布十分复杂,成矿作用又有多期性特征,二者相互影响,共同造成了多种成矿类型,比如与造山作用有关的铜镍硫化物矿床,与岩浆热液有关的斑岩型铜钼矿床,与沉积作用有关的层控型铅锌矿床等^[1]。

2.2 新疆地质找矿现状

新疆现已查明 138 种矿产,石油,天然气,煤炭,铁,铜,镍,金,稀有金属等资源储量位于全国靠前列,地质填图,物探测量,化探取样,槽探钻探这些传统找矿手段在过去数十年收获了突出成果,喀拉通克铜镍矿,阿舍勒铜锌矿,土屋-延东铜矿等大型矿床都是传统方法取得成效的例子,但是从 21 世纪开始,表层矿的发现率大幅下滑,找矿工作只能向深层和覆盖区域推进,传统方法所产生的效益快速缩减。

2.3 面临的挑战

新疆地质找矿遭遇结构性难题,地质状况极为复杂,单一方法很难取得理想效果。物探异常也许由非矿因素造成,化探异常有可能因后期改造而被遮掩,遥感图像还可能受到地表覆盖影响,传统方法还存在另一重大局限,即数据利用效率低下,数十年来积攒起大量地质数据,这些数据储存在众多机构之中,其格式,精度以及时空分辨率均不相同,又缺少有效地整合与深入挖掘途径,所以,这些数据蕴含的价值尚未得到全面发挥^[2]。

3 大数据与 AI 人工智能技术概述

3.1 大数据技术

大数据的定义并非仅仅指“数据量大”,其本质特征体现在体量、速度、丰富性以及价值密度这四个维度的相互交织之中,地质数据本身就具有大数据的特性:在体量方面,一个中等规模的找矿项目所产生的数据量可达 TB 级别;

在丰富性方面,包含结构化的钻孔数据,半结构化的物探剖面以及非结构化的遥感影像等多种类型;至于价值密度,则是有效的成矿信息常常隐藏在大量的背景噪声之下,而分布式存储(Hadoop, Spark),流式计算,数据清洗与标准化这些技术就是大数据处理的关键技术所在,它们为后面的智能分析形成了基础^[3]。

3.2 AI 人工智能技术

人工智能在显著地质找矿方面具有应用潜力,这源于它具备强大的模式识别及非线性建模能力,随机森林,支持向量机以及 XGBoost 这样的机器学习算法,精通于从高维特征空间当中获取分类边界,卷积神经网络,循环神经网络以及生成对抗网络这些深度学习算法,则可以自主学习到数据的层级表现形式,地质找矿本质上属于一种空间预测任务,而空间数据所具有的结构性正好同深度学习的框架十分契合。

4 基于大数据与 AI 的新疆地质找矿信息挖掘方法

4.1 地质数据的收集与整合

智能找矿首先要执行数据整合,这一步骤往往被忽视,地质考察产生的数据,物探得到的数据,化探获取的数据,遥感得到的数据,它们产生于不同时期,不同项目,而且采集标准存在差异,如果直接把这些数据合起来,就会产生很大的系统误差。创建统一数据库需要做许多预处理工作,比如坐标系统转换,统一数据格式,筛选出有效数据,评价数据质量等级等等,这个过程往往要占用整个项目历时的三分之一^[4]。

4.2 大数据分析在地质找矿中的应用

大数据分析重点在于从大量信息当中获取与成矿有关的特征信号,数据挖掘算法可以识别异常元素的组合模式,部分元素在特定地质背景下存在共生关系,这也许表明成矿作用的存在。空间聚类算法能够划分出成矿有利的区域,地质构造,岩浆活动以及流体通道在空间上的相互配合常常是成矿的必要因素。

4.3 AI 人工智能算法在地质找矿中的应用

算法的选择要考虑到地质问题的特别之处,神经网络预测模型适合解决连续变量的回归问题,矿体品位预测便属于此类场景。支持向量机分类算法合适处理二元判别问题,比如判断是否含有矿产,而整合学习算法经由多模型融合来加强预测的稳定性,依靠 AI 算法创建起来的建模流程包含特征工程,模型训练,超参数调整以及交叉验证这些环节,最后会输出成矿概率在空间上的分布状况。

4.4 成矿潜力预测模型

模型创建把东天山成矿带当作试验区,输入层融合地质,物探,化探,遥感这四类数据,一共包含 52 个特征变量,训练算法选取 XGBoost,并且运用带有地质限定的正则化项

来防止过拟合,输出结果是 $1\text{km} \times 1\text{km}$ 栅格的成矿概率分布图,经由已知矿床加以验证,其 ROC 曲线下的面积为 0.87,相比于传统专家评分法的 0.71 有了很大的改善。

4.5 深部隐伏矿预测突破

传统深部构造分析依靠人工反演重力剖面,效率低下而且极易受到解释者的经验左右,3D-CNN 反演技术把重力数据当作三维输入,经由卷积层自行获取深部密度异常特征,从而预测出东天山三处隐伏镁铁质岩体靶区,LSTM 用于处理磁测时间序列数据,改良磁源体的空间定位精度^[4]。

4.6 智能靶区优选

传统靶区优选依靠专家打分法,其中构造,岩浆,异常等要素的权重分配存在较强的主观性,DQN 强化学习算法经由模拟勘查决策流程,动态兼顾“找矿潜力”和“勘查成本”这两个目标,做到帕累托改进的靶区排序,GAN 生成对抗网络借助模拟成矿过程来验证靶区的地质合理性,并去除因算法偏差而产生的伪异常。

5 基于大数据与 AI 的新疆地质找矿应用案例分析

5.1 案例选取与数据准备

案例研究选定阿尔泰山南缘多金属成矿带作为试验区,此区域处于中国—哈萨克斯坦—蒙古跨境成矿带的东段延长线上,已识别出铜,铅,锌,金等多种矿产,已知矿床数量较为恰当,成矿类型丰富,数据汇集量在新疆各成矿带中位居前列,具备较好的代表性和可操作性,数据搜集工作融合了 1:5 万区域地质考察成果,1:20 万重力测量数据,1:10 万航磁测量数据,1:20 万水系沉积物地球化学测量数据以及 Landsat-8 多光谱遥感影像,时间范围覆盖 20 世纪 80 年代至 2023 年期间,数据总容量达 47GB。数据预处理在整个项目当中耗时最长,地质图数字化需执行图元矢量化及属性编码,物探数据要转为网格并用克里金插值法统一成 $500\text{m} \times 500\text{m}$ 分辨率,化探数据经标准化去除不同批次采样的系统误差,还要利用主成分分析法从遥感影像融合结果中获取铁染,羟基蚀变等找矿信息层,最后形成包含 52 个特征变量的空间数据库。

5.2 信息挖掘过程与结果

XGBoost 模型把研究区内 23 个已知矿床当作正样本,随机选取同样数量的未矿化区域作负本来执行训练,经过 500 次迭代之后,损失函数渐渐变得稳定,模型所输出的成矿概率分布图把研究区划分为五个级别,在高概率区域内(其概率值超出 0.75),总共存在三处,加起来大约是 180 平方千米。在这三个高概率区域当中,有两个正好和阿舍勒,可可塔勒这些已知矿床的空间位置相符合,这表明模型具备学习已知成矿规律的能力;第三个高概率区域处于当前勘查尚未涉足的克孜勒塔斯地区,这个地方以前仅仅做过区域性调查,并没有开展过系统的勘查活动。深部预测模块针对克

孜勒塔斯靶区展开分析,3D-CNN 反演重力数据显示,该区域下方 800 到 1200 米深度存在密度异常体,异常体体积约为 $2.5\text{km} \times 1.5\text{km} \times 0.4\text{km}$,这大概就是隐伏的镁铁质-超镁铁质岩体,结合地质背景和物探特征来看,这里属于 A 级找矿靶区。

5.3 应用效果评估

项目组想要验证模型预测是否可靠,于是就在克孜勒塔斯 A 级靶区布置了三个验证钻孔,这些孔的孔深都设计成 1000 米。钻探情况显示,ZK01 孔从 860 米到 920 米这段露出了辉长岩体,岩芯里能看见零星分布的黄铜矿化和磁黄铁矿化,铜的含量在 0.12% 至 0.28% 之间;ZK02 孔和 ZK03 孔没有明显的矿化现象,但是它们揭露出来的岩性组合类似于 ZK01 孔,这表明 3D-CNN 反演预测的隐伏岩体确实存在,不过铜的含量没有达到边界品位 0.3% 的工业标准,钻孔仅覆盖了异常体的边缘部分,核心区域还需要做进一步的验证,这样来看,模型的指示意义得到了初步的验证。应用过程中出现了一些问题,部分历史化探数据存在采样密度不均的问题,在数据稀疏区域,模型预测的不确定性显著增大。算法对于小样本矿床类型的学习能力比较有限,比如研究区内仅有 2 处的矽卡岩型矿床,存在漏判的风险,而且,深部预测的验证成本比较高,单孔钻探费用就超过 80 万元,大规模推广会遭遇经济方面的限制。

6 结论与展望

6.1 研究结论

大数据和 AI 技术在新疆地质找矿领域的应用已步入操作阶段,成矿预测模型的精准度得以提升,深部隐伏矿的识别取得超越,智能靶区的优选功能也得以达成,这些均表明该技术路线具备可行性和有效性。

6.2 研究不足与展望

当下研究存在一些局限之处,多源地质数据质量良莠不齐,小样本矿床类型建模较为棘手,深部预测验证成本偏高,日后的研究方向要着眼于多源数据融合算法的改良,采用小样本学习和迁移学习技术,并形成依靠数字孪生技术作支撑的虚拟验证体系。

参考文献

- [1] 丁恋. 基于绿色约束下勘探靶区优选的 AI+地质大数据找矿模型研究[J].中国矿业,2025,34(S2):155-158.
- [2] 王琼杰. 利用最低的成本、最快的速度、较高的准确率,为地质找矿提供选取和靶区 SR 智算一体机:引领地矿行业迈入 AI 新时代[J].资源导刊,2025,(S1):56-58.
- [3] 刘恒,韩媛,王伟,等. 促进新一轮找矿突破战略行动与生态保护的协同发展——AI+青藏高原地质信息化建设[J].科技创新与应用,2024,14(32):91-93+98.
- [4] 崔雪文,常亮,王冰,姜津珊,闫岩. 人工智能在自然资源管理与地质找矿中的前沿应用[J].地质论评,2025,71(S1):345-346.