

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 智能矿产潜力评价

左仁广\*

中国地质大学(武汉)地质过程与矿产资源国家重点实验室, 武汉 430074

一方面,矿产资源保障仍然是我国新时期高质量发展所面临的重大战略问题。另一方面,大数据带来了科研方法的变革,正成为科学发现新引擎,将改变地质学家的思维方式,为地质学的发展带来全新的面貌。近年来,大数据和深度学习在矿产勘查领域开始受到重视,使得矿产潜力评价进入智能化时代成为可能。智能矿产潜力评价是指基于大数据思维和机器学习(尤其是深度学习)对地质找矿大数据进行深度挖掘与集成融合,圈定找矿远景区并评价其资源潜力。它包括智能提取和智能决策两部分。智能提取主要是指从地质找矿大数据中自动地挖掘与矿床时空分布相关的信息。智能决策是指深度集成融合找矿信息并转化成知识,进行找矿决策。本文主要介绍了智能矿

产潜力评价的概念,分析了基于 GIS 智能提取岩体、构造、地层、断裂等与矿床时空分布相关的信息,以及基于深度学习(包括卷积神经网络、深度信念网络、循环神经网络、深度自编码网络和生成对抗网络)开展深层次找矿信息挖掘与集成研究进展,剖析当前基于深度学习开展智能矿产潜力评价存在训练样本少、超参数优化和模型收敛难,以及可解释性差等 3 大挑战与解决方案。智能矿产潜力评价未来应重点研究(1)如何把地质认知和成矿规律构建成深度学习的隐含层,实现深度学习模型的硬约束,提升深度学习模型的可解释性,以及(2)如何把深度学习模型可视化,进而分析智能决策背后的逻辑与推理,实现深度学习反哺成矿机理研究。

基金项目:国家自然科学基金面上项目(41972303)

第一作者简介:左仁广(1981-),教授,博士生导师,研究方向:数学地球科学与地学大数据. E-mail: zrguang@cug.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于深度学习的中沙群岛海山形态精细识别

张焱<sup>\*1</sup>, 杨永<sup>1</sup>, 周永章<sup>2</sup>

1. 广州海洋地质调查局, 广州 524057;
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, 广州, 510275

采用动态测量的多波束测深系统获取的水深数据值可能受到海浪起伏风流等海洋环境效应干扰, 进而影响海底地形形态的精度, 制约了对海山海丘的识别及形态刻画, 会影响到海底矿产资源的评价结果。针对这一问题, 本文拟分别基于中沙海域多波束的测深数据和三维地形数据, 采用分别具有非线性和特征学习能力的多重分形模型和深度学习方法对海底大数据进行更本质的刻画, 从而促进海山可视化或分类, 提高海山识别精度, 大数据手段可以避免受测量

及外界环境等条件因素影响而导致的水深数据精度问题。多重分形模型对中沙群岛海底地形进行识别可以有利的提取海山特征形态参数、分离出海山精细形态; 卷积神经网络模型可以获取未知海山信息, 提高海底地形图的分辨率和海山识别率, 并构建深度学习目标函数, 研发卷积神经网络新方法, 识别和圈定海山空间分布范围, 为后期矿产资源评价提供依据, 并为在复杂的海况环境下如何快速高效地识别海山提供新思路。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 动力学数值模拟算法及其并程序实现

何婷<sup>2</sup>, 章清文<sup>1</sup>, 刘耘<sup>1,2,3\*</sup>

1. 中国科学院地球化学研究所矿床地球化学国家重点实验室, 贵阳 550002;

2. 成都理工大学行星科学国际研究中心, 成都 610059;

3. 中国科学院比较行星学卓越创新中心, 合肥 230026

在地球和行星动力学领域, 由于难以对研究对象进行直接取样或就位观测, 且复杂地质活动具有非线性、多尺度、多场耦合的特性, 数值模拟发挥着不可替代的作用。尽管国内外可用超算资源 (详见全球超级计算机 TOP500 榜单: [www.top500.org](http://www.top500.org)) 日益丰富, 但相应的模拟算法研究及应用程序开发却十分滞后, 在地球动力学领域尤为明显。同时, 动力学模拟需要对复杂的非线性多场耦合数值模型 (尤其是高分辨率三维模型) 进行高效求解, 后者严重依赖大规模并行

计算, 现有程序对此往往支持不足。由此, 我们基于有限体积法和大规模并行技术自主开发了一套动力学模拟程序。该程序求解物质守恒、动量守恒 (即 Stokes 方程)、能量守恒方程、对流-扩散输运方程和其他控制方程。其中, Stokes 方程的求解基于 SIMPLE 算法, 而采用通量限制器的 TVD 格式处理各控制方程中的对流项, 使程序具有二阶收敛精度。该程序有望在早期地球和类地天体的构造演化和物质分异等关键过程的研究中发挥重要作用。

基金项目: 国家自然科学基金项目 (41804092)

第一作者简介: 何婷 (1997-), 硕士研究生, 主要从事计算地球动力学研究. E-mail: [heting6984@outlook.com](mailto:heting6984@outlook.com)

\*通信作者简介: 刘耘 (1968-), 研究员, 主要从事地球化学、天体化学研究

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 地质大数据与人工智能算法研究进展

周永章, 肖凡, 曹伟, 刘艳鹏, 余晓彤

1. 中山大学地球环境与地球资源研究中心, 珠海 510000;
2. 广东省地质过程与矿产资源探查重点实验室, 珠海 510000

最近十年, 数学地球科学的最显著发展是大数据与人工智能算法的引入。

地质数据以指数形式增长, 这是不容忽视的现实。基础地质、矿产地质、水文地质、工程地质、环境地质、灾害地质调查、勘查, 产生大量的数据。各类地基、空基对地遥感观测, 更产生了大量的数据。图件编绘、分析计算、模拟仿真、预测评价、管控调控, 同样产生大量的数据。并且这些数据可以是结构化的, 如地球化学分析和地球物理探查获得的数据; 更多是非结构化的、半结构化的。

在现实面前, 大数据挖掘和机器学习是地质学科跨越的必须选项。尽管依托大数据的人工智能地质学还远不成熟, 但已成为这个时代的绚丽浪花。

深度学习, 即多层神经网络的方法, 是一种实现机器学习的技术, 是过去几年大数据与数学地球科学研究的最重要的热点。贝叶斯网络是贝叶斯公式和图论结合的产物, 可用来建立矿床地质的成因网络, 进而理解矿床成因。地质大图形问题可以转化为大型的复杂网络空间问题和社区结构问题, 社区分析技术可用于地震预报、地质网络分析、特殊地质现象识别、矿床预测。关联规则和推荐系统算法在地质研究中已有成功的应用实例。化探数据及其异常经常包含复杂和非线性模式, 深度学习在智能识别与提取复杂地质条件下地球化学异常具有优异的能力, 卷积神经网络、堆叠自编码机等是较为常用和有效的方法。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于 GIS 的西藏同位素时空分析与找矿指示

张明明<sup>1,2,3,4</sup>, 李思洋<sup>1</sup>

1. 合肥工业大学 资源与环境工程学院, 合肥 230009;

2. 合肥工业大学 安徽省矿产资源与矿山环境工程技术研究中心, 合肥 230009;

3. 合肥工业大学 空间信息集成与综合分析平台, 合肥 230009;

4. 合肥工业大学 矿床成因与勘查技术研究中心, 合肥 230009

西藏高原地区的地质构造复杂, 矿产资源十分丰富, 目前该地区已积累了大量的地质年代情况、岩石特征、同位素、成矿元素特征等方面的数据。西藏高原地区的成矿过程与各个时期发生的地质构造运动有着密不可分的关系(李华健, 2017), 地质构造运动为板块地体带来了大量的成矿元素, 在板块地体交界处地球化学数据具有明显的起伏变化。因此, 通过同位素填图和异常地球化学数据分析能够反演板块构造变动带来的地质体成矿元素聚集(杨文采, 2017), 同时也能准确的反映成矿体在空间上的分布(侯增谦, 2015)。从地球化学数据与时空分析层次的结合来挖掘西藏高原地区的成矿潜力, 对于西藏高原的找矿采矿工作具有重要意义。

通过锆石 Hf 同位素的分析可以从时空上推测地壳的性质(Yan et al., 2015)。锆石 Hf 是一个样品的客观同位素平均值(Wang et al., 2017), 有利于开展大范围内的同位素填图。而全岩 Nd 同位素应用也十分广泛, 可示踪岩浆岩源区特点和估算源区的模式年龄(Wang et al., 2009)。首先通过系统分析锆石 Hf 和全岩 Nd 同位素组成, 确定其源岩组成特征和地壳模式年龄(TDMc)(王涛等, 2020), 然后制作岩浆

岩锆石  $\varepsilon_{\text{Hf}}(t)$ 、全岩  $\varepsilon_{\text{Nd}}(t)$  和  $T_{\text{DM}}^{\circ}$  等值线图(张立雪等, 2013), 反映地壳源岩空间分布和模式年龄空间变化, 依据锆石  $\varepsilon_{\text{Hf}}(t)$ 、全岩  $\varepsilon_{\text{Nd}}(t)$  和  $T_{\text{DM}}^{\circ}$  值变化, 推断不同地壳块体的空间展布和时间演变。同时将主要的多金属矿床也投在这些填图中, 以显示不同矿床与特定地质特征的空间关系。再结合多元统计分析, 进行合理的元素组合, 并通过元素组合反演出相似的地质作用过程

通过 GIS 分析结合地球化学数据和 Hf、Nd 同位素数据填图分析对研究区进行了找矿分析预测, 得出了以下结论: 西藏高原拉萨地体南部及西羌塘地体西部缝合带及断裂带两侧的  $\varepsilon_{\text{Hf}}(t)$  值、 $\varepsilon_{\text{Nd}}(t)$  值较高,  $T_{\text{DM}}^{\circ}$  年龄值较低, 对成矿元素的空间分布具有约束作用, 根据综合成矿特征, 划分了 3 处具有铜金找矿潜力的远景区, 对西藏高原铜金矿床的勘探具有重要参考意义。GIS 分析技术结合地球化学元素含量高的区域分析及矿点分布, 可获得不同层次的有利找矿区, 可为成矿带铜、金、钼、锌等多金属矿床找矿提供重要参考。通过同位素分析来快速圈定找矿战略靶区是一种可行且有效的方法, 开展地球化学元素与同位素填图, 可为现代找矿预测提供有力支持。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于知识驱动的铁铜多金属矽卡岩矿床三维远景建模 ——以安徽省繁昌盆地为例

李贺<sup>1</sup>, 李晓晖<sup>1,2\*</sup>, 袁峰<sup>1,2</sup>, 张明明<sup>1,2</sup>

1. 合肥工业大学资源与环境工程学院矿床与勘查中心, 合肥 230009;  
2. 合肥工业大学安徽省矿产资源与矿山环境工程技术研究中心, 合肥 230009

随着中国经济的发展,对矿产资源的需求日益增加。然而地表和浅层矿床的数量逐步减少,深部矿产勘查逐渐成为了找矿勘查的重点,与此同时,深部找矿预测随着深度的增加难度也随之上升。长江中下游成矿带是位于我国东部,发育多处火山岩盆地和大量铜铁金多金属矿床,成矿条件优厚。繁昌盆地位于长江中下游成矿带中段,其构造运动、地层条件以及岩浆活动与相邻的宁芜、庐枞等矿集区相似,但是针对该地区的找矿工作一直没有较大突破,仅在该地区浅部发现了几处中小型矿床,盆地深部有待开展进一步工作实现找矿突破。

矽卡岩类型的矿床在工业发展中扮演者重要的角色,是我国富铜矿和富铁矿的主要矿床类型,也是长江中下游成矿带的重要矿床类型。本研究利用

三维隐式地质建模方法,首先融合多源多维地质和地球物理解译成果,开展繁昌盆地三维地层、岩浆岩和构造模型构建研究;之后以繁昌盆地三维地质模型为基础,针对该区缺少已知矿化信息的特点,使用多级指数叠加、优劣解距离法、模糊综合评价三种知识驱动方法开展该地区的三维成矿定量预测,并对几种知识驱动方法的预测性能进行对比研究。结果表明:对于缺少已知矿化信息的地区,使用多级指数叠加、优劣解距离法、模糊综合评价方法均具有较好的成矿预测效果,圈定的有利靶区可为进一步的找矿勘探提供方向和依据;模糊综合评价方法预测效果优于优劣解距离法和多级指数叠加方法,在研究区矽卡岩型铁铜多金属成矿预测中具有更好的准确性和可靠性。

基金项目: 国家自然科学基金(41820104007)

第一作者简介: 李贺(1993-), 博士研究生, 研究方向: 数学地质研究. E-mail: 3014580365@qq.com

\*通信作者简介: 李晓晖(1986-), 副教授, 研究方向: 数学地质研究. E-mail: lxhlixiaohui@163.com

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于迁移学习和孪生神经网络的勘查地球化学异常识别

吴帮财<sup>1,2</sup>, 李晓晖<sup>1,2\*</sup>, 袁峰<sup>1,2</sup>, 李贺<sup>1,2</sup>, 张明明<sup>1,2</sup>

1. 合肥工业大学资源与环境工程学院, 矿床成因与勘查技术研究中心, 合肥 230009, 中国;
2. 安徽省矿床资源与矿山环境工程技术研究中心, 合肥 230009, 中国

地球化学异常的识别在矿产勘查中具有重要作用。地球化学背景与异常的划分可视为一个二元分类问题, 目前深度学习在各个分类领域都取得了很多的成果。相关研究表明, 使用深度学习方法能够提取地球化学数据的空间特征和内部关系, 从而有效识别地球化学异常。在地球化学异常识别的应用中, 矿化是罕见事件, 导致深度学习模型受到训练样本数量的限制, 需要新的改进算法来提高深度学习模型在异常识别中的可靠性。

本研究提出了一种利用迁移学习和孪生神经网络来减少对训练样本数量依赖的方法。首先, 采用基于 Inception-V3 网络的迁移学习方法, 提取与矿床相关的地球化学元素在不同尺度上的晕、重叠和空间分布特征。其次, 采用了孪生神经网络方法, 这是一种

具有相似性度量的特殊神经网络, 利用相似性度量函数确定任意两个样本之间的距离, 以评估它们之间的相似性。使用新疆北部准噶尔地区的地球化学数据进行实例研究, 结合这两种方法的模型准确率达到 85%, 表明改进的深度学习方法可以有效提高模型的异常识别能力。考虑到地球化学空间模式通常具有嵌套或分层的多尺度特征, 因此整合局部和区域地球化学数据可能会提高模型的性能。多尺度地球化学数据的融合使模型精度达到 88%, 说明该模型结构对不同尺度的地球化学异常信息也能够很好的提取和融合。因此, 深度学习可以有效地描述复杂的地球化学空间模式和潜在的异常, 并通过不断改进, 更好地应用于地球化学异常的识别。

基金项目: 国家自然科学基金项目 (42072321; 41820104007; 41872247)

第一作者简介: 吴帮财, 硕士研究生, 研究方向: 地理大数据分析. E-mail: 1766741721@qq.com

\*通信作者简介: 李晓晖, 副教授. 主要从事数学地质与 GIS 方面研究. E-mail: lxhlixiaohui@163.com

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 铜地球化学元素的高光谱数据挖掘研究

王珊珊<sup>1,2,3,4\*</sup>, 周可法<sup>1,2,3</sup>, 王金林<sup>1,2,3,4</sup>, 周曙光<sup>1,2,3,4</sup>

1. 中国科学院新疆生态与地理研究所, 荒漠与绿洲生态国家重点实验室, 乌鲁木齐 830011;
2. 新疆矿产资源与数字地质重点实验室, 乌鲁木齐 830011;
3. 中国科学院新疆矿产资源研究中心, 乌鲁木齐 830011;
4. 中国科学院大学, 北京

地球化学异常是找矿的重要标志, 其中铜的地球化学异常在地质勘探中有着至关重要的作用(张必敏等, 2016)。铜元素的地球化学异常主要与镁铁质岩和斑岩体有关, 也与共锌-铬-镍-铜组合的成矿元素有关。目前, 斑岩矿床和块状硫化物矿床是铜的最重要来源, 其中斑岩铜矿床提供了世界上近四分之三的铜(孙卫东, 2010)。随着高光谱遥感的发展, 通过高空间分辨率绘制矿物中元素吸收谱带的光谱特征, 为精确和详细地重建地球化学异常(表面)提供了新的手段。与传统地球化学勘探技术相比, 高光谱技术提供了快捷、廉价和无损的方法来表征地表矿化, 从而通过光谱特征识别元素含量(Greenberger 等., 2015; Van Ruitenbeek 等, 2012; 王晋年等, 2012)。这也为缺乏地球化学数据的地区找矿提供了数据和线索, 对找矿具有重要意义。

本研究利用光谱指数对新疆西准噶尔部地区的

斑岩矿区的铜元素含量进行了分析研究。根据不同 Cu 含量水平 (0~100 ppm, 100~1000 ppm, 1000~10,000 ppm 和 >10,000 ppm) 的岩石矿物和相对应含量的实测光谱反射率, 建立多种光谱指数。结果显示, 从 0.02~0.75 开始的所有指标的测定系数具有统计学意义。通过筛选光谱指数, 确定了归一化差异 (ND) 型光谱指数 (750~850 nm vs. 750~850 nm) 为最佳光谱指数和最佳反演波长范围, 其中对大于 10,000 ppm 高含量 Cu 的元素岩矿相关性最好 ( $R^2=0.63$ ,  $P < 0.005$ ), 表现最敏感, 且能够很好的反演岩矿中 Cu 元素的丰度。因此, ND 作为 Cu 元素含量的光谱指数, 提供了一个快速、准确和有效的方法来反演岩石矿物中不同含量 Cu 元素的地球化学异常。未来, 需要进一步研究岩石中铜元素与反射率、岩石类型、矿物学和矿物化学等岩石的特征之间的关系, 以便该指数更广泛的应用。

表 1 指数类型

指数	类型	含义 (中文名称)	公式
Reflectance (R)	反射率	反射率	$\rho_{\lambda 1}$
First Derivative (D)	导数	一阶导数	$\rho_{\lambda 1} - \rho_{\lambda 2}$
Simple Ratio (SR)	反射率	简单比值	$\rho_{\lambda 2} / \rho_{\lambda 1}$
Normalized Difference (ND)	反射率	归一化差值	$(\rho_{\lambda 1} - \rho_{\lambda 2}) / (\rho_{\lambda 1} + \rho_{\lambda 2})$
Modified Simple Ratio (mSR)	反射率	改良简单比值	$(\rho_{\lambda 1} - \rho_{\lambda 2}) / (\rho_{\lambda 1} + \rho_{\lambda 2} - 2\rho_{\lambda 3})$
Modified Normalized Difference (mND)	反射率	改良归一化差值	$(\rho_{\lambda 1} - \rho_{\lambda 3}) / (\rho_{\lambda 2} - \rho_{\lambda 3})$

基金项目: 自治区人才专项计划-天山优秀青年 (2019Q033)

第一作者简介: 王珊珊 (1982-), 副研究员, 研究方向: 定量遥感, 数学地质, 遥感应用研究. E-mail: wangshanshan@ms.xjb.ac.cn

\*通信作者简介: 王珊珊 (1982-), 副研究员, 研究方向: 定量遥感, 数学地质, 遥感应用研究. E-mail: wangshanshan@ms.xjb.ac.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 关键金属资源快速辨识技术研究

周可法<sup>1,2\*</sup>, 王金林<sup>1,2</sup>, 张楠楠<sup>1,2</sup>, 王珊珊<sup>1,2</sup>, 汪玮<sup>1,2</sup>, 张志欣<sup>1,2</sup>

1. 中国科学院新疆生态与地理研究所新疆矿产资源研究中心, 乌鲁木齐 830011;

2. 中国科学院大学, 北京, 100049

关键锂矿产资源是新能源、新材料、信息技术等新兴产业和国防军工等行业重要的金属原材料, 具有不可替代的重大用途。昆仑成矿带是研究和探测锂矿资源的天然实验室, 西昆仑成矿带出露大面积碱性花岗岩, 分布大量花岗伟晶岩脉, 是重要稀有金属成矿带, 具有巨大的锂资源成矿潜力。昆仑成矿带锂矿资源研究面临重大挑战(李荣社等, 2007; 陈守建等, 2008; Zhang et al., 2018); 大数据技术具有识别低丰度弱缓信息的显著优势(zhou et al., 2020)。西昆仑大红柳滩一带分布有大量花岗伟晶岩脉, 是我国传统的稀有金属成矿带。自 1958 年发现阿克塔斯锂矿以来, 其外围一直没有大的找矿突破(李荣社等, 2007; 莫宣学, 2007; Hu et al., 2016); 由于地质演化的漫长性、阶段性、复杂性及各种地质作用的影响, 地质大数据存在着参数信息不完全、结构信息不完全、关系信息不完全和演化信息不完全的特征, 并呈现异构多模态、复杂关联、动态涌现等特点(甘甫平等, 2018; 吴冲龙等, 2016)。大数据技术的应用为成矿过程模拟、找矿信息挖掘和矿产预测评价带来机遇(赵鹏大, 2015; 周永章等, 2017; 张旗和周永章, 2017)。

本研究从不同空间尺度(航天、航空、地面)、不同分辨率和不同精度遥感数据来识别和提取锂铍矿蚀变矿物及相关元素, 进行伟晶岩锂铍矿信息提取, 建立了地面、航空和航天光谱伟晶岩锂铍矿天-空-地成像高光谱探测技术; 确定伟晶岩锂铍矿的分

带及矿物组合, 构建了伟晶岩锂铍矿的图谱数据库, 通过数据挖掘技术分析伟晶岩锂铍矿的分带及矿物组合的波谱特征(强度、整体形态、吸收特征等), 筛选出伟晶岩锂铍矿稳定、敏感响应特征谱段, 构建了伟晶岩锂铍矿分带及矿物组合的图谱识别技术; 理清构造、岩性、图谱和模式(结构-纹理-色调)四大特征, 建成了伟晶岩锂铍矿图谱特征识别技术体系, 通过波谱特征区分花岗岩、含锂辉石伟晶岩、不含锂辉石伟晶岩的信息。

研究得出 1413nm、1900nm 和 2207 nm 三处吸收特征, 是区分含锂辉石伟晶岩、不含锂辉石伟晶岩及围岩的波谱吸收特征, 1900 nm 的特征吸收可以很好地含锂辉石伟晶岩与不含锂辉石伟晶岩及围岩区分, 是含锂辉石伟晶岩的指示波段; 图谱识别技术中, 辨识含矿岩体围岩颜色为黑黄色, 露头面积较大, 层理较为明显; 花岗岩颜色为亮白色, 纹理较为粗糙, 露头面积较大, 与周边其它岩性界线较为明显; 伟晶岩呈白色、灰白色色调和条带状影纹特征, 形体一般呈现不规则形状, 边角光滑, 呈方向性, 得出伟晶岩的反射率最高, 花岗岩次之, 而片岩和千枚岩的反射率最低; 图像特征中含锂辉石与锂云母的花岗伟晶岩脉呈淡蓝色, 黑云二长花岗伟晶岩呈亮白-粉白色, 识别出的浅蓝绿色为含绿柱石与锂辉石的花岗伟晶岩脉。开展关键金属矿产智能辨识技术研究具有重要意义, 可以促进交叉学科与新兴学科发展。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于监督学习算法在阿舍勒地区的找矿预测

郑超杰<sup>1,2</sup>, 袁峰<sup>1\*</sup>, 罗先熔<sup>2</sup>, 刘攀峰<sup>2</sup>, 文美兰<sup>2</sup>

1. 合肥工业大学 资源与环境工程学院 合肥 230009;

2. 桂林理工大学 地球科学学院 桂林 541006;

随着国内经济的高速发展,我国在基础地质及矿产勘查等方面投入了大量的研究经费,积累了大量宝贵的、种类繁多的基础地质数据;伴随着人工智能和机器学习研究的兴起,机器学习算法在矿产资源评价中的应用得到了格外的重视(左仁广等, 2021)。其中,监督学习算法通过对由待测研究区地质信息构成的训练、测试、验证集进行建模,能够识别数据集中控矿地质要素与见矿钻孔间存在内在模式关联,并将未知区存在相同/类似模式的控矿地质要素的区域,认定为有利找矿地段。由该类算法构建的预测模型具备较好的成矿预测效果,因此常被运用于矿产资源预测评价中(De Boissieu F. et al, 2018; Zuo R. et al, 2019; 李苍柏等, 2020)。新疆阿舍勒铜锌矿位于阿尔泰造山带西南缘阿舍勒盆地内,是典型的 VMS 型

矿床(杨富全等, 2016)。鉴于阿舍勒铜锌矿床具有埋藏深、开采难度大、采矿维护成本高等特点,伴随着矿山开采对探明资源储量不断消耗,深边部矿体品位下降,对阿舍勒铜锌矿床深部及外围找矿任务已迫在眉睫。基于此,本研究选用支持向量机、随机森林及 K 邻近分析三种监督式学习算法构建机器学习模型,对阿舍勒铜锌矿区进行找矿预测。结果表明,三类基于监督式学习算法构建的预测模型具备较高的精确度和召回率,各类模型预测结果与矿区地质特征吻合度高,并在新疆阿舍勒铜锌矿区外围指示出多个高潜力成矿区域,待进一步工程验证;相比常规方法圈定的找矿靶区,运用机器学习方法进行的找矿预测研究具备更高的准确性、客观性及合理性,可将其应用到今后的找矿预测工作中。

基金项目: 地电化学方法成晕机理及技术改进研究与应用示范(2016YFC0600603)

第一作者简介: 郑超杰(1993-), 博士, 从事勘查地球化学、成矿预测方向的研究. E-mail: zcj@glut.edu.cn

\*通信作者简介: 袁峰(1971-), 教授, 博士生导师, 从事成矿作用、成矿规律、成矿预测等方向的研究

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

# The Construction and Application of Knowledge Graph Method in Management of Soil Pollution in Contaminated Sites——Case Study in Guangdong Province, China

## 城市土壤污染知识图谱的构建与应用研究——以广东省为例

Han Feng<sup>1</sup>, Zhou Yongzhang<sup>1</sup>, Cheng Guangliang<sup>2</sup>, Bian Jing<sup>2</sup>, Zhang Qianlong<sup>1</sup>

1. School of Earth Science and Engineering, Sun Yat-Sen University, Zhuhai, 519082, China;

2. School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, 510006, China

Soil pollution of construction land is typical data-intensive issue, since the types of pollutants are various. Secondly, there are many sources and ways, and many related factors. And the secondary diffusion under natural conditions will form a wider range of contamination. Furthermore, the data collected in the process of soil pollution monitoring, prevention and early warning is huge with the features of multi-source heterogeneity, fuzzy relationship and numerous association.

Therefore, scientific analysis and management of urban soil pollution is an extremely complex task. It is necessary to organize the data, relationships, rules, logic, knowledge and models of urban soil pollution process into multi-scale, multi-temporal database to construct the models of urban soil pollution target at intelligent monitoring, simulation, early warning and control.

Knowledge Graph (KG for short) has the great advantages of obtaining structured knowledge from

massive data to realize unified expression and efficient storage, and is an effective means to solve the problem of management of massive data in soil pollution process.

This study selected over 2000 sites from the 14 municipal cities in Guangdong as the research Cases. The research work of this article includes data mining and extraction, ontology model building and knowledge graph construction. this study transforms the information, collected from investigating report, governmental official document, yearbook, and related website, into the format of property graph to promote the readability and computability for computer and intuitiveness for human.

In the end this study applied the constructed KG with the function of pollution information query, relationship reasoning and knowledge update aim to explore the application of KG technology in soil pollution prevention, control and risk assessment in urban area

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 钦-杭成矿带斑岩铜多金属矿床智能找矿研究

余晓彤<sup>123</sup>, 周永章<sup>123\*</sup>, 张前龙<sup>123</sup>

1. 中山大学地球环境与地球资源研究中心, 广州, 510275;

2. 中山大学地球科学与工程学院, 广州, 510275;

3. 广东省地质过程与矿产资源探查重点实验室, 广州, 510275

前人研究发现, 钦-杭成矿带沿线广泛分布斑岩铜多金属矿床 (毛景文, 2011), 是国内重要的斑岩-矽卡岩铜多金属矿成矿区。这些斑岩铜多金属矿床成矿时间大致都处于中晚侏罗纪, 可能是受古太平洋板块俯冲到华南板块之下的影响形成的 (毛景文, 2011)。之前一般是通过水系沉积物地化数据的综合分析来寻找圈定找矿远景区 (余晓彤, 2019; 王堃屹, 2019), 但这种方法一般是在小区域的找矿勘察上, 在大范围的构造区域上进行找矿评价就有些捉襟见肘。与此同时也忽略了构造对成矿的影响, 而同位素和稀土元素的结合似乎更有利于定位特定类型矿床的成矿构造, 因为他往往能够反映同期次的构造运动和成矿物

质来源进而发现由于其影响而生成的一系列同类型的矿床, 从而对圈定找矿远景区提供帮助。例如在中亚造山带利用  $\epsilon\text{Nd}$  与  $\ln(\text{La}/\text{Yb})$  的比值区别出了 Cu-Mo 矿床和 Cu-Au 矿床产出的区位 (Wu, et.al, 2021)。

本研究应用深度学习的方法选取合适的同位素评价指标, 将 GEOROC 数据库以及前人研究中的钦-杭成矿带斑岩铜多金属矿床及花岗岩中的同位素和稀土元素的数据提取出来作为研究对象, 运用深度学习算法定位钦-杭成矿带斑岩铜多金属矿床的成矿构造, 由此智能圈定钦-杭成矿带该类型矿床的找矿远景区。同时也为探寻该类型矿床的形成机制和成因及建立矿床模型提供了一定的方向。

基金项目: 国家自然科学基金 (U1911202)

第一作者简介: 余晓彤 (1994-), 博士研究生, 研究方向: 地球化学、智能找矿. E-mail: yuxt3@mail2.sysu.edu.cn

\*通信作者简介: 周永章 (1963-), 教授、博士, 研究方向: 地球化学、大数据与数学地球科学. E-mail: zhouyz@mail.sysu.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 四川省阿坝甘孜森林火灾风险评估及预测

李兵, 陈翠华\*

成都理工大学 地球科学学院, 成都 610059

森林火灾是森林生态系统的主要干扰因素, 不仅会破坏生态系统的动态平衡, 还会影响人类的生存环境, 给人类社会造成巨大经济损失。阿坝州, 甘孜州位于四川西北部, 森林资源丰富, 其森林覆盖率位于全国前列。阿坝州, 甘孜州是我国森林火灾频繁发生地区, 也是森林火灾的重点防范区, 每年会发生规模较大的森林火灾, 给森林资源、生态环境和社会经济等造成巨大影响(张翔, 2020)。研究阿坝州, 甘孜州森林火灾的时空分布特征和驱动因素, 预测森林火灾的发生和对森林火险进行等级区划, 是有效提高阿坝州, 甘孜州防灾减灾水平的关键, 为防火工作提供决策依据。

在电子科技大学团队的研究下, 已经通过对影响森林火灾的三大影响因素下, 做出了四川阿坝甘孜的以天为单位的火灾风险评估及火灾预测, 基于深度学习模型计算森林火灾风险, 以中国西南地区为例, 实现了对森林火灾风险在空间上的定量表征。基于 Logistic 回归模型计算野火风险指数, 实现对野火风险程度的定量表征, 内部验证结果表明模型具有很好的评估性能, 而外部验证结果表明当模型在时间维度上扩展时, 评估性能仍然卓越(文崇波, 2019)。基于深度学习模型计算森林火灾风险, 模型以四川省历史森林火灾事件及诱发因子数据库及其对照数据库为基础, 此外利用 Logistic 回归模型作为传统方法进行比较研究。验证结果表明构建的深度学习模型对森

林火灾风险具有很好的评估性能, 且性能好于 Logistic 回归模型(骆开苇, 2020)。

以阿坝甘孜为研究区, 基于电子科技大学团队的风险评估产品得到的将可燃物、天气、地形结合所得到的火灾风险因子进行以月为周期的火灾风险评估, 及应用长短期记忆模型 LSTM 模型与深度学习模型 ARIMA 模型。评估不同模型的预测精度, 主要的工作及结果可以概括为:

通过火险产品得到的是四川省阿坝甘孜的 70 多万个像元的火险值, ①通过 arcgis 栅格转点, 得到一个月的火险数据, 以火灾风险分级分为不同级别, 再点转栅格得到月尺度的火险等级区划图。②运用 python 编程实现了对它的定级, 结果相同。通过 python 编程实现分级更简单, 可以批量操作。最后得到的风险等级图与实际很接近。

在研究初期, 笔者是以 120 天的数据作为实验数据, ①首先利用 ARIMA 模型进行预测, 实现 ARIMA 预测也利用了 spss 软件和编程代码实现, 在实验过程中发现 ARIMA 模型中的 p,q 值选取存在一定困难, 最后预测出的结果与实际对比存在一定误差。实际结果与预测结果出现滞后现象。②利用 LSTM 模型, 采用 75% 的样本为训练集, 25% 的样本为验证集。得到的结果与实际误差很小。所实验的数据的性质更适合 LSTM 模型, 更符合长短期的要求。

第一作者简介: 李兵 (1996-), 硕士研究生, 研究方向: 资源与环境 (地质工程)。E-mail: 754385318@qq.com

\*通信作者简介: 陈翠华 (1972-), 教授, 博士生导师, 主要从事矿相学, 地球化学工作。Email: chencuihua@cdut.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 大兴安岭森林火灾风险评估及预测

尹衍亮, 陈翠华\*

成都理工大学, 地球科学学院, 成都 610059

森林火灾是指失去人为控制, 在林地内自由蔓延和扩展, 对森林资源、森林生态系统和人类带来一定危害和损失的林火灾害。森林火灾是全球性、突发性、破坏性极强的自然灾害, 森林火灾一旦发生, 则很难对林火进行控制和扑救, 因此需要及时有效的对森林火灾的发生进行预测预防。森林火灾预测预报研究, 可以及时有效的预防林火的发生, 有效地保护森林资源, 最大限度的减少林火灾害造成的影响和后果, 是当前森林火灾研究的重要内容。

结合已有森林火灾天气等级模型, 森林火灾预测模型研究与应用(于茜, 2019)中首先对森林火灾气象因子和植被因子数据的获取及处理方式进行了设计, 获取地面气象台站观测的日最高温度、日最小相对湿度、降水量及最大风速的逐时数据, 和卫星影像植被覆盖度数据, 对两种类型数据进行处理计算, 并对全国火灾案例进行火灾等级验证, 构建森林火灾预测模型, 结合 Matlab 开发语言, 研发森林火灾预测系统, 自动解算研究区森林火灾等级分布和各影响因子分布结果, 为森林火灾预测模型的应用奠定基础。然而, 在国家尺度内(如中国)或更大尺度内的森林火灾预警预测研究主要面临三个挑战:(1)由于无法在大尺度下时空连续地获取可燃物含水率信息以及缺乏其与森林火灾的关系研究导致森林火灾诱发因子数据库不完整;(2)森林火灾风险时空挖掘方法的欠缺使得大尺度的森林火灾风险评估的准确性有待提高;(3)像素级的高精度森林火灾风险预警预测方法亟待探索。

根据论文的研究目的, 电子科技大学复杂环境定

量遥感团队拟采用 Logistic 回归模型计算野火风险指数(Wildfire Risk Index, WRI)。野火风险指数是对野火风险程度定量化的表征, 取值在 0~1 之间变化, 值越大代表着野火风险程度越高。采用野火风险指数的目的在于通过对野火风险程度定量化的表征, 取代传统的野火风险评估分类规则, 比如低风险、中风险、高风险等十分模糊的评估描述。

本研究基于 2020~2021 的全球森林火灾指数(FDI)产品(时间分辨率:8 天, 空间分辨率:500 米), 对同一像元在 2020~2021 年的 FDI 进行时间序列模型建模, 并进一步预测 2021~2022 年森林火灾趋势。这符合 ARIMA 模型的基本思想即根据序列的过去值及现在值对序列的未来值进行预测。ARIMA 模型全称为差分滑动平均自回归模型(Autoregressive Integrated Moving Average Model, 记为 ARIMA(p, d, q))。现阶段, 笔者将近 90 天的大兴安岭 70 多万个像素点作为一个三维数组, 每个像素点都将有一个月的数据进行 ARIMA 模型训练, 并且对于缺失值可以利用 SPSS 进行补充修正, 将未来的 7 天进行预测, 并与实际情况相比较, 测试出预测的准确性。再利用 Python 进行 auto-ARIMA 模型对大兴安岭森林火灾的自动预测, 使整个预测过程更加简便快捷有效。

因此本研究建立了森林火灾风险时空挖掘及预警预测方法体系, 为森林火灾预警预测系统的开发提供新的思路和方法, 促进地方消防管理规划和基于地理位置的资源配置优化, 并贡献于森林火灾风险预警, 抑制, 和响应, 以及提高人们对生命财产的安全意识。

第一作者简介: 尹衍亮(1996-), 硕士研究生, 研究方向: 资源与环境. E-mail: 504970457@qq.com

\*通信作者简介: 陈翠华(1972-), 教授, 主要从事矿床学、矿相学、地球化学研究工作. E-mail: chencuihua@cdu.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 地质云立体交互终端建设与应用示范

焦守涛<sup>1,2\*</sup>, 朱月琴<sup>1,2</sup>, 刘荣梅<sup>1,2</sup>, 高振记<sup>1,2</sup>, 袁玲玲<sup>3</sup>, 郭艳军<sup>4</sup>, 洪韬<sup>4</sup>

1. 中国地质调查局发展研究中心, 北京 100037;
2. 自然资源部地质信息工程技术创新中心, 北京 100037;
3. 中国地质科学院地质研究所, 北京 100037;
4. 北京大学地球与空间科学学院, 北京 100871;

地质云立体交互终端是一个基于人机立体交互的平台, 包含了一系列硬件设备和软件, 是地质云平台的组成部分, 该平台的建设丰富了地球系统科学中地质作用、地质演化过程和地学数据可视化的方式。扩展现实 XR (Extended Reality) 涵盖了虚拟现实 (VR, Virtual Reality), 增强现实 (AR, Augmented Reality) 和混合现实 (MR, Mixed Reality) 等技术, 以及其它相关配套技术, 真正实现了人机交互, 是互联网、物联网和人机交互技术的完美融合。地学数据通过扩展现实技术可视化, 结合适宜的引擎工具进行

工程开发, 在不同的平台进行部署展示的技术路线, 将是未来立体交互可视化的一个重要发展方向。

本文以微软的 HoloLens2 MR 头显为基础, 构建了三维的中国电子沙盘, 并与雄安新区的地质三维模型进行了融合, 达到了多机协同立体交互的可视化效果。

地学领域扩展现实技术的应用可以更好的实现几何维度的地球深部信息和时间维度的深时信息的动态可视化模拟表达和交互, 可为未来地质大数据、全空间、高维度的全信息地学知识表达、知识发现和知识服务提供技术支撑。

基金项目: 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)、《基于“地质云”平台的深部找矿知识挖掘》(2016YFC0600510)、国家自然科学基金面上项目《大数据环境下的滑坡危险性评估模型构建方法研究》(41872253) 和地调项目《国家地质大数据汇聚与管理(中国地质调查局发展研究中心)》(DD20190318)

第一作者/通信作者简介: 焦守涛 (1988-), 助理研究员, 研究方向: 地质信息化和地质大数据研究. E-mail: jiaoshoutao88@163.com

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于循环一致生成对抗网络的三维数字岩石重构效果分析

闫伟超<sup>1, 2, 3\*</sup>, 蒋熙梅<sup>1, 2, 3</sup>, 邢会林<sup>1, 2, 3</sup>, 迟蓬<sup>4</sup>, 孙建孟<sup>4</sup>

1. 深海圈层与地球系统教育部前沿科学中心, 海底科学与探测技术教育部重点实验室, 中国海洋大学海洋地球科学学院, 山东青岛 266100;

2. 青岛海洋科学与技术国家实验室深海多学科交叉研究中心, 山东青岛 266237;

3. 中国海洋大学海底科学与工程计算国际中心, 山东青岛 266100;

4. 中国石油大学(华东)地球科学与技术学院, 青岛 266580

三维数字岩石是表征地下岩石三维微观孔隙、矿物等结构特征的虚拟仿真技术, 利用构建的数字岩石模型可以准确的模拟岩石孔隙度、渗透率、电阻率等岩石物理特性。目前, 构建三维数字岩石主要采用 CT 扫描的方式。然而, 由于岩石孔隙结构的多尺度特征, 单一 CT 扫描分辨率的数字岩石图像难以完整描述其特征。尽管多分辨率 CT 扫描技术能有效解决岩石组分结构的多尺度特征, 但多分辨率 CT 扫描价格昂贵, 且仅对某一块岩石有效, 无法将多尺度信息拓展到采集的全部岩石样品。针对这一问题, 本研究提出了一种基于循环一致生成对抗网络(CycleGAN)的多源多尺度三维数字岩石重构方法, 旨在充分融合低分辨率图像的大视场优势和高分辨率图像的高精度特征。为

了验证该方法的准确性, 本研究使用了两块不同类型的砂岩和碳酸盐岩的 CT 图像, 并将相对均质的 Berea 砂岩的低分辨率 CT 扫描图像和 SEM 高分辨率扫描图像进行了多尺度图像融合, 构建出具有分辨率提高不同倍数的三维数字岩石, 并进行了验证。研究结果表明, 使用不同分辨率的图像进行融合可以构建出具有较高准确性的多尺度数字岩石, 其统计特性和岩石物理属性方面与真实结果温和较好。同时, CT 扫描图像之间可实现相差 4-8 倍的多尺度信息融合, 而 CT 扫描图像和 SEM 扫描图像之间可实现相差 4-16 倍的多尺度信息融合。因此, 基于 CycleGAN 的多源多尺度数字岩石重构方法在多尺度数字岩石建模和岩石物理属性数值模拟方面具有广阔的应用前景。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于机器学习计算含水瓦兹利石的热导率

王冬<sup>1</sup>, 吴忠庆<sup>1\*</sup>, 邓忻<sup>1</sup>

1. 中国科学技术大学地球和空间科学学院 合肥 230026

地幔通过热传导的方式不断地对俯冲板片进行加热, 俯冲板片的热导率可以通过控制板片的温度结构而影响一系列的俯冲过程。俯冲板片将其中赋存的水输运至地幔深部, 最近的研究表明水的存在可以显著地降低橄榄石和林伍德石等矿物的热导率。而瓦兹利石作为俯冲板片中的重要含水矿物, 水对其热导率的影响尚不清楚。高温高压下测量矿物热导率仍面临着巨大挑战, 基于第一性原理计算得到的三阶力常数被广泛用于计算材料的热导率。然而对于某些材料, 更高阶的力常数是不可忽略的, 但其计算量又十分巨大。通过分子动力学计算热导率的方法考虑了所有非谐相互作用, 但面临着有限尺寸效应, 足够大的模拟体系才能产生可靠的结果。第一性原理分子动力学由于计算量的限制只能模拟数百个原子。基于第一性原理数据训练出来的机器

学习势能模型在基本保持第一性原理计算精度的同时大大降低了计算成本, 可以帮助我们进行长时间大规模的分子动力学模拟。

本研究中, 我们利用第一性原理数据训练了瓦兹利石的机器学习势能模型, 并结合非平衡分子动力学模拟获得了高温高压下无水以及含水条件下瓦兹利石的热导率。无水瓦兹利石的计算结果与现有实验十分吻合, 我们发现水的存在有效地降低了瓦兹利石的热导率, 并且由于橄榄石和瓦兹利石的晶格优选定向和热导各向异性, 热流可能沿着具有较低热导率的方向流向板片。通过对板片温度结构数值解和解析解的分析, 表明水的存在和热导各向异性有效地降低了板片的温度, 使得其中亚稳态矿物以及含水矿物的稳定深度增加, 提高了板片输运水的能力以及深源地震的最大发震深度。

基金项目: 国家重点研发计划 (批准号: 2018YFA0702703), 国家自然科学基金 (批准号: 41925017)

第一作者简介: 王冬 (1999-), 男, 硕士研究生, 研究方向: 高温高压矿物物性研究. E-mail: wangdong01@mail.ustc.edu.cn

\*通信作者简介: 吴忠庆 (1974-), 男, 教授, 研究方向: 高温高压矿物物性研究. E-mail: wuzq10@ustc.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于机器学习方法的地层粒度分布预测

蒋熙梅<sup>1, 2, 3</sup>, 闫伟超<sup>1, 2, 3\*</sup>, 邢会林<sup>1, 2, 3</sup>, 孙建孟<sup>4</sup>

1. 深海圈层与地球系统教育部前沿科学中心, 海底科学与探测技术教育部重点实验室, 中国海洋大学海洋地球科学学院, 山东青岛 266100;
2. 青岛海洋科学与技术国家实验室深海多学科交叉研究中心, 山东青岛 266237;
3. 中国海洋大学海底科学与工程计算国际中心, 山东青岛 266100;
4. 中国石油大学(华东)地球科学与技术学院, 青岛 266580

地层粒度分布是指地下不同地层中, 颗粒物质在其尺寸或粒径上的分布情况。地层粒度分布对于地质学、石油工程、地下水资源研究等领域都有着重要的意义。在地质学中, 地层粒度分布是研究岩石成因、沉积环境、地层演化等方面的基础数据。通过对地层中颗粒物质的尺寸、形态、成分等进行分析, 可以了解该地层的沉积环境和历史, 进而推断出区域地质历史和资源潜力。粒径是颗粒的一个重要特征, 它可以用不同的方法测量, 如光学显微镜、电子显微镜、激光散射等。然而, 粒度分布的测量需要耗费较多的时间和人力物力成本, 特别是对于大批量的样品测量来说, 时间和成本的消耗较大, 且在一些井段中由于取芯收获率低, 导致获得的粒度数据在深度上不连续等局限性。为了解决这些问题, 结合测井曲线和机器学习方法可以节约时间和成本, 并实现连续深度预测地层粒度分布的目的。目前已有很多关于使用机器学习方法进行地层粒度预测的研究, 但这些研究都只是针

对粒度特征值进行预测, 属于一维特征分析, 而单一的粒度特征值不能完全反映问题, 也会影响对岩性、沉积环境和水动力条件的判断。因此, 本研究采用了自动超参数优化框架和梯度提升算法相结合的机器学习方法, 将其应用于渤海湾盆地某区块的测井数据。本研究测试了线性回归、支持向量机、K-最近邻、随机森林、梯度提升决策树和卷积神经网络等常规机器学习方法, 以自然电位、声波、井径、补偿中子、自然伽马、地层真电阻率、深侧向电阻率、微侧向电阻率和浅侧向电阻率等 9 种测井参数作为输入数据, 以二维粒度分布作为输出数据进行测试。通过对决定系数的大小和均方根误差的分析, 选出了最佳模型, 用于预测实际井的连续深度二维粒度分布。对比分析证明, 自动超参数优化框架和梯度提升算法相结合的机器学习方法对地层粒度分布预测的精度明显高于传统的机器学习方法预测效果, 该研究成果对其他地区地层的二维粒度分布的预测具有借鉴意义。

基金项目: 国家自然科学基金(页岩油储层核磁共振响应机理与储层参数评价方法研究, 42004098)

第一作者简介: 蒋熙梅(2001-), 女, 研究方向: 测井解释. E-mail: 1520937159@qq.com

通信作者简介: 闫伟超(1991-), 男, 副教授, 研究方向: 数字岩石与测井解释. E-mail: yanweichaoqz@163.com

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于汇水盆地和关联规则算法的地球化学异常识别及其找矿预测——以冀北新拔地区 1: 5 万水系沉积物测量为例

梁鸣<sup>1</sup>, 罗先熔<sup>1</sup>, 文美兰<sup>1</sup>, 欧阳菲<sup>1</sup>, 高文<sup>1</sup>,  
刘攀峰<sup>1</sup>, 王晓东<sup>2</sup>, 陈皓<sup>2</sup>, 石佳磊<sup>1</sup>, 孙瑶<sup>1</sup>

1. 桂林理工大学 地球科学学院 隐伏矿床预测研究所, 桂林 541006;

2. 河北省地矿局 第四地质大队, 承德 067000

水系沉积物地球化学测量是寻找表生矿、半隐伏矿和难识别矿种的重要手段(罗先熔等, 2007; 龚庆杰等, 2020), 通常应用于大面积的地质扫面工作(王学求, 2013)。它具有相对明显的优缺点: 在中、小比例尺调查时, 能够较为准确地圈定出异常, 并对原生异常的源头进行初步追索(孔焜辉, 2020); 然而, 大比例尺调查时, 圈定的异常易受采样点上游岩性影响和沟谷地形地貌特征的制约; 同时, 水系沉积物采样介质通常会随水系的方向进行迁移, 圈定的异常在空间上与其源头脱节, 真正的找矿目标可能在上游的汇水区(赵阳等, 2021)。因此, 在大比例调查时, 绘制水系网络并生成汇水盆地系统, 是寻找真实异常源的关键技术, 加强分析基于汇水盆地的数据组成和结构对于提取矿致地球化学异常至关重要(黄啸坤等, 2021)。

新拔地区位于河北省承德市围场县西北部, 属内蒙一大兴安岭褶皱系的内蒙华力西晚期褶皱系棋盘山中凹陷构造单元内。区域内构造-岩浆活动强烈, 具有多期次成矿作用叠加的特征, 是冀北重要的成矿地段(梁鸣等, 2022)。对研究区进行微地球化学景观划分, 大部分区域处于半干旱中浅切割中低山区, 海拔在 1100~1800 米, 沟谷发育, 切割侵蚀强烈, 保障水系沉积物样品具有较高的质量。

本文对冀北新拔地区 1: 5 万水系沉积物测量数据进行挖掘和再开发。研究步骤为: (1) 选用新拔地区等高线数据, 转化为 DEM 数据; 通过水文分析方法, 划定分水岭, 以分水岭为流域边界, 构建研究区汇水盆地模型; (2) 计算研究区各地质体的加权平均含量和单个汇水盆地背景值, 进行残差校正; 从而确定单个汇水盆地的局部背景值; (3) 对剔除局部背景值的数据进行“七级异常划分”(Gong 等, 2018), 绘制单元素异常图, 评价单元素异常特征; (4) 运用关联规则算法探讨元素间的高阶相关性(刘心怡等, 2019), 构建指示找矿的成矿元素关联规则数据库。

研究目的是构建一套客观、局部背景值不固定、分布无关、多元的异常识别和找矿预测方法。尝试应用“汇水盆地法”和“七级异常划分”方案将基于空间识别和频率识别的不同数据编制技术结合起来, 更为精密的圈定单元素异常; 应用“关联规则算法”挖掘数据的结构特征, 厘清地质极端事件中元素的高阶相关关系, 以此为依据分离出异常盆地; 并对区域内有相同或相似高阶相关关系的汇水盆地进行找矿预测。后续异常查证阶段, 共发现 3 条 Ag 矿体、2 条 Zn 矿体、1 条 Au 矿体、1 条 Pb 矿体和多条矿化蚀变带, 表明该方法在新拔地区具有较好的适用性。

基金项目: 广西科技计划青年基金项目(2022GXNSFBA035548; 2021JJA150037); 广西高校中青年教师科研基础能力提升项目(2022KY0261)

第一作者简介: 梁鸣(1993-), 硕士研究生, 研究方向: 勘查地球化学。E-mail: 2783213586@qq.com

\*通信作者简介: 罗先熔(1953-), 博士生导师, 教授, 研究方向: 矿床学及地球化学。E-mail: lxr811@glut.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于机器学习构建富硒土壤预测模型——以江西省信丰县为例

杨兰, 王运, 赵碧波, 张安

东华理工大学 地球科学学院, 江西 南昌 330013

硒被称为“长寿元素”, 是人体健康必需的微量元素之一[1]。中国乃至世界大部分地区都是缺硒的, 硒元素被加工成为各种营养品, 以弥补人体所需的硒元素。随着中国人民生活水平的不断提高, 人们更加追求健康长寿。目前, 全国掀起了寻找富硒土壤的浪潮。土壤中的 Se 与地球化学元素、pH 等之间关系密切[2], 因此可通过这些指标基于机器学习构建富硒土壤预测模型[3-4]。

为了寻找富硒土壤, 开发富硒农产品, 中国部分地区已完成了 1:5 万土地质量地球化学调查, 圈定了大量富硒土壤, 积累了几百万个土壤化学元素数据。然而这些大量数据, 通常仅作简单的特征描述及相关性分析, 很少开展深入研究, 忽视了数据的宝贵价值。另外, 土壤找矿地球化学调查、耕地地球化学调查及环境监测等工作, 积累了大量不含硒的土壤化学数据。因此, 为了降低找富硒土壤的成本, 急需基于机器学习利用 1:5 万土地质量地球化学调查数据, 构建富硒土壤预测模型, 之后可将其他工作获取的不含硒元素的数据输入模型, 开展富硒土壤的预测。

由于富硒土壤与 pH、元素含量等指标数据, 不符合正态分布, 且存在较强的非线性联系, 适合采用应用范围广, 扩展性强, 对自变量要求低的方法开展建模。因此, 本次通过无需编程即可利用统计和数据挖掘算法构建预测模型的 SPSS Modeler 18 软件开展预测, 该方法被用于医学诊断, 技术较为成熟, 结果较为准确。

以 1277 组江西省信丰县油山地区 1:5 万土地质量地球化学调查数据为例, 根据地质背景及数据相关性及其共线性分析, 优选了 8 种参数 (Zn、pH、D、K<sub>2</sub>O、P、Mo、Mn、Cr) 作为预测变量, 以是否富硒为因变量, 运用 SPSS Modeler 18 软件开展预测富硒土壤的二元 Logistic 回归模型、多层感知器神经网络模型和支持向量机模型的构建。在对这些模型进行准确性和稳定性评价的基础上, 利用 28 组土壤样品的实测数据, 对模型开展验证, 遴选最优模型。

研究结果表明: (1) 二元 Logistic 回归模型、多层感知器神经网络模型和线性支持向量机模型对富硒土壤预测效果较好, 正确率达 100%, 其中多层感知器神经网络模型模型的正确率和稳定性最好。(2) 二元 Logistic 回归模型、多层感知器神经网络模型、随机森林模型和支持向量机模型 (线性、多项式、径向基、Sigmoid) 建模时的预测正确率可以作为优选模型的依据, 但与实际数据验证的结果并非完全一致。(3) 通过斯皮尔曼相关性分析发现元素 Mo、Cr 与土壤富硒具显著正相关性, 且相关系数为 0.735 和 0.518; 不同的预测模型均表明 Mo、Cr 对土壤富硒影响较大, 故可将 Mo、Cr 作为富硒土壤的指示元素。

本研究基于机器学习构建富硒土壤预测模型, 为快速、高效、精准的圈定富硒土壤提供依据, 且可进一步拓展到地质找矿及环境监测等领域。

本研究的快速、高效、精准的圈定富硒土壤提供依据, 且可进一步拓展到地质找矿及环境监测等领域。

### 参考文献

- [1] Jiao L S, Zhang L, Zhang Y, Wang R, Lu B, Liu X. 2022. Transcriptome analysis provides new insight into the distribution and transport of selenium and its associated metals in selenium-rich rice[J]. Environmental Pollution, 301: 118980.
- [2] 王运, 邹勇军, 王鹤, 汪成钵. 江西信丰油山地区土壤硒及重金属元素地球化学特征[J]. 华东地质, 2019, 40(2): 152-160.
- [3] 周永章, 左仁广, 刘刚, 袁峰, 毛先成, 郭艳军, 肖凡, 廖杰, 刘艳鹏. 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[J]. 矿物岩石地球化学通报, 2021, 40(3): 556-573.
- [4] 成秋明. 什么是数学地球科学及其前沿领域[J]. 地学前缘, 2021, 28(3): 6-25.

基金项目: 东华理工大学江西省数字国土重点实验室开放研究基金资助项目 (DLLJ202205); 江西省研究生创新专项资金项目 (YC2022-S600)。

第一作者简介: 杨兰 (1997-), 硕士研究生, 研究方向: 主要从事土壤地球化学研究. E-mail: 1483054242@qq.com

通信作者简介: 王运 (1985-), 博士, 高级工程师, 研究方向: 主要从事农业地质研究. E-mail: 201960125@ccut.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 岩石薄片图像智能分析研究进展

苏程<sup>1\*</sup>

1. 浙江大学地球科学学院, 杭州 310058

分析岩石薄片显微图像中的矿物组成、结构构造、生成顺序、围岩蚀变和次生变化等特征, 可以进行岩石种类鉴定、地质构造分析、古地质环境反演等相关研究。当前岩石薄片显微镜下的分析与鉴定工作主要依赖人工手动完成。由于受制于专家知识、经验和效率, 针对大规模样品的研究工作, 不同人员的研究成果往往缺乏统一标准而难以进行横向对比及综合使用, 同时还会由于时间或成本等因素而难以完成。随着图形图像采集设备的发展, 镜下岩石薄片可以被数字化, 形成数字岩石薄片图像, 便于对其进行记录、观察和分析。因此, 如何利用计算机对岩石薄片图像进行自动化定量分析, 快速产出大量、系统、客观、准确的岩石薄片图像分析结果, 在“大数据+地球系统科学”的时代背景与研究范式转变中, 将具有重要的研究意义与应用价值。

本文梳理总结了利用卷积神经网络深度学习方法对岩石薄片图像进行智能分析的研究工作, 从技术实现的角度归结为图像分类、语义分割和目标检测 3

等种模式, 并结合具体研究案例阐释不同模式下岩石薄片图像智能分析的具体模型的选择与构建, 训练样本的制作方式, 以及样本标签的形式等技术实现环节。结合实际应用案例和技术方法特点, 提出了当前岩石薄片图像智能分析面临: 岩石薄片图像数据规模相对较小、图像获取稳定性相对较低, 智能分析模型构建难度较大, 智能分析模型推广性相对较差, 应用场景复杂等困难与挑战, 并给出了对应的解决思路。

本文所总结的关于岩石薄片图像智能分析的一般原理与方法不但可以应用在光学图像上, 也同样适用于如诸如微区 X 射线荧光光谱(micro-XRF)、扫描电镜-能量色散 X 射线光谱仪(SEM-EDX)、电子探针等分析技术基于矿物成分产生的岩石样本图像, 从而快速产生丰富、客观、准确的定量岩石特征信息。随着有关数据的不断积累和机器学习技术的发展, 岩石薄片图像智能分析将会为地球科学研究与应用提供更加丰富多样的定量化岩相信息。

基金项目: 国家自然科学基金资助项目(42050103)

第一作者/通讯作者简介: 苏程 (1985-), 副教授, 研究方向: 地学大数据与机器学习理论与应用. E-mail: sc20184@zju.edu.cn

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于知识图谱和机器阅读的岩石描述文本质量检测

陈忠良<sup>1, 2\*</sup>, 袁峰<sup>1</sup>, 李晓晖<sup>1</sup>

1. 合肥工业大学资源与环境工程学院, 合肥 230009;

2. 安徽省地质调查院, 合肥 230001

岩相学关注岩石颜色、结构、构造和矿物成分等特征的描述和岩石分类。岩相学研究形成的岩石描述文本是野外地质观测获取的第一手调查资料, 其质量对地质调查、矿产勘查和地表基质调查等具有重要影响。在数字地质调查过程中, 岩石描述文本是除照片和影像外, 重要的非结构化数据之一。当前不同领域人工智能正由“感知智能”过渡到“认知智能”, 以知识图谱和机器阅读技术为代表的认知智能技术已在多个行业取得进展, 也为非结构化的岩石描述文本质量检测提供了技术基础。同时, 以大规模预训练语言模型为基础的智能聊天机器人 ChatGPT 也在全球范围内掀起一阵“人工智能”热潮。本文充分利用大规模预训练语言模型, 针对非结构化的岩石描述文本, 开展基于知识图谱和机器阅读的岩石描述文本质量检测方法研究。首先, 基

于岩相学理论开展岩石描述文本命名实体与关系定义; 然后, 基于大规模预训练语言模型, 在人工标注岩石描述预料的基础上, 开展基于深度学习方法的岩石描述文本知识提取方法研究; 同时, 基于先验知识开展用例驱动的迭代式岩石知识图谱构建; 最后, 基于岩石知识图谱计算提取的岩石知识三元组与岩石知识图谱的一致性。经多专家协作确认的不一致岩石知识三元组进入先验知识库, 可实现岩石知识图谱的迭代增新。文中选择安徽 1:5 万宣城县幅开展了实证研究。实验结果显示, 本文提出的技术框架能够实现岩石描述文本的实体与关系提取。基于岩石知识图谱的岩石知识一致性计算能够快速检测出岩石描述文本中结构、构造和物质成分的描述错误。本文所提方法能够实现岩石描述文本的智能检测, 可有效提高岩石描述文本质量。

• 专题 29: 地学大数据挖掘、机器学习与人工智能算法应用 •

## 基于多信息融合技术的未测区土壤元素含量预测

董健彪<sup>1</sup>, 万翔<sup>2</sup>, 蒋兴念<sup>3</sup>, 高婕妤<sup>2</sup>, 汪新庆<sup>3</sup>, 谢淑云<sup>1\*</sup>

1. 中国地质大学(武汉) 地球科学学院, 武汉 430074;

2. 湖北省地质调查院, 武汉 430034;

3. 中国地质大学(武汉) 资源学院, 武汉 430074

近年来,机器学习等多信息融合技术成为地质学领域的一大热点,为地学的研究提供了诸多助力,利用多信息融合技术可以更好地挖掘地质数据中的隐藏信息。土壤是土地的重要组成部分,为作物生长提供营养元素以及生长环境,是整个生态系统不可或缺的一部分,掌握土壤中元素含量的分布情况对生产生活具有重要意义。多信息融合技术预测土壤理化性质已被广泛使用。目前,我国多生去地球化学调查工作广泛开展,但依然存在大面积未测区。随着生态环境地球化学调查、土壤质量地球化学调查、区域地球化学找矿等多项目的联合开展,缺少地球化学测试分析数据的未测区尚有大量工作可作。多信息融合可为未测区区域地球化学分布特征等提供技术支撑。本文拟

采用模糊证据权法和生成对抗网络,利用湖北省地质图、土壤类型图、水系缓冲区以及已测区土壤地球化学元素作为训练数据,预测湖北省未测区土壤中元素分布特征,并利用典型示范区地球化学数据验证预测结果。结果表明:两种方法都能有效预测示范区元素分布特征,其中,模糊证据权对 As、Cu 两元素的预测准确率达到 60%以上。这些工作采用地质、地球化学、计算机、数学等多学科交叉和地质观测与数值模拟相互印证的技术路线,系统识别地表单指标、多指标地球化学元素的分布特征及其与不同类型土壤分布的关系,通过已知的地球化学调查数据与地层属性的对应关系,填补采样空白区数据,为地表环境中典型环境指标的定量预测提供技术方法支撑。

作者简介:董健彪(1998-),男,在读硕士,主要从事农业地球化学研究工作。E-mail: 15849571958@163.com。

\*通讯作者简介:谢淑云(1976-),女,教授,博士,主要从事应用地球化学与数学地球科学相关研究工作。E-mail: tinaxie@cug.edu.cn。