

引用格式:蔡惠慧,朱伟,李孜轩,等.基于深度学习的钨钼找矿靶区预测方法研究[J].地球信息科学学报,2019,21(6):928-936. [Cai H H, Zhu W, Li Z X, et al. Prediction method of Tungsten-molybdenum prospecting target area based on deep learning[J]. Journal of Geo-information Science, 2019,21(6):928-936. ] DOI:10.12082/dqxxkx.2019.190032

# 基于深度学习的钨钼找矿靶区预测方法研究

蔡惠慧<sup>1,2</sup>, 朱伟<sup>3\*</sup>, 李孜轩<sup>4,5</sup>, 刘园园<sup>2</sup>, 李龙斌<sup>3</sup>, 刘畅<sup>2</sup>

1. 中国地质大学(北京), 北京 100083; 2. 中国地质调查局发展研究中心, 北京 100037; 3. 陕西省矿产地质调查中心, 西安 710068; 4. 中国地质大学(武汉)信息工程学院, 武汉 430074; 5. 中国地质大学(武汉)国家地理信息系统工程技术研究中心, 武汉 430074

## Prediction Method of Tungsten-molybdenum Prospecting Target Area based on Deep Learning

CAI Huihui<sup>1,2</sup>, ZHU Wei<sup>3\*</sup>, LI Zixuan<sup>4,5</sup>, LIU Yuanyuan<sup>2</sup>, LI Longbin<sup>3</sup>, LIU Chang<sup>2</sup>

1. China University of Geosciences (Beijing), Beijing 100083, China; 2. Development and Research Center of China Geological Survey, Beijing 100037, China; 3. Shanxi Center of Mineral Geological Survey, Xi'an 710068, China; 4. Department of Information Engineering, China University of Geosciences, Wuhan 430074, China; 5. National Engineering Research Center of Geographic Information System, Wuhan 430074, China

**Abstract:** With the exploration of minerals from shallow mines to deep concealed mines, from easy-to-identify mines to difficult-to-identify mines, the difficulty of prospecting is increasing, and geological experts are paying more and more attention to the application of new theories, new methods, and new technologies. As a frontier field and technology of artificial intelligence, deep learning has a unique advantage in realizing the intelligent forecasting and evaluation of mineral resources. The method uses normalized geochemical data as the training data to extract outliers by a neural network called Autoencoder and identify the favorable mineralization areas, and then realizes the qualitative prediction of mineral resources prospecting prospect. The research results show that after classifying the original data of 957 single elements geochemical anomalies and labeling of the model, the whole process automatically completes the learning and prediction in the "black box" of the computer, compared with the traditional prediction research method, this method of research is highly automated and objective. In addition, this paper uses the known mine sites to construct the training dataset, and uses the random forest method to predict the mineral resources prospecting target area in the prediction area, which provides a scientific basis for further narrowing the scope of the prospecting target area.

**Key words:** random forest method; deep learning; tung-molybdenum polymetallic mineral resources; big data; prediction; evaluation; Western Zhenan; Shaanxi

**\*Corresponding author:** ZHU Wei, E-mail: 165324824@qq.com

**摘要:**随着矿产勘查工作由浅部矿向深部隐伏矿、由易识别矿向难识别矿发展,找矿难度日益增大,地质专家越来越重视新理

收稿日期:2019-01-18;修回日期:2019-04-22.

基金项目:智能地质调查系统开发与推广(DD20160355)。[ **Foundation item:** Development and promotion of intelligent geological survey system, No.DD20160355. ]

作者简介:蔡惠慧(1986-),女,江苏盐城人,博士生,主要从事地球探测与信息技术研究。E-mail: caihuihui6@126.com

\*通讯作者:朱伟(1985-),男,陕西旬阳人,高级工程师,主要从事地质调查与矿产勘查,成矿预测等工作。

E-mail: 165324824@qq.com

论、新方法、新技术的应用。深度学习作为人工智能的前沿领域/技术,对于实现矿产资源预测“智能化预测评价”具有得天独厚的优势。本文以陕西省镇安县西部钨钼矿集区单元化探异常原始数据为基础,提出了基于深度学习的钨钼矿产评价方法。该方法以归一化地球化学数据作为模型训练数据,通过深度学习中深度自编码网络方法实现异常值提取进而识别重点成矿有利地段,实现矿产资源找矿远景区定性预测。研究结果表明,在对957条单元化探异常原始数据分类且做好模型标签后,整个过程在计算机的“黑盒子”中自动完成学习和预测,相较于传统预测研究方法,本文方法具有自动化程度高和客观性强的特征。此外,本文利用已知矿点构建训练数据集,采用随机森林方法对预测区进行矿产资源找矿靶区预测圈定,为进一步缩小找矿靶区范围提供科学依据。

**关键词:** 随机森林方法;深度学习;钨钼多金属矿产资源;大数据;预测;评价;陕西镇安西部

## 1 引言

随着信息技术以及人工智能的快速发展,矿产资源定量预测评价工作也进入了一个新的阶段,即借助计算机强大又智能的数据处理分析优势,帮助厘清积累已久的地质“大数据”以及地质专家形成的地质模型,为矿产资源定量预测提供很好的辅助决策依据。

地质调查工作多来源、多模态的特点使地质数据呈现数据规模大、数据关联关系复杂等特点。目前,地质调查数据处理技术主要集中在前处理和检索查询方面,而对数据的深度分析与挖掘是地质大数据分析挖掘面临的重要技术问题<sup>[1]</sup>。在大数据时代,对信息量极大的数据来说,数据分析过程是数据处理的关键性环节。大数据分析处理主要分为2大类:①统计分析,主要是使用传统关系型数据库的联机分析处理技术和方法,通过使用SQL语句完成各种查询、统计分析等;②智能化分析,大数据的深度价值仅通过简单分析是难以发现的,因此需要使用基于机器学习和数据挖掘的智能化复杂分析才能实现<sup>[2]</sup>。

机器学习是一种使获取知识自动化的计算方法的学习<sup>[3]</sup>,其应用已遍及人工智能的各个分支,如专家系统、自动推理、自然语言理解、模式识别、计算机视觉、智能机器人等领域<sup>[4]</sup>。阴江宁等<sup>[5]</sup>通过构建系统知识库和规则库,并结合专家对一系列先验概率等参数赋值等方式,用统计学原理实现单一成因矿床及复合矿床类型的推理。于萍萍等<sup>[6]</sup>以地质理论指导地质大数据分析和计算机技术实现地质大数据挖掘两条主线展开研究,结合青海祁漫塔格铁铜多金属矿床、山东焦家金矿床、云南个旧锡铜多金属矿床等不同地区、不同成矿类型和矿种开展了应用研究,实现了面向地质大数据的数据挖掘与矿产资源的定量预测评价。朱月琴等<sup>[7]</sup>从地质大数据语义模型定义和基于知识图谱技术的地质大数

据语义模型构建入手,提出了一套面向地质大数据的语义检索和聚类分析方法,完成了面向地质大数据语义检索系统平台的开发、实施及实验验证。大数据环境下,无论是数据的数量还是种类都巨幅增加,新型的数据类型需要被分析,如文本情感上的分析、图像的搜索和理解、图像数据的分析等,这使机器学习的研究方向变得更加多元化,很多衍生出来的学习方法也得到了较多的应用<sup>[8]</sup>。例如,如何合理地使用半监督学习的方法<sup>[9-10]</sup>来提高学习泛化能力的集成学习方法<sup>[11]</sup>。王惠英等<sup>[12]</sup>通过构建深度学习卷积神经网络水体提取模型,进而自动提取河道和自动发现变化图斑。徐永洋等<sup>[13]</sup>通过随机森林方法实现开放街区数据(OpenStreetMap, OSM)城市主干线自动提取。周龙等<sup>[14]</sup>通过深度学习检测方法检测雷达图像目标,并在原YOLOv检测精度上提高了6.07%。胡滨等<sup>[15]</sup>通过极限学习机算法分析铝羟基和镁羟基矿物空间分布信息,用匹配滤波算法处理Hyperion数据获取的3种热液蚀变矿物,进而建立高光谱数据识别热液蚀变矿物系统方法达到预测目的。王圆圆等<sup>[16]</sup>利用高光谱数据,研究了决策树算法特征选择的特点以及特征选择对决策树分类结果的影响。上述研究均体现出了机器学习以及深度学习在数据的特征提取及预测识别方面有着广阔的应用前景,但并没有以深度学习作为成矿预测及评价工作的主体。传统的找矿预测评价工作主要使用证据权法,需要满足各找矿预测要素条件独立性,成矿过程本就是复杂且漫长的过程,为了探索各找矿预测要素数据本身的非线性关系以及对致矿的影响,本文以深度学习作为主要方法,利用2种典型的神经网络对南秦岭山脉中段的钨钼矿产进行了找矿预测及评价工作,以神经网络训练的结果作为矿产的异常指示。结果表明,将深度学习神经网络应用在成矿预测工作中能够取得良好的效果。

## 2 研究区概况与数据源

### 2.1 研究区概况

研究区位于陕西省镇安县西部,属秦岭山脉中段南坡,地势西高东低,海拔530~3000 m,水系发育,植被覆盖较为严重。该地区地质构造复杂,岩浆活动强烈,成矿地质条件优越,资源潜力巨大,具有形成大型-超大型钨钼、金、铜矿床的前景,现已成为陕西省找矿热点和重点地区之一。

研究区北部为北秦岭复合岛弧杂岩带,中部为商丹复合板块构造结合带,南部为南秦岭边缘海盆(图1)。区内沉积地层发育较为齐全,出露古元古界结晶基底杂岩、中-新元古界过渡性褶皱基底岩系和震旦系至上石炭统沉积盖层。断裂较为发育,主要有印支主造山期的近东西向断裂带和形成于燕山期的北北东向断裂带,两组不同方向断裂形成了区内五纵四横的断裂格局,并对区内各种矿产的形成和赋存起着决定性作用。侵入岩(基)体规模大,为区域成矿提供充足的热动力条件和物质来源。岩体主要形成于印支期-燕山期,主要有东江口、胭脂坝、懒板凳、四海坪等中酸性复式岩体。中酸性侵入体与碳酸盐岩地层的接触部位形成数十米至几十米的矽卡岩(化)带,与钨钼、铅锌成矿较为密切。现已发现的金属矿产种类有钨、钼、金、银、铅锌、铁、钒等12种,以钨、钼、铜、铅锌、金为优势矿种。矿产地126处,其中大型矿床2处(钨2处)、中型矿床8处(金1处、铅锌2处、钨4处、钒1处)。矿床类型主要为与岩浆作用密切相关的矽卡岩型、岩浆热液型、层控热液型、构造热液型,次为构造蚀变岩型、海相沉积型、沉积-变质型。在空间

上,矿床(点)温度梯度分带明显,从岩体向外到围岩,成矿具有从高温到低温的钨钼、铜、铅锌、金的矿床分带性,具有良好的成矿空间。

### 2.2 数据源

已有地质研究资料表明,该矿集区主要表现为钨钼多金属矿化活跃,具有良好的钨钼成矿潜力,因此本文实验研究目的是定性的圈定钨钼找矿靶区,为地质专家和研究人員提供辅助决策依据。该矿集区研究范围约为9个五万标准图幅面积,由于南秦岭的植被覆盖严重,影像图的智能解译证据不足,故采用1:25万水系沉积物测量化探原始数据作为本实验的试验数据。39种单元素(或氧化物)数据一共包含957个原始采样点,通过查阅该研究区大量的已有研究资料,本文选择了9种与钨钼多金属矿致矿相关的金属元素作为模型训练数据,即Mn、Bi、Ag、As、Co、Cr、Ni、Ti、Zr共9种元素;通过深度学习中深度自编码网络方法实现异常值提取,并用已知矿区文件作为验证数据进而达到预测目的。

## 3 研究方法

地质数据具有数据量庞大、类型多样、关联关系复杂等特点,使地质研究和矿产预测工作不仅需要地质专业人员的野外现场勘探和长时间的专业知识积累,还需要借助计算机等手段快速高效的分析计算。为了提高矿产预测结果的可靠性,研究者们常使用相同的观测数据应用多种统计预测模型来预测矿产资源靶区<sup>[17]</sup>。本文通过对研究区中的

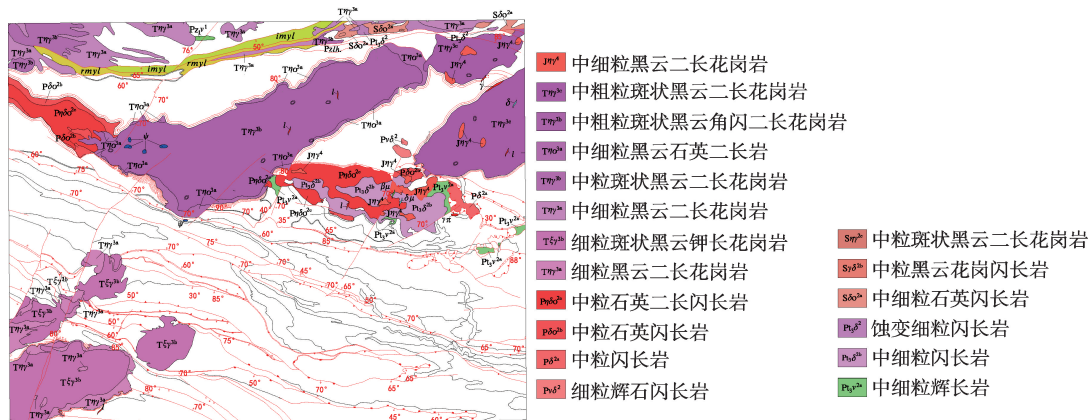


图1 研究区岩体与构造

Fig. 1 Rock mass and structural lines in the study area



地球化探元素信息进行分析,预测可能存在钨钼矿点的找矿靶区。根据研究区钨钼成矿特征和主要控矿因素,结合大量文献资料,最终选取了与钨钼成矿相关的化探元素信息,分别为Mn、Bi、Ag、As、Co、Cr、Ni、Ti、Zr共9种。在已知钨钼矿区中,这9种元素均显示出不同程度的异常。据此信息,本研究将上述9种元素连同钨钼一起,作为钨钼成矿的指示性元素,用来预测钨钼矿点。

### 3.1 矿产预测学习模型确定

通过水系沉积物测量的化探元素强度进行定性的矿产预测,针对不同地形能够采用不同的方法。由于遥感影像光谱可以充分反映地表物质特征,在地表裸露的研究区可以充分利用研究区内遥感影像数据,通过先验知识进行训练样本人工标注,然后利用监督的方法(如随机森林(RF)、极限学习机(ELM)等)进行像素级分类研究,从而得到研究区内成矿预测。这种监督学习的方法成本低、自动化程度高。

在地表被植被覆盖的研究区,由于研究区内被大量植被覆盖,光谱特征不能直接反映地表矿物特性信息,因而需要定点测量一些地球化学数据。通过数据预处理,主成份分析(PCA)等方法选择合适的成矿元素。利用深度学习中非监督学习方法(深度自编码网络(AE)、深度对抗神经网络(GAN)等方法)找到研究区内化学元素异常分布,再异常值提取进而实现成矿预测目的。非监督学习方法除了自动化程度高的优势以外,还具有客观性强的特点。具体定性验证流程如图2所示,通过化探原始数据和深度学习得出的预测数据进行对比验证,圈定找矿重点预测区,达到定性矿产资源预测目的,为矿产勘查工作部署提供科学依据。

传统的统计方法、分形/多重分形方法在处理单变量的地球化学异常时,都能取得比较好的结果。已有的多变量统计方法(如PCA)和分形相结合的方法也已广泛地应用于多元地球化学异常识别之中。由于成矿的复杂性,导致地球化学数据往往并不满足线性分布,所以本研究使用深度自编码网络这一非线性算法来处理。深度自编码网络作为一个新的异常识别方法,可以很好地用于多元地球化学异常的识别。对小样本数据,深度自编码网络的编码重构的能力比较弱,而对大样本数据编码重构能力比较强,可利用重构误差作为地球化学异常识别的指示变量。重构误差的值比较高的区域,跟已

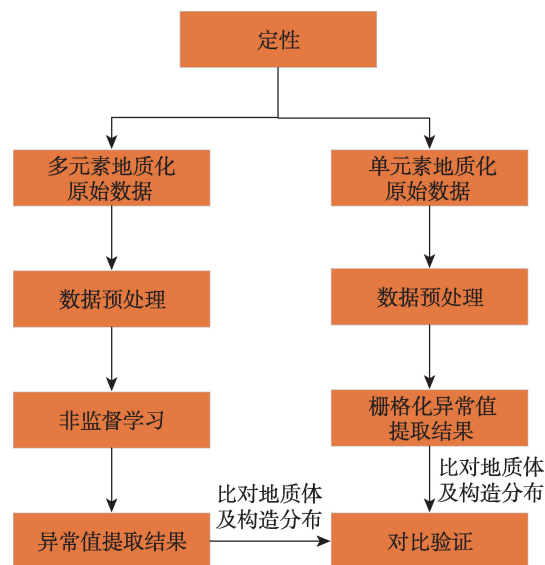


图2 定性矿产资源预测流程

Fig. 2 Flowchart of qualitative prediction of mineral resources

知矿点具有比较高的空间相关性,说明提取出的异常区域可能跟矿化区域相关,该异常区域可以作为进一步的重点勘查区域。

### 3.2 自编码网络模型构建

Hinton提出的自编码神经网络,该网络被认为是一种新的神经网络模型(图3),通过多层神经网络对输入数据进行编码和解码<sup>[17]</sup>。深度自编码神经网络具有对称结构,通过训练使输入和输出之间的差异最小化。输出数据是对输入的估计,它们之间的差异定义为重构误差,常用于检测异常<sup>[18-19]</sup>与

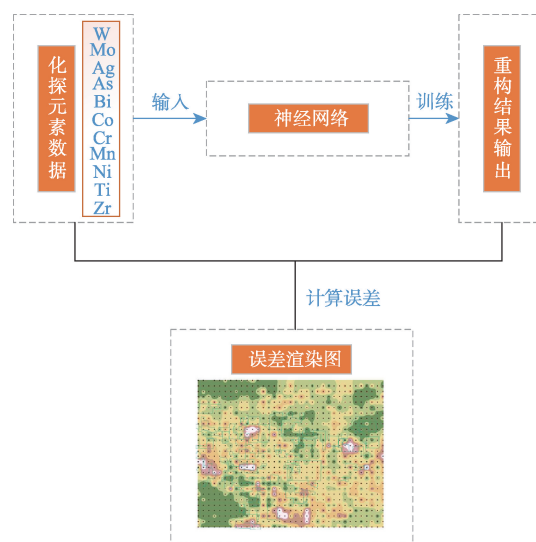


图3 自编码网络深度学习预测模型

Fig. 3 Deep learning prediction model diagram of autoencoding networks



数据特征分析<sup>[20]</sup>。较大的重构误差对应着明显异常的数据输入,重构误差计算如下:

$$E = \sqrt{\sum_{i=1}^n (O_i - I_i)^2} \quad (1)$$

式中: $E$ 表示重构误差; $O$ 表示输出数据; $I$ 表示输入数据; $n$ 表示输出层的大小。

由于受限玻尔兹曼机(RBM)输入只能是0或者1的离散型数据,然而实际当中地化数据都是连续型,因此在构建深度自编码网络时,本实验引入了连续性受限玻尔兹曼机(CRBM)。连续性受限玻尔兹曼机是在受限玻尔兹曼机基础之上,加入标准化的高斯噪声,即在待操作的数据上加入一项标准正态分布,从而使RBM能更好地处理连续数据,得到构建本次深度自编码网络的CRBM模型。

为使重构误差最小,对深度自编码神经网络的训练可以分为3步:①在预训练步骤中,CRBMS<sup>[21]</sup>被逐个训练;②在展开步骤中将所有经过训练的CRBMS堆叠起来,形成一个深层的自编码神经网络;③引入了一种反向传播算法来调整网络中的所有参数,这个步骤称为微调(图4)。

在预训练过程中,自编码神经网络的激活功能定义为:

$$\varphi(x_j) = \theta_L + (\theta_H - \theta_L) \frac{1}{1 + e^{-\alpha_j x_j}} \quad (2)$$

式中: $\varphi(x)$ 是一个sigmoid函数,其渐近线是常数 $\theta_L$ 和 $\theta_H$ 。参数 $\alpha_j$ 控制了sigmoid函数的斜率,确保了

当 $\alpha_j$ 增加时从无噪声的确定状态到二元随机状态的平稳过渡<sup>[22]</sup>。

### 3.3 实验设置

#### 3.3.1 深度自编码网络方法

首先将研究区内1:25万化探采样点做预处理,然后针对每一个点上对应上文所述的9种地球化学元素做为深度模型的数据输入。利用深度自编码网络对该网格位置的变量数据进行重编码,针对每个网格计算其重构误差,对于异常的区域,其重构误差会比较大。图5为采样点坐标数据在ArcGIS中的显示,其中每个点中已包含所需各种元素的数据。

研究区原始采样点共有957个,每个采样点包含11项化探元素的数据,因此输入神经网络的数据集为 $957 \times 11$ 的矩阵,即该深度学习网络的样本尺寸是 $957 \times 11$ 。网络学习率(Learning Rate)设置为0.1。训练epoch设置为1000,batch size设置为50。本文将其中的80%作为训练集,剩下20%的数据作为测试集。本次深度自编码网络由4层CRBM构成,各层CRBM尺寸见表1。每层CRBM的输出作为下层CRBM的输入,不同连接层CRBM的权重 $W$ 与偏置 $b$ 单独训练。

实验中深度自编码神经网络迭代训练了1000次(图6)。由图6分析可得,前100次迭代过程中,重构误差收敛速度较快;200次迭代之后,重构误差

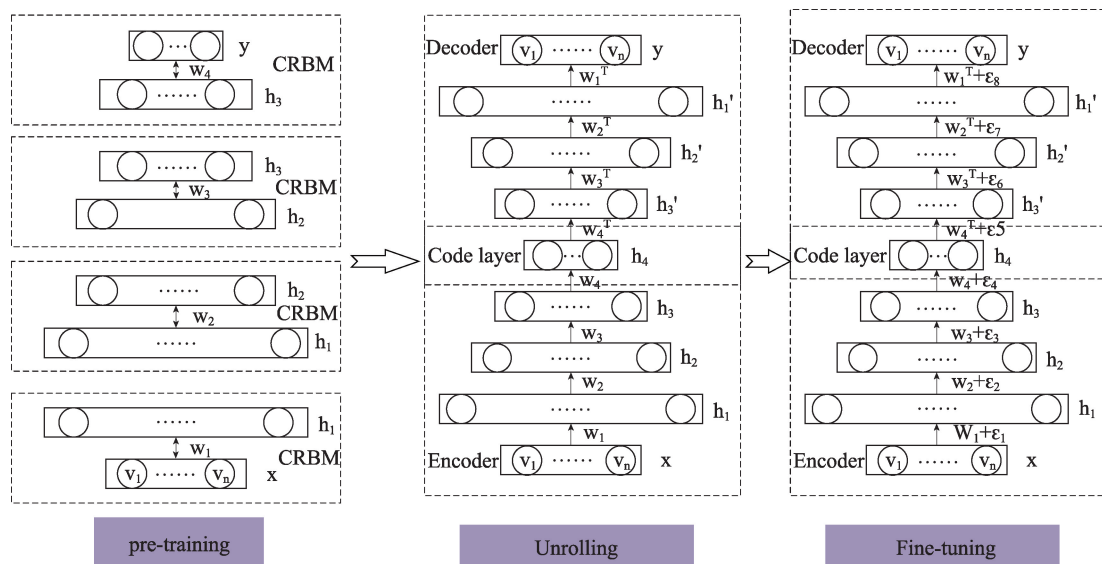


图4 自编码神经网络流程

Fig. 4 Deep learning process diagram of autoencoding networks

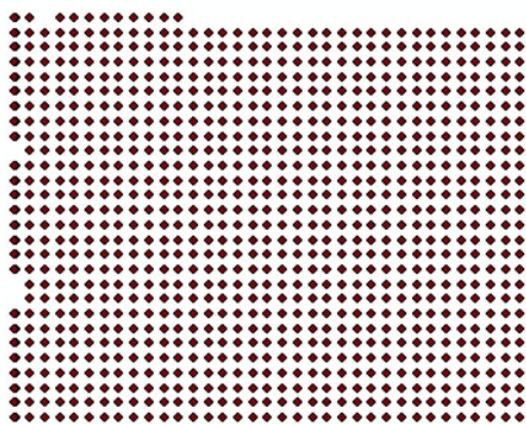


图5 研究区化探采样点坐标范围

Fig. 5 Coordinate ranges of the sampled chemical points in the study area

表1 深度自编码网络每一层CRBM尺寸

Tab. 1 CRBM size of each layer of the deep autoencoder networks

网络层编号	CRBM 尺寸
Layer1	11×100
Layer2	100×50
Layer3	50×20
Layer4	20×5

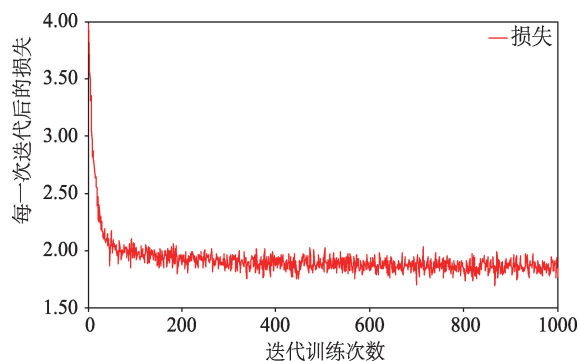


图6 深度自编码神经网络迭代损失

Fig. 6 Iterative losses of the autoencoder neural networks

基本趋于稳定(或振幅较小)。针对训练数据在1000数量级之内,训练迭代次数500次就可以达到较好的学习效果。

### 3.3.2 随机森林方法

随机森林是一种集合学习方法,可用于分类和回归。它由一系列的决策树组成,每一类输出是这一类的属性(用于分类)或森林中所有树的平均预测(用于回归)。在随机森林的训练过程中,利用装袋随机生成训练数据集。在树的训练过程中,引入了基尼指

数作为随机森林中的属性选择度量,并对表示类的特征的异质性进行了度量。基尼指数可以写为:

$$Gini(T) = \sum \sum \left( \frac{f(C_j, T)}{|T|} \right) \times \left( \frac{F(C_j, T)}{|T|} \right) \quad (3)$$

式中:  $T$  表示训练数据集;  $f(C_j, T)/|T|$  所选示例隶属于  $C_j$  类的概率。

随机森林需要2个参数:属于一棵树的每个节点的特征数量,森林中的决策树的数量。对于一个新的数据集,通过遍历随机林中的每一棵树来预测数据集中的每一种预测情况,然后选择在所有树中投票最多的类。对于分类模型,计算结果如下:

$$H(x) = \arg \max_c \sum_{i=1}^T h_i^c(x) \quad (4)$$

式中:  $T$  表示森林中决策树的数量;  $h_i^c(x)$  是样本  $x$  通过决策树  $i$  分类到类别  $j$  的输出。

## 4 实验结果与分析

### 4.1 深度自编码神经网络

深度自编码神经网络是一种无监督的学习算法,它通过前向操作提取输入的特征,后通过“解码”的过程反向还原出原始数据。本文以研究区内各点所包含的9种与钨钼多金属矿致矿相关的金属元素作为输入数据,输入到深度自编码神经网络中进行训练,比较还原出的数据与输入的原始数据之间的差异,从而得到研究区内各点的成矿异常情况(图7)。

将原始数据与训练得到的元素  $W$  与  $Mo$  依照

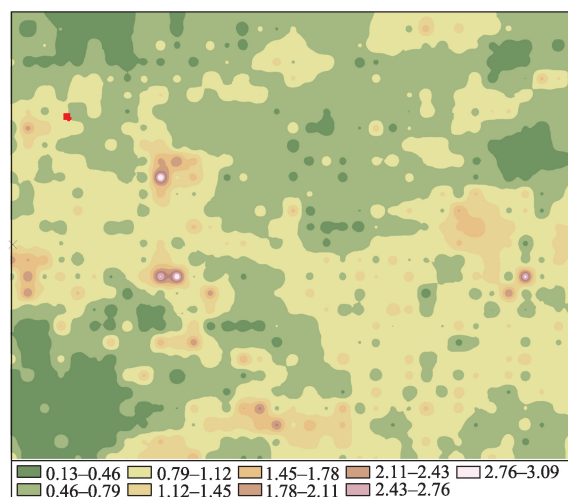


图7 化探采样点自编码深度误差

Fig. 7 Self-coding error map of the geochemical sampling points

图2的步骤进行操作,最终能够得到一组与原始数据规模相同的输出数据,这些输出数据与每个点的坐标也能够一一吻合。将这组输出数据与输入数据做差,并将每个数据取平方后线性拉伸至0~255的范围,能得到一组按采样点分布的标准化数据。在ArcGIS中对这些数据进行反距离加权(IDW)操作,能够得到各个采样点的误差大小图像。

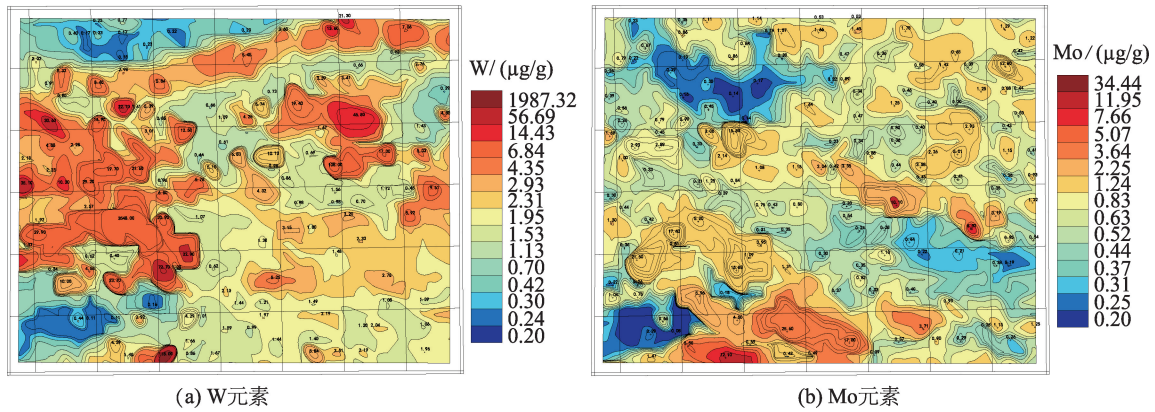


图8 W和Mo的单元素异常等值线  
Fig. 8 Anomaly contour map of W and Mo

比较化探采样点误差示意图与单元素异常等值线图,可见前者高值点的分布与Mo元素化探异常分布特征高度一致,较好的反应了Mo异常分布情况( $>2.25 \mu\text{g/g}$ ),因此,本文选择的非监督学习方法在处理野外实际化探数据方面效率高、效果好,可以较好的帮助地质工作人员快速高效地缩小找矿范围,为矿产勘查工作节约人力物力和时间成本。

#### 4.2 随机森林方法

随机森林是一种有监督的学习算法。它通过构建一定数量的决策树,并将生成的多棵分类树组成随机森林,用随机森林分类器对数据进行分类,按多棵树分类器投票决定最终分类结果。训练完成后,能得到预测区域内各点(图9)的训练误差值,误差越大则说明该点的异常情况越明显,成矿可能性越大。同时,通过F1得分(它是模型准确率和召回率的一种加权平均,它的最大值是1,最小值是0,是衡量二分类模型精确度的一种指标)选择合适的决策树数目,使模型更加高效、完善。

由图10中的误差示意图可看到,通过对已有矿区的化探原始采样点标本学习,预测区内的高值异常区与自编码深度学习结果有多处预测点吻合,结合构造及地质体分布图分析得出,预测区域内的矿产总体与构造和地质体的空间分布规律一致,进而

由图8可知,研究区元素异常总体呈近东西向和北北东向带状或串珠状展布,与区域构造保持一致,元素w、Mo异常高值区与该区三叠纪岩体(胭脂坝岩体、四海坪岩体、东江口岩体等)分布范围套合较好,在岩体与碳酸盐岩地层接触部位、断裂构造发育区异常强度更高,即研究区钨钼成矿受印支-燕山期构造-岩浆活动联合控制。

也间接证明了该方法在矿产资源预测领域的首次探索研究成功,在前期的矿产勘查阶段,能够为地质研究者提供辅助决策依据。

在随机森林训练过程中,随机森林中决策树数目对训练结果起至关重要的作用。当决策树较少时,随机森林不能有效的学习到数据特征,从而影响到预测结果的精度;当决策树较多时,会增加训练的复杂度,使得学习效率降低。在本实验中,通过观察不同决策树数量对F1结果的影响,选择最优的值,构建稳定性较高并且可以有效避免过度拟合的随机森林(图11)。由图分析可得,当随机森林结



图9 预测区域化探点位数据分布  
Fig. 9 Geodetic points distribution map in the predicting area



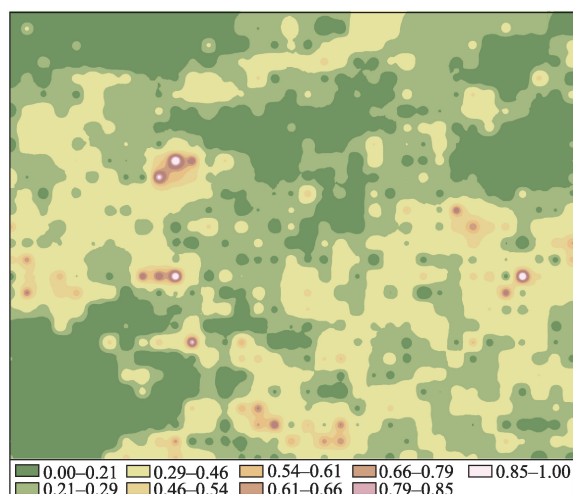


图10 预测区域化探采样点误差

Fig. 10 Error map of the geochemical points in the predicting area

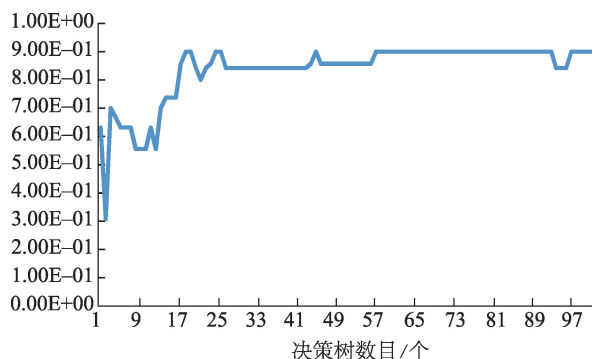


图11 随机森林决策树数目与其F1得分间的关系

Fig. 11 Relationship between the number of random forest decision trees and their F1 scores

果达到一定稳定水平,即使决策树的数量发生改变,F1得分几乎不发生变化。因此,本实验中决策树数量设置为58,F1得分达到最高值趋于稳定。

需要特别说明的是,由于本研究区的已知矿点数据为30个,相对于957个原始采样点数据,标本数据量太少,因此在实验中按照样本数据与预测数据为1:4比例设置,随机挑选了120个预测数据加上30个样本数据训练,发现在决策树数目为58左右的时候,F得分趋于稳定。实验中,随机森林方法是一种探索,当样本数据量增大,训练数据也足够多的情况下该系统是否能保持高精度且维持稳定,有待进一步验证。

## 5 结论与讨论

本文在“大数据”背景下,尝试性地通过分析数

据本身非线性关系,采用深度自编码网络以及随机森林的机器学习方法,实现了对南秦岭中段南坡地区可能存在的钨钼矿产资源的预测,并对预测结果进行了评价。本文结论如下:

(1)利用深度自编码神经网络,提取各个勘测点相对于W、Mo的单元元素异常值,能够得到各地化勘测点自编码深度误差,误差值越大表示致矿异常越高,表明成矿概率越大,图7的误差高低值分布显示,误差大于2.43的高值分布区域与W、Mo单元元素异常高值分布区位置吻合,此误差能较好地反映出W、Mo的成矿分布。

(2)利用随机森林,得到的化探样点误差图与自编码深度学习结果有多处预测点吻合。一方面,证明随机森林方法在矿产资源预测领域的首次应用成功;另一方面,表示化探原始采样点数据对找矿指示的直接作用,在前期的实地矿产勘查阶段,能够为地质研究者提供辅助决策依据。

本实验选取的是化探原始采样点957条数据,由于该研究区的遥感影像数据植被覆盖较多,因此本实验选取的化探原始数据做了方法上的探索,下一步的方向会在方法探索的基础上,结合地质,地球物理以及高光谱等多元数据,构建多元地质大数据挖掘模型,对矿产资源预测评价提供支撑。

通过以上深度学习方法的探索尝试,随机森林等监督学习的方法适应于样本基数较大且样本明确的试验,而自编码网络方法在样本数据较少的情况下分析结果较好。本实验的随机森林方法和自编码网络方法结果某种程度表现出一致性,主要归因于该矿区的样本明确、构造和岩体分布规律明确等因素。但是,矿产资源潜力预测工作因其特殊的知识背景和自然规律,需要提取更多的证据帮助指示成矿预测,因此,今天将会着力于加大原始数据集的输入类型,通过研究传统的专家知识背景,结合地质要素建立专家知识库,改进矿产资源预测深度学习模型,以便更好地为矿产资源预测工作提供辅助决策依据。

### 参考文献(References):

- [1] 严光生,薛群威,肖克炎,等.地质调查大数据研究的主要问题[J].地质通报,2015,34(7):1273-1279. [Yan G S, Xue Q W, Xiao K Y, et al. An analysis of major problems in geological survey big data[J]. Geological Bulletin of China, 2015,34(7):1273-1279. ]
- [2] 黄宜华.大数据机器学习系统研究进展[J].大数据,2015,

- 1(1): 35-54. [ Huang Y H. Research progress on big data machine learning system[J]. Big Data Research, 2015,1(1):35-54. ]
- [ 3 ] John Zukowski. Java2从入门到精通[M].北京:电子工业出版社,2000. [ John Zukowski. Java 2: From entry to mastery[M]. Beijing: Electronic Industry Press, 2000. ]
- [ 4 ] 黄林军,张勇,郭冰榕.机器学习技术在数据挖掘中的商业应用[J].中山大学学报论丛,2005(6):145-148. [ Huang L J, Zhang Y, Guo B R. Commercial application of machine learning technology in data mining[J]. Sun Yatsen University Forum, 2005(6):145-148. ]
- [ 5 ] 阴江宁,肖克炎,何凯涛,等.铜矿数字矿床模型专家系统的原理与技术实现[J].地质论评,2009,55(3):449-456. [ Yin J N, Xiao K Y, He K T, et al. Principle and technology realization of copper mine digital deposit model expert system[J]. Geological Review, 2009,55(3):449-456. ]
- [ 6 ] 于萍萍,陈建平,柴福山,等.基于地质大数据理念的模型驱动矿产资源定量预测[J].地质通报,2015,34(7):1333-1343. [ Yu P P, Chen J P, Chai F S, et al. Research on model-driven quantitative prediction and evaluation of mineral resources based on geological big data concept [J]. Geological Bulletin of China, 2015,34(7):1333-1343. ]
- [ 7 ] 朱月琴,谭永杰,张建通,等.基于Hadoop的地质大数据融合与挖掘技术框架[J].测绘学报,2015,44(S1):152-159. [ Zhu Y Q, Tan Y J, Zhang J T, et al. A framework of hadoop based geology big data fusion and mining technologies[J]. Acta Geodaetica et Cartographica Sinica, 2015,44(S1):152-159. ]
- [ 8 ] 王晓.大数据环境下机器学习算法趋势研究[J].哈尔滨师范大学:自然科学学报,2013,29(4):48-50. [ Wang X. Research on trends of machine learning algorithms in big data environments[J]. Natural Sciences Journal of Harbin Normal University, 2013,29(4):48-50. ]
- [ 9 ] Oliver C, Bernhard S, Alexander Z. Semi-Supervised Learning[J]. The MIT Press, 2006,172(2):530-530.
- [ 10 ] Zhu X J. Semi-Supervised Learning Literature Survey [D]. Madison:University of Wisconsin, 2008.
- [ 11 ] Zhou Z H. Ensemble Methods:Foundations and Algorithms Boca Raton[M]. Boca Raton: CRC Press, 2012.
- [ 12 ] 王惠英,孙中平,孙志伟,等.基于深度学习的河道提取与变化监测应用——以永定河为例[J].北京测绘,2019,33(2):173-178. [ Wang H Y, Sun Z P, Sun Z W, et al. The Application for extraction of river channels and change detection based on deep learning: A case study from Yongding River[J]. Beijing Surveying and Mapping, 2019,33(2):173-178. ]
- [ 13 ] Xu Y Y, Xie Z, Wu L, et al. Multilane roads extracted from the Openstreet Map urban road network using random forests,wiley[J].Translation in GIS, 2018. DOI: 10.1111/12514.
- [ 14 ] 周龙,韦素媛,崔忠马,等.基于深度学习的复杂背景雷达图像多目标检测[J].系统工程与电子技术,2019,41(6): 1258-1264. [ Zhou L, Wei S Y, Cui Z M, et al. Multi-objective detection of complex background radar image based on deep learning[J]. Systems Engineering and Electronics, 2019,41(6):1258-1264. ]
- [ 15 ] Hu B, Xu Y Y, Wan B, et al. Hydrothermally altered mineral mapping using synthetic application of Sentinel-2A MSI, ASTER and Hyperion data in the Duolong area, Tibetan Plateau, China[J]. Ore Geology Reviews, 2018,101: 384-397.
- [ 16 ] 王圆圆,李京.基于决策树的高管狗数据特征选择及其对分类结果的影响分析[J].遥感学报,2007,11(1):69-76. [ Wang Y Y, Li J. Analysis of Feature Selection and Its Impact on Hyperspectral Data Classification Based on Decision Tree Algorithm[J]. Journal of Remote Sensing, 2007,11(1):69-76. ]
- [ 17 ] 陈永良,周斌,李学斌.基于 Boltzmann 机的矿产靶区预测[J].地球物理学进展,2012,27(1):179-185. [ Chen Y L, Zhou B, Li X B. Mineral target prediction based on Boltzmann machines[J]. Progress in Geophysics, 2012,27(1):179-185. ]
- [ 18 ] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006,313(5786):504-507.
- [ 19 ] Xu Y, Chen Z, Xie Z, et al. Quality assessment of building footprint data using a deep autoencoder network[J]. International Journal of Geographical Information Science, 2017,31(10):1929-1951.
- [ 20 ] Fiore U, Palmieri F, Castiglione A, et al. Network anomaly detection with the restricted Boltzmann machine[J]. Neurocomputing, 2013,122:13-23.
- [ 21 ] Chen H, Murray A F. Continuous restricted Boltzmann machine with an implementable training algorithm[J]. Vision Image and Signal Processing, IEE Proceedings, 2013, 150:153-158.
- [ 22 ] Sun J, et al. Application of deep belief networks for precision mechanism quality inspection[M]. Berlin: Springer, 2014.