

<https://doi.org/10.3799/dqkx.2020.297>



基于半监督神经网络的铜矿预测方法

徐永洋^{1,2}, 李孜轩¹, 谢忠^{1,2*}, 冯斌^{1,3}, 陈浩⁴

1. 中国地质大学地理与信息工程学院, 湖北武汉 430078

2. 中国地质大学国家地理信息系统工程技术研究中心, 湖北武汉 430078

3. 中国地质科学院地球物理地球化学勘查研究所, 河北廊坊 065000

4. 武汉中地先进技术研究院, 湖北武汉 430074

摘要: 将人工智能技术引入成矿预测研究中, 可以提高预测效率, 挖掘探测数据与结果之间的隐藏信息. 利用半监督学习方法对样本构建要求低的优点, 结合其在异常识别方面的应用效果, 设计了基于分割准则的孤立森林与深度自编码网络的神经网络结构; 基于西藏冈底斯地区的化探元素数据, 对研究区内的铜矿进行了成矿预测工作, 预测结果与已知矿区数据叠加效果较好, 说明本文的神经网络结构能够完成成矿远景区的预测工作.

关键词: 铜矿; 成矿预测; 深度学习; 半监督神经网络.

中图分类号: P3

文章编号: 1000-2383(2020)12-4563-11

收稿日期: 2020-07-20

Prediction of Copper Mineralization Based on Semi-Supervised Neural Network

Xu Yongyang^{1,2}, Li Zixuan¹, Xie Zhong^{1,2*}, Feng Bin^{1,3}, Chen Hao⁴

1. School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China

2. National Engineering Research Center of Geographic Information System, China University of Geosciences, Wuhan 430078, China

3. Institute of Geophysical and Geochemical Exploration, Chinese Academy of Geological Sciences, Langfang 065000, China

4. Wuhan Zondy Cyber Technology Co. Ltd., Wuhan 430074, China

Abstract: The introduction of artificial intelligence technology into metallogenic prediction can improve the prediction efficiency and research the hidden information between the exploration data and the results. In this paper, it designs a brand new semi-supervised neural network structure based on isolation forest with split-selection criterion (SCiForest) and autoencoder, using the advantage of semi-supervised learning method that the samples are easy to construct and the outstanding application effect in anomaly recognition. Based on the geochemical element data of Gangdise area in Tibet, the copper deposits in the study area is predicted. The prediction results are basically within the known mining area, which shows that the neural network structure in this paper can be used to predict the metallogenic prospective area.

Key words: copper deposit; metallogenic prediction; deep learning; semi-supervised neural network.

基金项目: 国家自然科学基金(No.41671400); 国家重点研发计划项目(Nos.2018YFB0505500, 2018YFB0505504); 开放基金(No.CY119R015); 地质探测与评估教育部重点实验室开放基金(No.GLAB2020ZR05).

作者简介: 徐永洋(1989—), 男, 副教授, 主要从事智能空间认知方面的研究. ORCID: 0000-0001-7421-4915. E-mail: yongyangxu@cug.edu.cn

* 通讯作者: 谢忠, E-mail: xiezhong@cug.edu.cn

引用格式: 徐永洋, 李孜轩, 谢忠, 等, 2020. 基于半监督神经网络的铜矿预测方法. 地球科学, 45(12):4563-4573.

1 研究现状简介

在国家经济和社会的快速发展进程中,矿产资源的作用极其重要,基础地位很难被其他资源所代替.我国矿产资源种类丰富,总量也处于世界领先地位(麻志周,2009).铜矿在我国分布较为普遍.自然界中出现的含铜矿物约有280多种,其中16种具有工业意义(刘小舟,2007),其在电器、机械、车辆、船舶工业和民用器具等方面都有着广泛的应用.因此,对于铜矿的成矿预测研究就显得尤为重要.

随着我国发展速度逐渐加快,矿产资源的需求也不断加大,而储量却在日益减少,找矿难度也日益加大.矿产远景制图(mineral prospectivity mapping, MPM)是应用较为广泛的成矿预测方法,它根据研究区的各种地球物理、地球化学资料,综合分析研究区地质特征与各种矿物类型、规模的关系,阐明成矿规律,预测该地区可能存在的矿产的有利条件,并为下一阶段的一般矿产预测提供依据.自20世纪20年代以来,矿产远景制图引起了许多研究者的关注(Allais, 1957; Cargill and Clark, 1978; Zhao, 2007; Tessema and Abera, 2017).

随着信息技术的不断发展,机器学习被逐步引入到矿产远景制图,如神经网络(Xiong and Zuo, 2016)、逻辑回归(Carranza and Laborte, 2015; Xiong and Zuo, 2018)、随机森林(Carranza and Laborte, 2016; Hariharan *et al.*, 2017)等,这些方法大多采用分类或回归的思想对研究区内的数据进行处理.近年来,人工神经网络开始在各个领域展示出独特的优势(潘和平和刘国强, 1997; 刘文龙等, 2000; Moser *et al.*, 2013; Wang *et al.*, 2015; Xu *et al.*, 2017; 程国轩, 2018; Xu *et al.*, 2018).识别和预测是深度学习的两个基本功能(Lecun *et al.*, 2015),因此,一些深度学习方法也开始被用于处理地球化学勘探数据和挖掘地球化学模式(Zuo *et al.*, 2019),利用神经网络进行成矿预测也取得了长足的进展. Shao *et al.* (2007)利用BP神经网络实现了金矿床的成矿预测; Chen *et al.* (2014)使用连续限制玻尔兹曼机(continuous restricted boltzmann machine, CRBM)模型来学习已知矿床与证据图之间的关系.深度学习方法与地理信息科学(geographic information science, GIS)相结合的方法也开始出现,用于分析地球化学元素之间的关系; Quan *et al.* (2013)同时考虑了地球化学元素的异常和空间分布,利用基

于GIS的人工神经网络(artificial neural network, ANN)对塔里木盆地北缘铀成矿进行了预测.如此诸多的研究也证明了深度学习神经网络在成矿预测领域广阔的应用前景.

金属矿产在形成过程中,会伴随着或多或少的分布异常.元素在进行分布时,往往具有富集特征,这种特征在绝大多数的找矿方法中都是不能被忽视的重要指示信息.监督学习与非监督学习是深度学习的两大热门领域,监督学习以其卓越的准确性在各大领域应用甚广,而非监督学习大幅缩减了人工成本,提高了训练效率,也成为深度学习重要的组成部分.非监督神经网络最为典型的应用领域就是数据异常识别,该应用与成矿中的关键信息——矿物元素异常相对应,这也说明将非监督学习用于成矿预测是完全可行的.

为了提高成人工智能在成矿预测的准确性与高效性,本文基于非监督神经网络中的自编码器(autoencoder, AE)与基于分割准则的孤立森林(isolation forest with split-selection criterion, SCiForest),同时借鉴生成对抗网络(generative adversarial networks, GAN)的思想,设计全新的半监督异常检测神经网络.设计的半监督神经网络能够综合监督学习与非监督学习的优势,在保证预测结果准确性的同时,极大程度地减少了人工成本,提高工作效率.西藏地区的地质条件复杂,蕴含信息丰富,近年来在该地区已经开展了针对不同角度的成矿研究工作(焦彦杰等, 2019; 黄勇等, 2019).采用本文的方法,利用化探数据对西藏冈底斯研究区的铜矿进行了成矿预测,预测的结果与成矿规律基本吻合;通过预测结果与已知矿区叠加验证,结果表明本文提出的半监督异常识别方法能够对实际生产中的成矿预测工作起到指导意义.

2 理论方法与实验验证

本文提出方法基于SCiForest与AE构造出一种全新的半监督学习神经网络结构,实现了对高维数据的异常检测.数据集首先需要经过SCiForest训练,训练值较小的数据被认为是背景数据.然后选择SCiForest中一定数量的背景数据对自编码器进行训练,经过背景数据训练后的自编码器对预测数据中的背景数据会有良好的适应性,增加了模型对异常数据的敏感性.最后将全部待预测数据带入经过训练后的自编码器,结果中重构误差较大的点

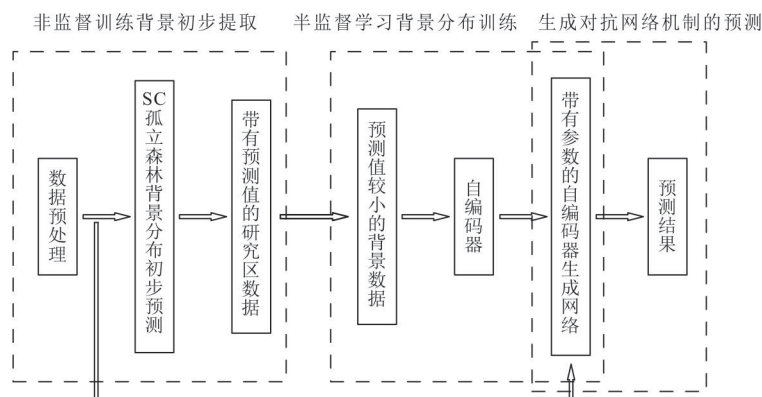


图1 基于非监督学习成矿预测流程图

Fig.1 Flow chart of metallogenetic prediction based on unsupervised learning method

即为异常点.本文提出的方法顾及了单一模型对数据弱异常检测能力低的不足之处,结合 SCiForest 与 AE 优势对化探数据进行深度分析,基于分析结果进行成矿预测.本文方法的流程图如下(图1).

2.1 基于非监督学习的背景预处理

为了提高对背景样本分布的无监督训练效率,首先需要对数据进行背景提取.本文中将有化探元素含量异常的区域认为是背景样本.首先利用非监督学习方法(SCiForest)进行背景识别.

SCiForest 是孤立森林(isolation forest)方法的发展,其能够应用于对高维数据的一元分类.在构建孤立树(isolation tree)时,孤立森林在子采样的样本中随机选取一个属性计算节点之间的距离,而 SCiForest 则构建了一个包含多个属性的超平面 f ,利用超平面对子采样集进行分割,由此实现对高维度数据的处理.相比于其他非监督聚类算法,SCiForest 避免了繁复的样本点距离或密度的计算,对于正负样本分布不均衡的小样本有很好的适应性,对聚类异常也有良好的识别效果.

化探数据中土壤水系样本是对多种元素探测的结果,可以理解为某个探测点高维度的特征描述.土壤水系采样法有确定的采样点布置原则,研究区内按照一定比例尺划分窗格布设采样点,用此采样法得到的采样点均匀分布在整個研究区内,采样点的特点与区域成矿信息相符合,正负样本分布不均衡;并且同一区域的成矿模式相近,一定范围内异常点的异常形式也相似,形成一定规模的聚类异常.充分考虑以上数据特点,SCiForest 被引入该研究进行背景识别.

假定 $X = \{x^1, \dots, x^n\}$ 为一个 d 维度的分布,

其中的每个实例 $x^i = \{x_1, \dots, x_d\}$, 其中 $i = 1, \dots, n$. 与孤立森林相同, SCiForest 不适用于大数据集的训练(Liu *et al.*, 2010). 因此,在创建孤立树时需要先对数据集进行采样: $X' \subset X, |X'| = \psi \dots$; 其中 X' 为采样后的子数据集, ψ 为一小常数,在孤立森林的经验下通常为 256(Liu *et al.*, 2010). 在构造一棵树时,其每个分裂点处都需要在 τ 个超平面中选择出一个最佳分裂点 p . 超平面 f 的定义如下:

$$f(x) = \sum_{j \in Q} c_j \frac{x_j}{\sigma(x'_j)} - p, \quad (1)$$

其中: Q 具有 q 个从 $\{1, \dots, d\}$ 中随机选择的属性索引, c_j 是从 $[-1, 1]$ 中随机选择的系数, x'_j 是采样集 X' 中的第 j 个属性值. 选择最佳分裂点时的准则被定义为 Sd_{gain} , 表示为:

$$Sd_{gain}(Y) = \frac{\sigma(Y) - avg(\sigma(Y^l), \sigma(Y^r))}{\sigma(Y)}, \quad (2)$$

其中: $Y^l \cup Y^r = Y$; Y 是通过将 X 投影到超平面 f 上而获得的一组真实值. $\sigma(\cdot)$ 是标准差函数, $avg(a, b)$ 返回 $(a+b)/2$. 分裂点 p 将 Y 分割成 Y^l 与 Y^r , 即 $y^l < p < y^r, y^l \in Y^l, y^r \in Y^r$.

与孤立森林相同, SCiForest 中样本点的异常都是根据其在孤立树上的路径长度 $h(x)$ 计算得到. $h(x)$ 表示从根节点到叶节点的过程中 x 经过的边(edge)的个数. 对于 t 棵树的路径长度的期望 $E(h(x))$ 能够用于表示异常. 路径长度较长的点代表正常点,而路径长度越短的点则越有可能是异常点.

经过 SCiForest 初步训练,能够得到针对于研究区内所有数据点的异常值大小. 异常值越大,说明该数据点的化探元素含量异常越大. 选取 SCiForest 训练后异常值较小的数据点作为背景样

本数据集,用来训练自编码器,从而实现更精确的描绘背景分布情况.

2.2 半监督学习背景样本分布学习

经过上一个步骤的训练,能够初步提取数据中的异常点.通过选取经过 SCiForest 训练后异常值较小的数据点制作成数据集构成半监督学习在训练时所需要使用的正常样本数据,去训练构建的深度自编码神经网络(autoencoder).自编码神经网络是一种应用十分广泛的非监督学习神经网络模型,通过多层神经网络对输入数据进行编码和解码.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.其结构主要分为两大部分:编码(encode)和解码(decode).在编码过程中,输入数据经过多层的限制玻尔兹曼机(RBM)降低维度,提取特征.输入数据 D_I 的编码过程可以用如下函数表示:

$$I=f(W^eD_I+\delta^e), \tag{3}$$

其中: I 代表编码过程中的隐藏层, W^e 和 δ^e 分别表示编码过程的权重矩阵和偏置向量.解码过程与编码过程类似,其作用是将编码过程提取出的特征进行重构,使维度还原至输入时的大小.解码过程也可以用函数表示为:

$$D_o=g(W^dI+\delta^d), \tag{4}$$

其中: D_o 为解码后的输出结果, W^d 和 δ^d 分别表示解码过程的权重矩阵和偏置向量.

自编码器的训练目标是通过优化参数减小重

建与初始值之间的误差.通常,基于自编码器的方法中都采用均方误差(MSE)作为损失函数,用于量化重构数据与原始输入数据的偏差.均方误差的计算方法如下:

$$MSE(y,y')=\frac{\sum_{i=1}^m(y_i-y'_i)^2}{m}, \tag{5}$$

其中: y_i 是每一个训练批次的数据中第*i*个值, y'_i 为其对应的预测值, m 为样本数.自编码器的重构误差用于确保神经网络的性能,其由MSE构建而来,定义为:

$$L_r=\|D_I-D_o\|_2. \tag{6}$$

较大的重构误差对应着明显异常的数据输入.背景样本数据经过自编码器训练后,整体的重构误差达到最小,意味着自编码器能够较准确地重构出背景分布.经过训练后的自编码器及其参数能够作为下一步生成对抗网络的生成器参与最后的异常检测工作.利用自编码器进行半监督学习背景分布训练的网络结构图如图2所示.

2.3 生成对抗网络机制的异常识别

在进行预测时,我们还需要一个鉴别器来对研究区内全部数据的预测结果进行评价.根据实际需要,结合自编码器自身特点,本文对神经网络的结构进行了改进,使其能够完成训练—鉴别的完整的半监督学习目的,该网络被定义为 encoder-decoder.类似于生成对抗网络(generative adversarial networks,GAN),自编码器编码和解码的训练过

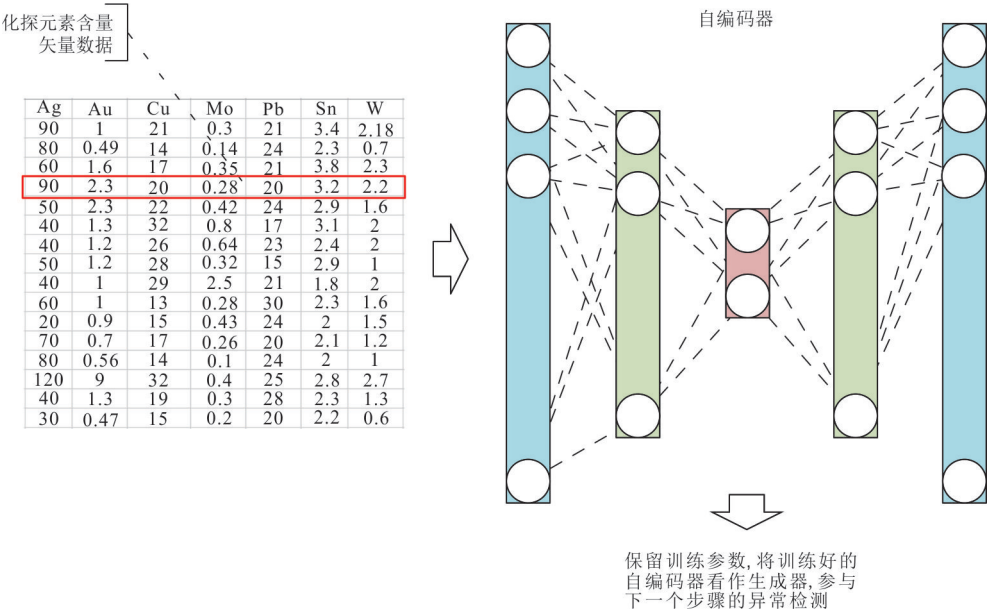


图2 自编码器神经网络结构
Fig.2 Network structure of Autoencoder

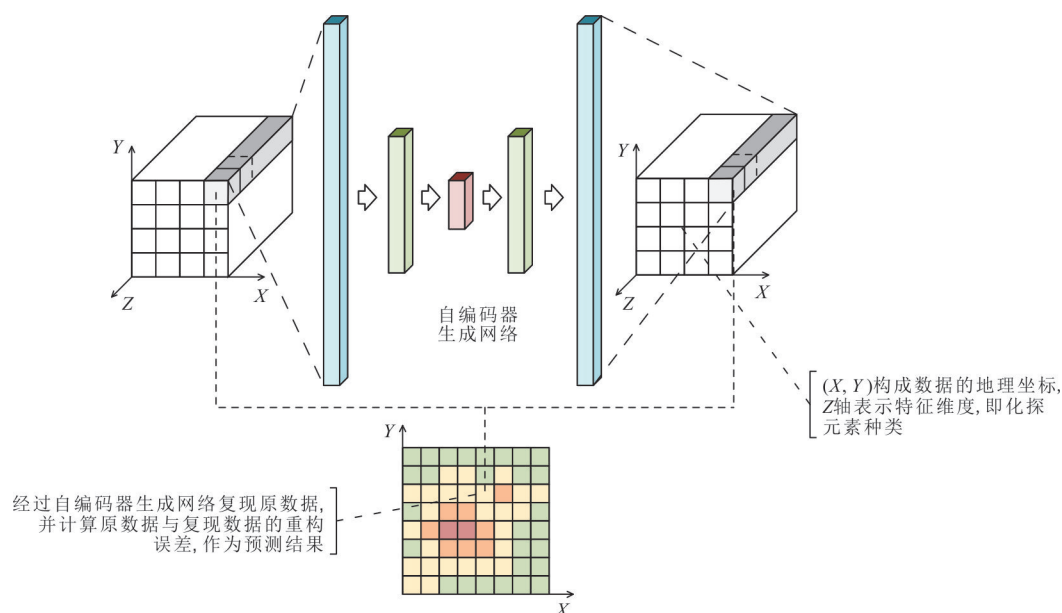


图3 生成对抗网络机制的自编码器网络结构

Fig.3 Structure of encoder-decoder GAN

程被用作生成器.该网络能够根据训练得到的参数重构出待预测数据的分布情况,并且利用重构误差进行异常检测.由于编码和解码过程中的参数是由之前的背景样本数据训练得到,因此在对所有的待预测数据进行重构时,属于背景的区域重构误差小,异常的区域重构误差较大.生成对抗网络机制的自编码器网络结构如图3.

3 实验结果

3.1 实验数据

研究区位于冈底斯成矿带,地处西藏自治区中部,内有3 135个采样点.研究区内每个采样点的经度间隔为 0.02° ,纬度集中在北纬 $29^\circ \sim 30^\circ$.东西向距离大约为1.92 km;南北向两采样点之间的经度差为 0.02° ,距离约为2 km.研究区内数据均由土壤水系采样法采样得到,包含Au、Ag等32种元素以及 Al_2O_3 、CaO、 Fe_2O_3 、 K_2O 、MgO、 Na_2O 、 SiO_2 等化合物的含量.由于地质特征在空间上变化具有无规律性,每个采样点的数据仅能代表该点的地质特征,相邻采样点间并没有可依据的变化趋势;并且采样点间距也较大,因此尽管本次研究的数据量较小,仍不能够对原始数据进行插值扩充数据集.因此在选择神经网络方法时,我们需要依据之前的讨论,选择合适的神经网络方法,尽可能避免小样本数据导致的问题.

冈底斯铜矿带内的铜矿床类型主要为斑岩型,次为矽卡岩型和热液脉型,矿床在空间上具有一定的分带性(李光明等,2002).与斑岩有关的矿化类型分为两种:单一的细脉浸染型矿化,斑岩型和矽卡岩型复合型矿化.细脉浸染型矿化的矿化规模一般较大,矿化元素含量普遍较低,除了伴生有Au、Ag、Pb、Zn、Mo、W和Sn等元素外,尚可综合回收Ag、Re、Co、S、Se和Te等元素.采样的数据集中包含有上述指示性元素中的Au、Ag、Pb、Zn、Mo、W、Sn、Co和Cu,因此我们选择这9种元素的研究区数据作为数据集进行预测工作.

3.2 实验结果

首先,将研究区内所有数据点带入SCiForest进行训练.根据经验指导,当孤立树的棵数大于10棵的时候,其训练的结果开始趋向收敛(Liu *et al.*, 2010).随着树的数目增加,训练时间会略有增加,但不会影响到训练的结果.因此,本文设置SCiForest的孤立树数为20.由于矿区数据点中异常数据偏少,SCiForest的训练结果中标签为1的数据理论上只占有很小一部分,因此为了将标签为-1的数据区分开以选择更为“纯净”的背景,我们将SCiForest的输出调整为连续的预测值.样本点的预测值分布情况如图4.

从图4中可以看出,样本点异常值分布的拟合曲线(图中红色虚线)基本与正态分布相吻合.相比于一般的一元异常检测方法,SCiForest对于异常更

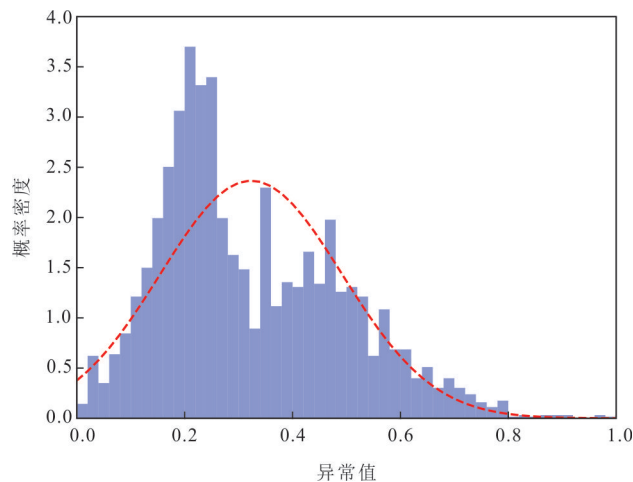


图4 SCiForest 预测结果的预测值分布
Fig.4 Value distribution of prediction results by SCiForest

为敏感,同时,SCiForest也能够完成对聚类异常的识别任务.这也说明,SCiForest的预测结果中预测值较小的点为非异常点的可能性更高,对于背景点的预测更为可靠.预测值越小,该点作为正常点的概率越高.考虑到预测值越小,所选取的数据点是正常点的概率越大,数据集越纯净,但预测值过小时,会导致选取的数据点个数过少,进而使得自编码器的训练数据集过小,增大过拟合的可能性,导致预测效果变差.因此,在充分考虑了这些因素之后,我们首先选取预测值小于0.4的数据作为自编码器的训练数据集,观察预测效果.

经过 SCiForest 的预处理,本文将筛选出的数据作为背景样本数据点对自编码器中进行训练,以得到神经网络的相关参数.由于研究区内的数据以各点为独立单位,每个采样点包含9种元素的含量数据,因此带入到神经网络中的每个数据维度为 1×9 .在本实验中,自编码器设置有3层RBM,维度分别为50-20-5.在训练时,为了使训练损失实现收敛,自编码器的epoch设置为5 000,每一个批次有50个数据进行训练.从训练损失图(图5)可以看出,在训练5 000次后,损失基本收敛.

自编码器的参数训练完成后,原数据中没有异常的数据点,重构误差较小;原数据中存在化探元素含量异常的点,重构误差较大.基于此,便能够找出研究区内化探元素异常的点,化探元素含量异常就代表了该点可能为矿点.图6展示了重构结果中研究区内各点的重构误差分布.

结果显示,重构误差与探测点数之间基本符合正态分布,这与实际生产生活中的经验相吻合,体

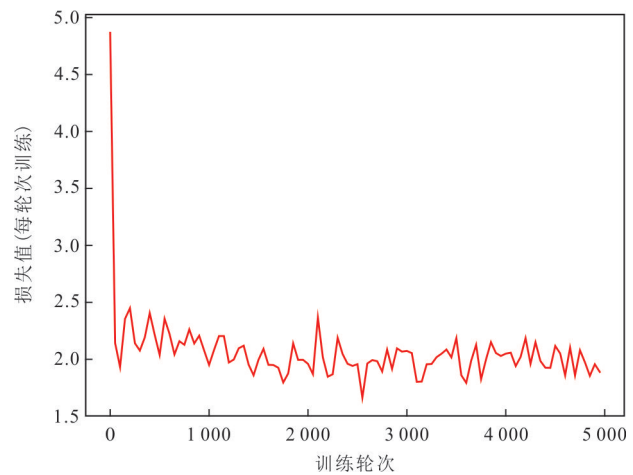


图5 基于 SCiForest 预测值小于0.4的数据训练的自编码器的迭代训练损失
Fig.5 Iterative training loss of Autoencoder based on data training with SCiForest prediction value less than 0.4

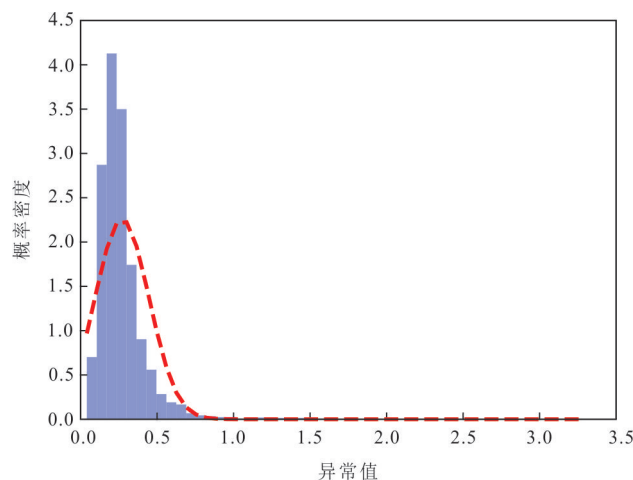


图6 自编码器生成对抗网络的重构误差分布
Fig.6 Reconstruction error distribution of the encoder-decoder GAN

现出预测结果的实际参考价值.绝大部分点的异常值均位于1以下,而采样点最大异常值达到了3.29,极差较大,因此由于图幅的原因,异常值较大的点未能在图中较好的显示.利用 ArcGIS 软件的反距离权重法(inverse distance weight, IDW)功能,对研究区数据的重构误差进行了插值可视化,效果如图7a所示.

初步的预测结果中,预测值达到0.23时,插值显示的结果就已经将这些区域标定为存在异常,说明自动插值的结果与实际情况不符;尽管异常值位于0.23~0.34的区域被自动插值为低异常区,大面积的低异常区也不利于凸显出更有意义的高值异

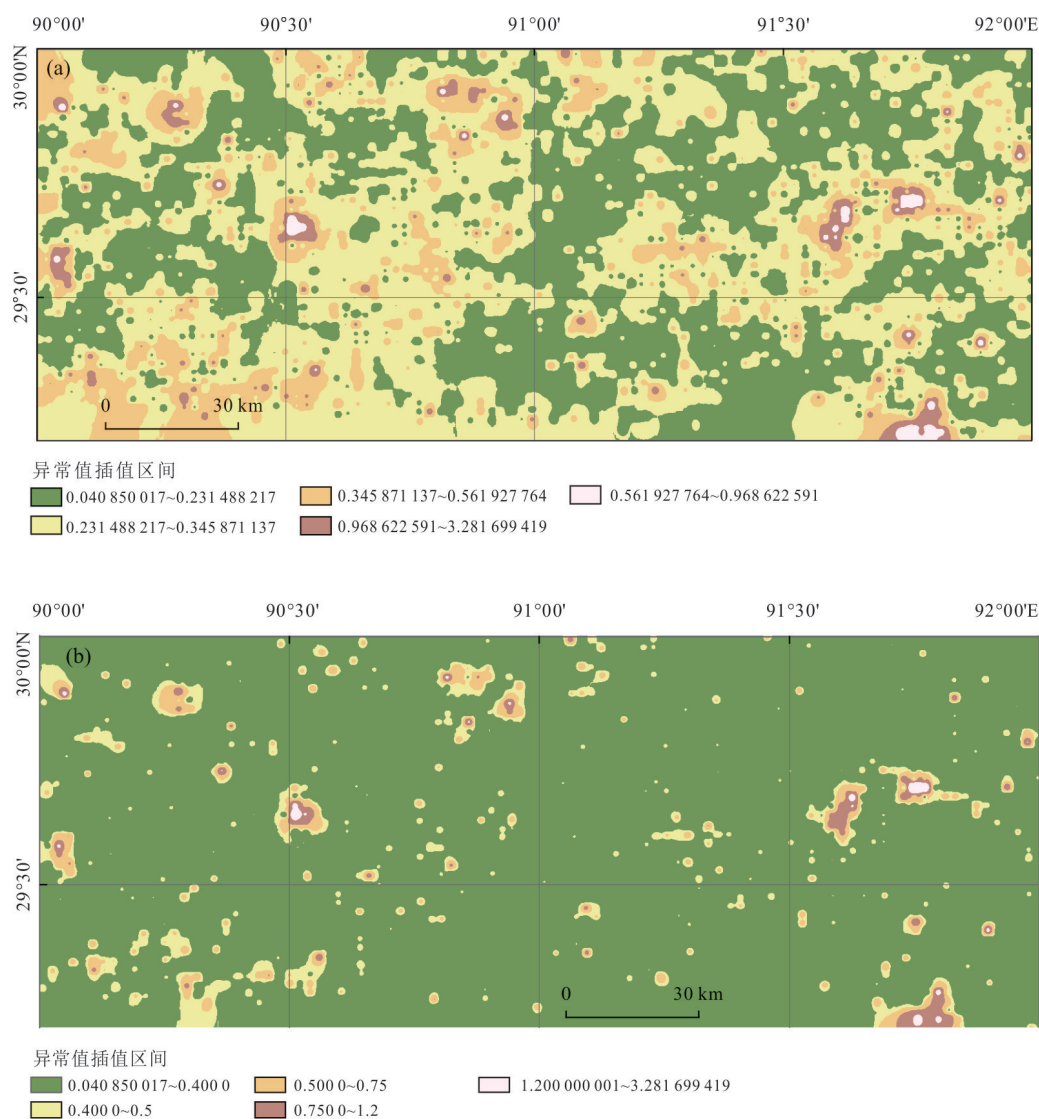


图 7 基于 SCiForest 预测值小于 0.4 的数据训练的半监督神经网络预测结果的插值可视化

Fig.7 Interpolation visualization of prediction results of semi-supervised neural network based on data training with SCiForest prediction value less than 0.4

a. 未插值优化; b. 插值优化

常. 因此, 我们在不改变神经网络预测值的前提下, 根据数据的分布情况, 结合成矿远景区的判别经验, 优化插值区间, 使结果的以更好的呈现. 优化后的插值可视化结果如图 7b 所示.

由于上述情况的出现, 为了更好地评价本次实验的结果, 我们继续将预测结果与研究区内的矿区信息进行叠加, 效果如图 8b 所示.

从矿区的叠加结果显示, 预测结果中的高值区在矿区的范围内均有分布. 实验结果分析表明, 半监督学习神经网络能够利用化探元素数据对研究区进行铜矿的成矿预测, 神经网络能够识别出采样点较明显的异常的化探元素含量, 预测的结果能够

对下一步成矿远景区的圈定起到指导作用.

通常情况下, SCiForest 判定正常值的标准阈值为 0.5, 因此我们继续选择 SCiForest 预测值小于 0.5 的数据作为自编码器的训练集, 经过同样的训练与复现、可视化步骤后, 得到预测结果叠加矿区的示意图如图 8c 所示. 插值区间优化的标准同样保持一致. 从预测结果中我们发现, 有些矿区内没有预测出异常, 这说明当在选择 SCiForest 预测值小于 0.5 的数据作为正常数据时, 正常数据的标准被降低, 训练数据混入了一定数量的异常值, 经过自编码器训练后, 这些异常区域也被较好的复现.



图8 预测结果叠加矿区示意

Fig.8 Schematic diagrams of prediction results superimposed mining area

a. 基于SCiForest预测值小于0.3的数据训练的半监督神经网络预测结果的插值可视化;b.SCiForest预测值小于0.4;c.SCiForest预测值小于0.5

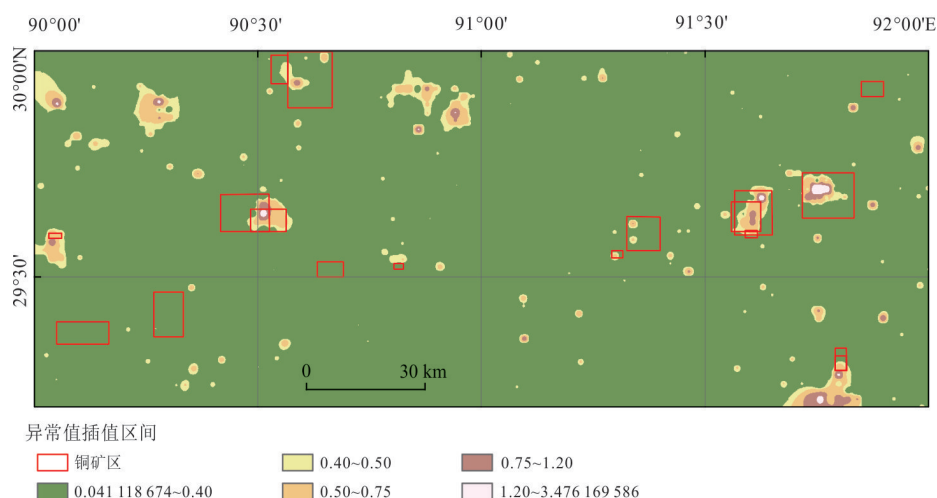


图9 基于DBSCAN的半监督神经网络预测结果的插值可视化

Fig.9 Interpolation visualization of prediction results of semi-supervised neural network based on DBSCAN

为了最优化自编码器的训练数据集,我们进一步提高正常数据的判定标准,选择了 SCiForest 预测值小于 0.3 的数据作为自编码器的训练集进行对比试验,结果如图 8a 所示.插值标准与之前的实验保持一致.从图 8 中可以看出,预测结果中出现了大面积的离散点,说明神经网络对训练数据产生了过拟合现象,参与训练的数据复现效果好,未参与训练的数据的重构误差被放大.

同时,为了衡量 SCiForest 在本次研究中的效果,我们继续选择了 DBSCAN 作为背景预处理的对比方案.根据成矿预测的特点,DBSCAN 聚类中采样点数量占绝大多数的类别被看作正常点聚类,因此我们选择了该类的采样点数据作为自编码器的训练数据.经过与上文相同的训练条件,在相同插值显示条件下,利用 DBSCAN 构建半监督神经

网络的预测结果如图 9 所示.

结果显示,尽管大部分矿区内确实存在预测高值区,但是位于研究区西南的两处已知矿区与东北角的一处矿区内并没有成功预测出异常,这就凸显出 SCiForest 能够预测聚类异常的优势,也说明选择 SCiForest 对数据进行背景预处理是更为合适的.此外,我们还利用了基于密度的非监督算法 LOF 进行尝试,其结果并不能准确预测出研究区内的异常,因此不做详细介绍.

经过以上对比试验,我们最终采用 SCiForest 预测值小于 0.4 的数据的预测值作为下一步工作的参考.对图 9 的预测结果进一步分析,发现除了预测结果与已知矿区的对应区域外,还存在有五处异常高值区,这些区域在图 10 中标出.

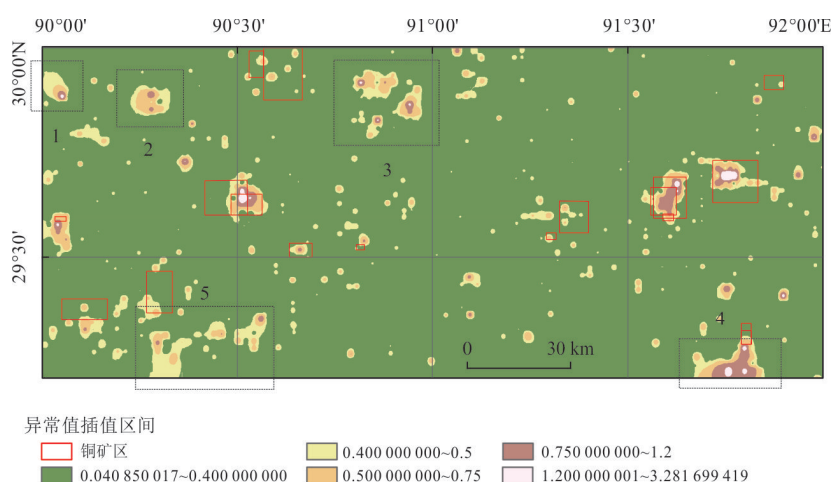


图10 预测出的潜在矿区示意

Fig.10 Predicted potential mining area

根据成矿远景区的定义,对于所有 3 个等级的成矿远景区,“具有一定的规模”都是判断的一项重要准则;同时对于 I、II 级成矿远景区来说,异常的强度也有一定的要求。因此,综合预测结果与成矿远景区的判断准则,本文认为图 10 中的 1、2、3 三处区域都有较大的成矿可能性;4 区域北部虽然有小面积矿区与矿点,但是从预测结果中来看,可能还存在更广阔的矿产资源开采潜力。5 区域与 4 区域类似,其附近除了存在已探明的矿区外,区域内同样存在其余成矿可能性。其余区域虽然存在一定数量的离散异常高值点,但是它们的规模普遍不够大,或是异常值并不突出。虽然从预测结果中已知矿区内的异常情况来看,并不能完全否定这些离散点的成矿可能,但相比于以上列出的 5 个预测远景区域,这些离散点的成矿情况还需要进一步实地勘探。

以上实验证明了本文所搭建的半监督学习异常识别神经网络对冈底斯地区的铜矿成矿有良好的预测效果,预测结果不仅体现出与已知矿区较高的吻合度,同时还根据预测结果指出了 5 个具有潜在成矿前景的预测区。这对于研究区内铜矿的探测与后续的开采工作起到了指导作用,对于冈底斯地区铜矿资源的利用具有一定的意义。

4 总结

本文提出了一种基于半监督学习的异常检测神经网络结构,将监督学习的准确性与非监督学习的高效性有机结合起来。该结构包括了 SCiForest 与 AE 两种非监督神经网络,同时结合生成对抗网络的思想,将自编码器应用于 GAN 的生成网络,利用自编码器的重构误差实现了对数据的异常检测。利用该神经网络对西藏冈底斯铜矿区内的 3 135 探测点化探数据进行异常识别,得到了研究区内各采样点处的异常值。本次研究采样的比例尺偏大,因此数据量偏少,在今后的工作中我们能够通过增大采样时的比例尺来扩充数据集,进一步提升预测结果的精度。化探元素含量的异常通常是成矿的指示性标志,因此该异常值的大小能够代表采样点处成矿的可能性大小。但是化探元素的异常也需要有系统性的体现,在高维度数据中出现的单元异常往往不容易被神经网络识别出来。结果显示,本文提出的半监督异常检测神经网络对于研究区内铜矿的成矿预测有着良好的效果,该预测结果也将为下一阶段成矿远景区的划分提供参考。

References

- Allais, M., 1957. Method of Appraising Economic Prospects of Mining Exploration over Large Territories: Algerian Sahara Case Study. *Management Science*, 3(4): 285–347. <https://doi.org/10.1287/mnsc.3.4.285>
- Cargill, S. M., Clark, A. L., 1978. Report on the Activity of IGCP Project 98. *Mathematical Geology*, 10(5): 7.
- Carranza, E. J. M., Laborte, A. G., 2015. Data-Driven Predictive Mapping of Gold Prospectivity, Baguio District, Philippines: Application of Random Forests Algorithm. *Ore Geology Reviews*, 71: 777–787. <https://doi.org/10.1016/j.oregeorev.2014.08.010>
- Carranza, E. J. M., Laborte, A. G., 2016. Data-Driven Predictive Modeling of Mineral Prospectivity Using Random Forests: A Case Study in Catanduanes Island (Philippines). *Natural Resources Research*, 25(1): 35–50. <https://doi.org/10.1007/s11053-015-9268-x>
- Chen, Y. L., Lu, L. J., Li, X. B., 2014. Application of Continuous Restricted Boltzmann Machine to Identify Multivariate Geochemical Anomaly. *Journal of Geochemical Exploration*, 140: 56–63. <https://doi.org/10.1016/j.gexplo.2014.02.013>
- Cheng, G. X., Niu, R. Q., Zhang, K. X., et al., 2018. Opencast Mining Area Recognition in High-Resolution Remote Sensing Images Using Convolutional Neural Networks. *Earth Science*, 43(Suppl. 2): 256–262 (in Chinese with English abstract).
- Hariharan, S., Tirodkar, S., Porwal, A., et al., 2017. Random Forest-Based Prospectivity Modelling of Greenfield Terrains Using Sparse Deposit Data: An Example from the Tanami Region, Western Australia. *Natural Resources Research*, 26(4): 489–507. <https://doi.org/10.1007/s11053-017-9335-6>
- Huang, Y., Fu, J. G., Li, G. M., et al., 2019. Determination of Lalong Dome in South Tibet and New Discovery of Rare Metal Mineralization. *Earth Science*, 44(7): 2197–2206 (in Chinese with English abstract).
- Jiao, Y. J., Huang, X. R., Li, G. M., et al., 2019. Deep Structure and Mineralization of Zhaxikang Ore-Concentration Area, South Tibet: Evidence from Geophysics. *Earth Science*, 44(6): 2117–2128 (in Chinese with English abstract).
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. *Nature*, 521: 436–444. <https://doi.org/10.1038/nature14539>
- Li, G. M., Wang, G. M., Gao, D. F., et al., 2002. Perspective of Copper Ore Deposit and Its Exploration Direction in Gangdise Metallogenic Belt, Tibet. *Mineral Deposits*, 21

- (Suppl.1):144—147(in Chinese with English abstract).
- Liu, F. T., Ting, K. M., Zhou, Z. H., 2010. On Detecting Clustered Anomalies Using SCiForest. *Machine Learning and Knowledge Discovery in Databases*. Springer, Berlin Heidelberg, 274—290. https://doi.org/10.1007/978-3-642-15883-4_18
- Liu, W. L., Li, S. T., Sun, D. J., et al., 2000. Prediction of Pore-Fluid Pressure in Deep Strata of Songliao Basin. *Earth Science*, 25(2): 137—142(in Chinese with English abstract).
- Liu, X. Z., 2007. Copper Resources in China: Today and Tomorrow. *Northwestern Geology*, 40(1): 83—88(in Chinese with English abstract).
- Ma, Z. Z., 2009. Thoughts on the Guarantee of Mineral Resources in China. *Land and Resources Information*, (3): 2—7(in Chinese with English abstract).
- Moser, G., Serpico, S. B., Benediktsson, J. A., 2013. Land-Cover Mapping by Markov Modeling of Spatial-Contextual Information in Very-High-Resolution Remote Sensing Images. *Proceedings of the IEEE*, 101(3): 631—651. <https://doi.org/10.1109/JPROC.2012.2211551>
- Pan, H. P., Liu, G. Q., 1997. Applying Back-Propagation Artificial Neural Networks to Predict Coal Quality Parameters and Coalbed Gas Content. *Earth Science*, 22(2): 210—214 (in Chinese with English abstract).
- Quan, X. D., Wang, Y. J., Wang, B., et al., 2013. Application of ANN and GIS in Uranium Metallization Prediction: A Case Study of Northern Tarim Basin. *Uranium Geology*, 29(6): 374—379.
- Shao, Y. J., He, H., Zhang, Y. Z., et al., 2007. Metallogenic Prediction of Xiangxi Gold Deposit Based on BP Neural Networks. *Journal of Central South University (Science and Technology)*, 38(6): 1192—1198.
- Tessema, A., 2017. Mineral Systems Analysis and Artificial Neural Network Modeling of Chromite Prospectivity in the Western Limb of the Bushveld Complex, South Africa. *Natural Resources Research*, 26(4): 465—488. <https://doi.org/10.1007/s11053-017-9344-5>
- Wang, J., Song, J. W., Chen, M. Q., et al., 2015. Road Network Extraction: A Neural-Dynamic Framework Based on Deep Learning and a Finite State Machine. *International Journal of Remote Sensing*, 36(12): 3144—3169. <https://doi.org/10.1080/01431161.2015.1054049>
- Xiong, Y. H., Zuo, R. G., 2016. Recognition of Geochemical Anomalies Using a Deep Autoencoder Network. *Computers & Geosciences*, 86: 75—82. <https://doi.org/10.1016/j.cageo.2015.10.006>
- Xiong, Y. H., Zuo, R. G., 2018. GIS-Based Rare Events Logistic Regression for Mineral Prospectivity Mapping. *Computers & Geosciences*, 111: 18—25. <https://doi.org/10.1016/j.cageo.2017.10.005>
- Xu, Y. Y., Chen, Z. L., Xie, Z., et al., 2017. Quality Assessment of Building Footprint Data Using a Deep Autoencoder Network. *International Journal of Geographical Information Science*, 31(10): 1929—1951. <https://doi.org/10.1080/13658816.2017.1341632>
- Xu, Y. Y., Wu, L., Xie, Z., et al., 2018. Building Extraction in Very High Resolution Remote Sensing Imagery Using Deep Learning and Guided Filters. *Remote Sensing*, 10(1): 144. <https://doi.org/10.3390/rs10010144>
- Zhao, P. D., 2007. Quantitative Mineral Prediction and Deep Mineral Exploration. *Earth Science Frontiers*, 14(5): 1—10. <https://doi.org/10.1007/s11442-007-0020-2>
- Zuo, R. G., Xiong, Y. H., Wang, J., et al., 2019. Deep Learning and Its Application in Geochemical Mapping. *Earth Science Reviews*, 192: 1—14. <https://doi.org/10.1016/j.earscirev.2019.02.023>

附中文参考文献

- 程国轩, 牛瑞卿, 张凯翔, 等, 2018. 基于卷积神经网络的高分辨率遥感影像露天采矿场识别. *地球科学*, 43(增刊 2): 256—262.
- 黄勇, 付建刚, 李光明, 等, 2019. 藏南拉隆穹窿的厘定及其稀有多金属成矿作用新发现. *地球科学*, 44(7): 2197—2206.
- 焦彦杰, 黄旭日, 李光明, 等, 2019. 藏南扎西康矿集区深部结构与成矿: 来自地球物理的证据. *地球科学*, 44(6): 2117—2128.
- 李光明, 王高明, 高大发, 等, 2002. 西藏冈底斯铜矿资源前景与找矿方向. *矿床地质*, 21(增刊 1): 144—147.
- 刘文龙, 李思田, 孙德君, 等, 2000. 松辽盆地深层孔隙流体压力预测. *地球科学*, 25(2): 137—142.
- 刘小舟, 2007. 我国重要有色金属资源——铜矿的现状 & 展望. *西北地质*, 40(1): 83—88.
- 麻志周, 2009. 我国矿产资源保障问题的思考. *国土资源情报*, (3): 2—7.
- 潘和平, 刘国强, 1997. 应用 BP 神经网络预测煤质参数及含气量. *地球科学*, 22(2): 210—214.