

基于 AlexNet 网络的二维找矿预测 ——以松桃—花垣地区沉积型锰矿为例

李 诗¹, 陈建平¹, 向 杰², 张志平³, 张 烨³

LI Shi¹, CHEN Jianping¹, XIANG Jie², ZHANG Zhiping³, ZHANG Ye³

1. 中国地质大学(北京)地球科学与资源学院, 北京 100083;

2. 自然资源部成矿作用与资源评价重点实验室/中国地质科学院矿产资源研究所, 北京 100037;

3. 页岩气勘探开发国家地方联合工程研究中心(重庆地质矿产研究院), 重庆 401120

1. *School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing 100083, China;*

2. *Key Laboratory of Metallogeny and Mineral Assessment, MNR/Institute of Mineral Resources, CAGS, Beijing 100037, China;*

3. *National and Local Joint Engineering Research Center for Shale Gas Exploration and Development (Chongqing Institute of Geology and Mineral Resources), Chongqing 401120, China*

摘要:在大数据的时代背景下,地质大数据逐渐趋于复杂化的模式与其间的空间关联性为基于机器学习算法的矿产资源定量预测带来了更大的挑战。利用深度卷积神经网络算法优异的分析性能来提取不同成矿条件下多种二维要素图层的空间分布特征与关联性是一项非常有意义的探索性实验。以松桃—花垣地区沉积型锰矿为例,利用深度卷积神经网络模型 AlexNet 挖掘 Mn 元素、沉积相、大塘坡组出露、断裂及水系的分布与锰矿矿床的就位空间的耦合相关性,以及不同的控矿要素之间的相关性,以此训练出二维矿产预测分类模型。经过训练后,可以得到验证准确率 88.89%,召回率为 66.67%,损失值 0.08 的深度卷积神经网络分类模型。应用该模型对未知区进行二维成矿预测,共圈定出 91、96、154、184 号 4 个成矿远景区,其中 91 号和 154 号的区域含矿概率为 1,96 号含矿概率为 0.5。由此可见,预测区具有很大概率存在尚未发现的矿床。

关键词:大数据;找矿预测;卷积神经网络;Alexnet 网络;松桃—花垣锰矿

中图分类号:P618.32;P628 **文献标志码:**A **文章编号:**1671-2552(2019)12-2022-11

Li S, Chen J P, Xiang J, Zhang Z P, Zhang Y. Two-dimensional prospecting prediction based on AlexNet network: A case study of sedimentary Mn deposits in Songtao-Huayuan area. *Geological Bulletin of China*, 2019, 38(12): 2022-2032

Abstract: There are many challenges in the task of predicting ore deposits from big data repositories. The data are inherently complex and have great significance to the intervenient spatial relevance of deposits. The characteristics of the data make it difficult to use machine learning algorithms for the quantitative prediction of mineral resources. There are considerable interest and value in extracting spatial distribution characteristics from two-dimensional ore-controlling factors' layers under different metallogenic conditions. In this paper, the authors conducted such an analysis by using a Deep Convolutional Neural Network (D-CNN) algorithm named AlexNet. Training on the two-dimensional (2-d) mineral prediction and classification model was performed using data from the Songtao-Huayuan sedimentary manganese deposit. The authors investigated the coupling correlation between the

收稿日期:2019-04-20; **修订日期:**2019-07-23

资助项目:国家深地资源勘探开采专项《深部资源预测系统技术与示范》(编号:2017YFC0601502)、自然资源部公益性行业专项经费项目《地质大数据技术研究与应用试点》(编号:201511079-02)、重庆市社会事业与民生保障科技创新专项《富水断裂裂缝系统分布综合预测新技术研究与应用》(编号:cstc2017shmsA90003)和中国地质调查局项目《资源环境重大问题综合区划与开发保护策略研究》(编号:DD20190463)

作者简介:李诗(1994-),女,博士,从事地质大数据与矿产预测方面的研究。E-mail: 1176629242@qq.com

通讯作者:陈建平(1959-),男,博士,教授,从事矿产资源勘查、遥感与GIS应用、月球构造研究、地质大数据等领域的教学及科研工作。E-mail: 3s@cugb.edu.cn

spatial distribution of manganese element, sedimentary facies, outcrop of Datangpo Formation, faults, water system and the areas where manganese orebodies are present, as well as the correlation between different ore-controlling factors by employing the AlexNet networks. After training, the deep convolutional neural network classification model with the verification accuracy of 88.89%, recall of 66.67% and loss value of 0.08 could be obtained. By applying this model to unknown areas for two-dimensional metallogenic prediction, four metallogenic prospective areas. i.e., No. 91, No. 96, No. 154 and No. 184, were delineated, in which the ore potential probability of No. 91 regional ore-bearing probability and No. 154 prospective area is 1, and that of No. 96 is 0.5, suggesting that the probability of existence of undiscovered deposits in prediction areas is large.

Key words: big data, prospecting prediction, convolutional neural network, Alexnet network; Songtao - Huayuan Mn deposits

近年来,大数据已经成为第四科学范式,在大数据思维下矿产资源定量预测取得了许多积极的进展。赵鹏大^[1]提出,将大数据概念引入地学领域,提出数字找矿的概念,实现了数学地质到数字地质的飞跃,弥补了传统定性找矿的缺陷,结合四大找矿预测理论,对矿产预测做出科学的定量评价与分析,是未来的必然发展趋势。肖克炎等^[2]以大数据时代的预测思维方法,结合重要矿产资源潜力评价具体工作,探索了矿产资源预测评价的基本理论基础,总结了在数字化、信息化时代矿产资源预测评价的主要工作流程。陈建平^[3]重点介绍大数据背景下的“地质云”构建理念与方法,以及大数据在地学领域的应用。于萍萍等^[4]提出基于地质大数据理念的模型驱动的矿产资源定量预测评价新方法和基于模型流程建模技术的新思路。郑啸等^[5]介绍基于大数据的国家地质信息服务系统助推地质信息共享,资料数据服务与信息知识服务并重,并服务于国家建设与社会经济发展的问题。

在过去的数十年间,矿产资源评价领域涌现了大量的算法模型,在地质大数据时代背景下,机器学习与矿产资源预测的结合日益紧密。机器学习作为人工智能的重要分支,不仅为处理大量与矿产预测相关的证据特征图层、大数据分析、模式识别和预测提供了有效手段,也为基于大数据的智能找矿模型构建提供了技术支撑。机器学习具有识别已知矿床与证据图层之间非线性关系的潜力,其重要特点是赋予研究人员强大的预测能力。在矿产预测评价中运用最广泛的算法包括人工神经网络(Artificial neural networks, ANNs)、回归树(Regression trees, RTs)、支持向量机(Support vector machines, SVMs)、随机森林(Random forest, RF)。Zuo 等^[6]对福建省西南部某区域河流沉积物地球化学调查的 39 个地球化学变量数据,通过深度自编码网络成功地识别出与铁多金属成矿相关的地球化

学异常。有学者也成功地将基于大数据的机器学习应用于异常信息提取,从而深入地进行矿产资源定量预测^[7-14]。由此可见,基于多元、海量、异构的地质大数据,满足“两条主线”,即需求主线和“数据链”主线^[15]的人工智能矿产定量预测,是当今矿产资源定量预测发展的必然趋势。

然而,随着数据量不断增长,数据模式趋向于复杂化,数据之间的关联性逐渐增加,给分类、预测工作带来了更高的挑战,导致传统浅层机器学习算法难以表现出良好的性能。在此情况下,提出了深度学习法。作为一种神经网络结构,深度学习法与人脑模式更加相似,主要以传统浅层结构为基础,并适度加入隐藏层数提高分析性能。在地球化学处理中,国内已经多次使用神经网络方法^[16]。陈三明利用证据权重法与弹性 BP 神经网络实现了矿田与外围成矿预测。陈剑平^[17]尝试通过神经网络模式识别非线性方法,由其他层面将地质、地球物理等信息集中于神经网络模型内,开展油气藏综合预测工作。杨浩^[18]通过深度学习与主成分分析融合的技术对攀枝花某矿区的土壤地球化学测量数据进行分析处理,并成功地构建出体模型,证明改进后深度学习法能够提高地球化学分析准确度与智能化水平。Muhittin 等^[19]将深度学习中的细胞神经网络(Cellular Neural Networks, CNN)算法用在重、磁异常分离方面,利用模型试算与资料分析得出神经网络方法可以突出浅层异常的结论。目前该算法已成功应用于对铬铁矿区的矿床,以及围岩的重力异常进行分异等^[20-21],为矿产预测提供了更合理的依据。刘艳鹏等^[22]以安徽省兆吉口铅锌矿床为例,利用卷积神经网络算法,不断挖掘元素 Pb 分布特征与矿床地下就位空间的耦合相关性,最终圈定一个未知区域的成矿远景区。由此可见,卷积神经网络在对化学元素的空间特征提取与找矿预测方面具有一定的优势。

Krizhevsky 等^[23]提出 AlexNet 网络算法,以远高

Krizhevsky等^[23]提出 AlexNet 网络算法,以远高于其他算法的 TOP-5(即前5次判断正确)错误率 15.3%实现了深度学习在图像识别方面的突破。由于地质大数据的数量庞大,但目前尚无一种既能提取空间分布特征,又能挖掘不同成矿条件下多种控矿要素图层关联性的方法。本文在前人研究的基础上,采用深度卷积神经网络模型 AlexNet,通过训练松桃—花垣地区已知锰矿的分布与 Mn 元素、沉积相、大塘坡组出露、断裂、水系之间的关系,得到基于深度卷积神经网络的分类模型。应用该模型对松桃—花垣地区的未知区进行二维预测,弥补了以往矿产预测方法仅侧重于异常的定量提取,而忽视空间分布特征及不同成矿条件关联性的缺陷。

1 地质背景

研究区位于湘黔两省毗邻区(图1),是中国最重要的锰矿资源聚集区,近年来锰矿找矿突破进展重大^[24]。松桃—花垣地区属于上扬子古陆块的南部被动边缘褶皱冲带北东缘,跨铜仁逆冲带、风冈滑脱褶皱带2个四级构造单元。本区构造线主要呈北北东及北东向分布,少数呈南北向,褶皱形态多舒缓开阔,断裂多为与地层走向一致的正断层或逆断层。该区出露的地层为蓟县系梵净山群一下志留统、中上志留统、上古生界,缺失中生界三叠系、侏罗系、下白垩统及新生界。该区锰矿主要为新元古代晚期沉积型锰矿,主要以菱锰矿形式赋存于南华系大塘坡组第一段黑色页岩底部。关于研究区南

华纪大塘坡早期的古地理环境及其控矿作用研究,前人也做了较多的工作,一方面通过沉积相研究,分析岩相组合及横向变化规律,恢复大塘坡早期的沉积环境和古地理格局;另一方面利用岩石地球化学方法分析大塘坡早期的海洋古环境条件,包括古气候、古温度、古盐度、氧化还原条件、生物有机质作用、海底热水活动等^[25-34]。近年来,周琦等^[24]认为,“大塘坡式”锰矿是由于 Rodinia 超大陆裂解导致南华裂谷盆地形成并演化,于湘黔渝毗邻地区形成一系列“堑-垒”式古地理格局,且受控于一系列同沉积断裂,而锰矿是在这一系列的次级地堑盆地中发生古天然气渗漏沉积成矿的产物;同时指出受南华纪冰期—间冰期气候影响;同沉积断裂不但是深部锰质和古天然气上升的通道,更是连接古天然气渗漏沉积型锰矿成矿系统中内系统和表层系统的纽带。

根据前人的研究成果,可以得到如下认识:①大塘坡早期的古地理格局控制锰质的迁移富集;②由海底同生断裂及海底火山喷发导致的热水活动,可能提供了大量的成矿物质;③生物有机质在促进锰质的沉淀成矿方面具有重要的作用;④海洋古环境中的各种物化条件,特别是氧化还原条件,是影响锰质沉淀及沉淀形式的重要因素。通过收集研究区大塘坡期的岩相古地理图、建造构造图、锰矿床分布图、Mn 地球化学元素分布图等,构建本次矿产预测的基础图件。

2 研究方法

本文提出的基于深度卷积神经网络的 AlexNet

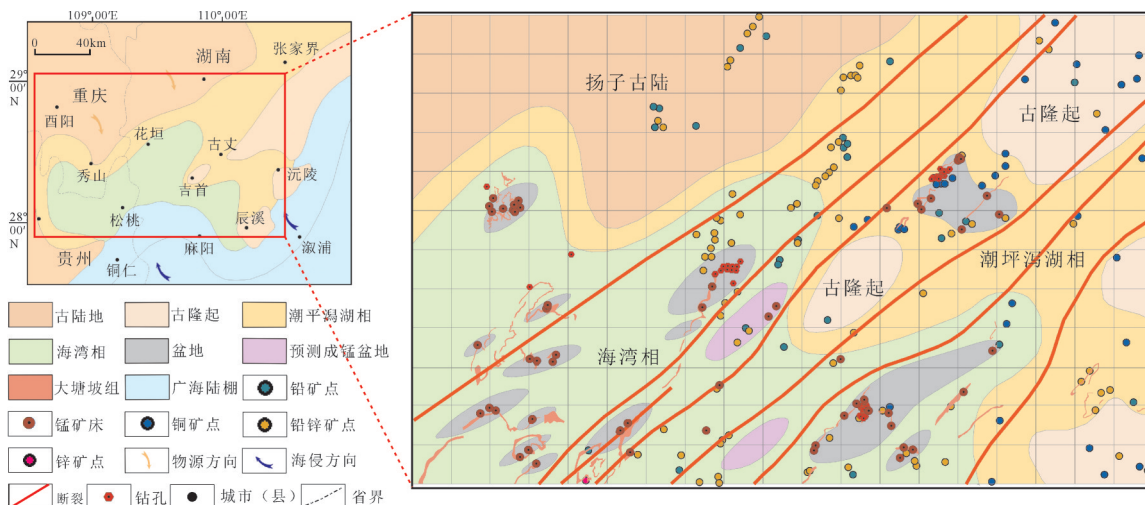


图1 松桃—花垣地区南华纪大塘坡期岩相古地理图

Fig. 1 Lithofacies-palaeogeographic map of Songtao - Huayuan area during the Datangpo Stage of Nanhua Period

模型的二维成矿预测整体框架如图 2 所示,主要包括 3 个主要组成部分。

(1)数据发现与预处理。大数据发现是实现数据化到信息化的必要技术手段,也是实现大数据挖掘找矿的前提。由于地质大数据具有数据量大、种类繁多、更新速度快等特点,需要通过需求分析建立逻辑结构树和网址结构树,并以此为依托开展数据收集工作。收集的数据包括文本数据与二维数据,但二维数据的收集应以基于文本挖掘的矿化模型构建为前提,在矿化模型的指导下进行二维数据的发现。由于收集到的二维数据具有格式不统一、坐标不一致等问题,因此需要将数据进行预处理。将地球化学元素含量进行 IDW 插值并转化为 TIFF 图像。将其他二维数据进行坐标统一与配准并进行格式转换,再将数据

进行切分。并将已知的正负样本作为训练集,未知待预测区作为测试集。

(2)AlexNet 模型预测。目前 AlexNet 网络结构已成功应用于图像处理中,但很少有研究者将其用于地质图像。本次研究将预处理之后的正负样本中的 80%作为训练集,20%作为验证集。将不同的要素图层以不同波段的形式输入到网络中。利用该网络结构优异的空间特征提取能力,将不同要素图层的空间分布特征用于训练预测模型。模型中同时涵盖了不同要素之间的潜在关联性。采用准确率和召回率两项指标对 AlexNet 模型的分精度进行检验。测试集经过 AlexNet 网络、flatten 层及全连接层,最终得到每个测试样本的 2 类分值,即含矿的得分与不含矿的得分。

(3)预测结果及含矿概率输出。将全连接层输出的得分经过 softmax 层的归一化,最终得到 0~1 之间的分值,即为含矿的概率与不含矿的概率,最终以 csv 格式的列表进行输出。通过已知正负样本的序号列表与预测结果的序号列表,在划分的网格中以不同的颜色代表不同的区域进行 JPG 格式图片输出。

通过上述技术方法流程,能更加智能化地处理多元、海量的二维要素图层数据,并以其中的空间分布特征及潜在关联性训练分类模型,达到找矿预测的目的,实现定位及定概率的二维预测过程,为基于地质大数据的找矿预测提供了新的思路。

2.1 数据发现与预处理

数据的发现从局域网和公网 2 个途径分别进行语料的获取。主要是包括文本数据与要素图层数据的收集。

针对文本数据发现的局域网数据的搜索与筛选是在 C#平台的基础上,对 Everything 软件进行二次开发,同时结合 MySQL 关系型数据库的方法实现。通过 P2P

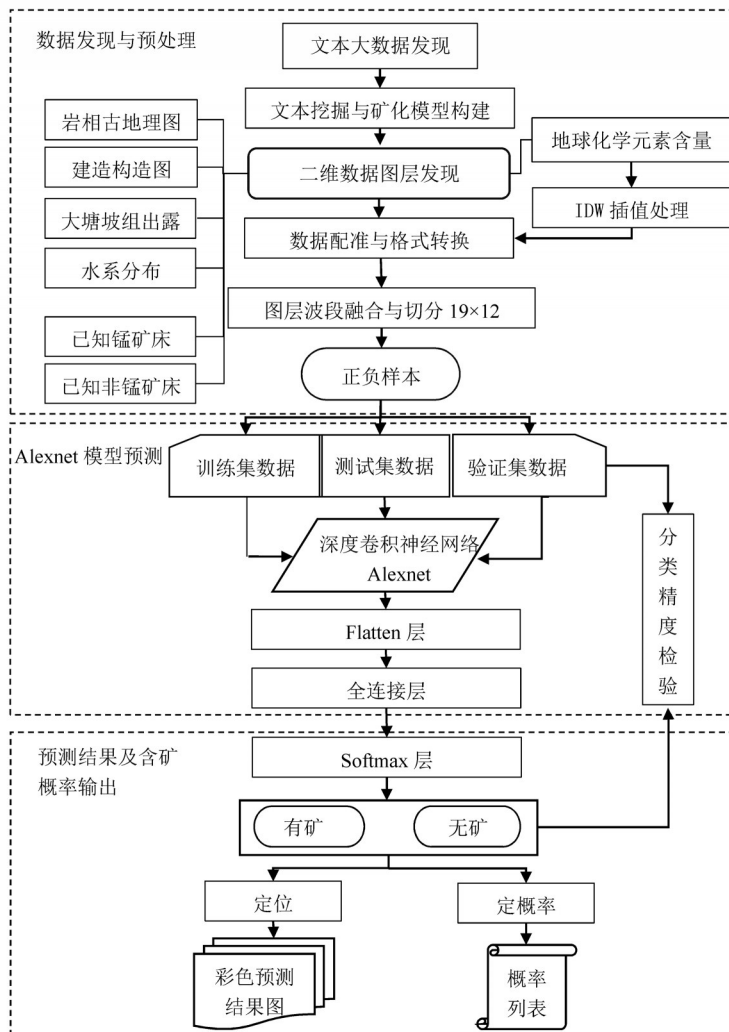


图 2 AlexNet 模型技术框架
Fig. 2 Technical framework of AlexNet

在线传输和FTP离线传输2种方法进行数据的获取。面对海量、多元、异构的公网数据,提出了基于检索关键词和网址URL的双迭代方案。在地质词典的关联词拓展结构网基础上建立逻辑结构树;通过百度、谷歌等主流搜索引擎对关键词的检索,生成初始URL种子网址,对种子网址的数据内容进行分析与信息提取,产生新的关键词,补充到结构树中。再利用关键词检索出新的网址,即通过网址链接种子与结构树的机器学习不断发现新的网址与关键词,形成网址结构树,二者相互迭代,形成正反两方向的全面搜索包围圈。本次研究收集的松桃—花垣地区的文本资料包括国内外及研究区的新闻、相关文献及区域报告。通过文本挖掘出的关键词与已有的矿床模型知识库,基于朴素贝叶斯概率进行匹配,最终得到松桃—花垣锰矿的矿化模型。由矿化模型结合研究区的实际二维图层数据发现结果形成该锰矿的找矿模型(表1)。

共收集到的找矿要素数据有39种地球化学元素含量统计表、水系分布图、岩相古地理图、建造构造图、大塘坡组出露图、已知锰矿床(点)分布图、已知铅锌矿床(点)分布图、钻孔数据分布图等。本次实验将39种地球化学元素进行IDW插值,并将其中相对重要的Mn元素分布图与水系分布图、大塘坡期岩相古地理图、断裂构造分布图、大塘坡组出露图进行坐标统一与配准,并统一转换为TIFF格式,如图3-a所示。将不同的要素图层作为每个样本区域图像的不同通道输入。并对松桃—花垣地区数据进行网格化,自西向东,自北向南分成228(19×12)个2×2km²的小格,编号依次为0~227(图3-b)。

2.2 AlexNet 网络结构

本文采用的AlexNet是2012年ImageNet竞赛冠军获得者Hinton和Alex Krizhevsky设计的。该网络模型前5层为卷积层,后3层为全连接层。AlexNet网络相对于传统神经网络的优势如下。

(1) 非线性 ReLU(Rectified Linear Unit) 激活函数

AlexNet采用ReLU激活函数,取代之前普遍采用的Sigmoid非线性激活函数,使整个过程的计算量节省很多,也可以使训练速度加快几倍。对于深层网络,Sigmoid函数反向传播时,很容易出现梯度消失的情况(在Sigmoid接近饱和区时,变换太缓慢,导数趋于0,这种情况会造成信息丢失),从而无法完成深层网络的训练。而ReLU会使一部分神经元的输出为0,这样就造成了网络的稀疏性,且减少了参数的相互依存关系,缓解了过拟合问题的发生。ReLU激活函数形式如下:

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

(2) Dropout层防止过拟合

AlexNet通过引入多种权值组合的DropOut方法控制过拟合。在训练过程中,隐层神经元的激活状态通过特定范数阈值进行控制,超过阈值的神经元在前向传播与反向传播中被抑制。本次训练时,用0.5的概率随机失活神经元。

本文采用的AlexNet模型各层超参数设置如图4-b所示。①第一层(卷积层):卷积核为[11,11], stride(步长)=4,输出通道=64, padding=valid(不扩充边缘)。②第二层(池化层):pool_size=[3,3], stride=

表1 松桃—花垣地区沉积型锰矿找矿模型

Table 1 Prospecting model for the sedimentary Mn deposits in Songtao - Huayuan area

预测要素	内容
成矿时代	南华纪大塘坡期
大地构造位置	张家界-花垣褶皱冲带
古地理	拉伸裂陷盆地
沉积相	半局限海湾含锰质页岩亚相
沉积序列	主要分布在南华纪第三层序(NHS ₃)和第四层序(NHS ₄)2个三级层序中,主要在海侵体系域(TST)和凝缩段(SS)2个体系域中
古气候	半局限海湾环境,气候转暖,为间冰期湿热气候,相应地出现消融性海侵
沉积建造	黑色页岩建造
构造	构造交会处,深大断裂和基底同生断裂
含锰岩系	黑色含粉砂质炭质页岩与条带状菱锰矿层组成,含锰岩系厚度大于35m时,锰矿体厚可达5~7m, Mn品位在24%以上(民乐);当含锰岩系厚度小于10m时,难以形成具规模的锰矿体(衫木寨)
岩层指示标志	厚层冰碛砾岩
地球化学	Mn元素异常值大于1200×10 ⁻⁶ 范围与矿化特征吻合

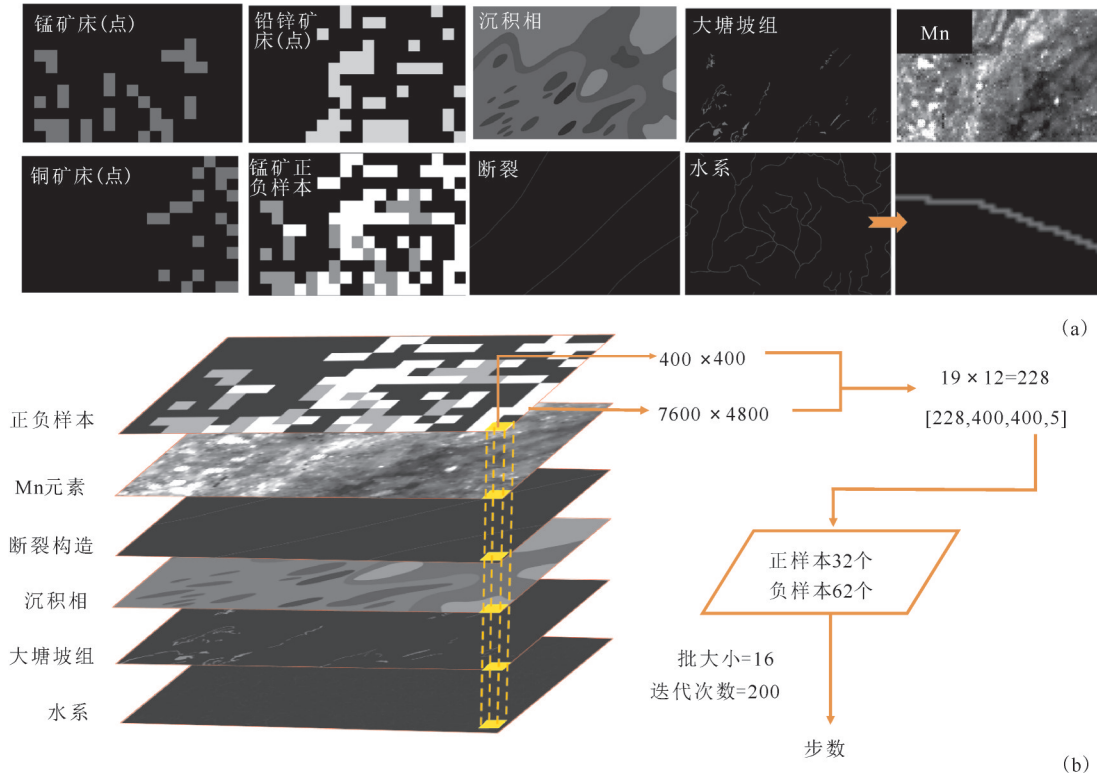


图 3 二维数据预处理

Fig. 3 Preprocessing of two-dimensional data

a—数据格式转换(其中正样本 32 个对应的网格即为已知锰矿矿床(点)所在网格区域(灰色),负样本 62 个即为其他矿种或钻孔未见矿区域的网格(白色),黑色为未知待预测网格);b—图像切分(图像共 7600×4800 像素,水平分辨率与垂直分辨率都为 96dpi,以 400×400 进行网格划分,共划分 228 个,其中 80% 为训练样本,每次随机取 16 个训练样本数据,循环次数为 200 次)

2, 输出通道=64(默认), padding=valid。③第三层(卷积层):卷积核为[5,5], stride=1,输出通道=192, padding=same(扩充边缘,使输入与输出大小相同)。④第四层(池化层):pool_size=[3,3], stride=2,输出通道=192, padding=valid。⑤第五层(卷积层):pool_size=[3,3], stride=1,输出通道=384, padding=same。⑥第六层(卷积层):卷积核为[3,3], stride=1,输出通道=384, padding=same。⑦第七层(卷积层):卷积核为[3,3],stride=1,输出通道=256, padding=same。⑧第八层(池化层):pool_size=[3,3],stride=2,输出通道=256, padding=valid。

每个卷积层包含了激活函数 Rectified Linear Unit(ReLU)处理,然后再经过降采样(pool 处理)。ReLU 作为 CNN 的激活函数,经过验证,其效果在较深的网络中超过了 Sigmoid,成功解决了 Sigmoid 在网络较深时的梯度弥散问题。后面的全连接层均进行了 dropout 防过拟合处理(除最后 FC8 网络输出层)。

本次研究用 3 个卷积层代替全连接层,可以提高 GPU 的运算效率。接着将 AlexNet 输出的结果经过 Flatten 操作后送入全连接层,最终通过 Softmax 操作输出每一个样本区域的含矿概率与不含矿概率(图 4-a)。

2.3 定位与定概率

本次通过获取预测结果的坐标将预测出的含矿区域与已知矿床(点)的区域赋予不同的颜色进行输出,将结果保存为 jpg 图像格式。同时,通过 Softmax 激活函数确定含矿概率。Softmax 函数计算如下:

$$P_k = \frac{\exp(S_k)}{\sum_{k=0}^k \exp(S_k)}$$

其中 S_k 是 Softmax 的输入,通常将 S_k 作为第 k 类的模型分数,也可以称为非归一化得分。 P_k 则是第 k 类的概率。

Softmax 激活函数只用于多于一个输出的神经元,保证所有的输出神经元之和为 1.0,所以输出的

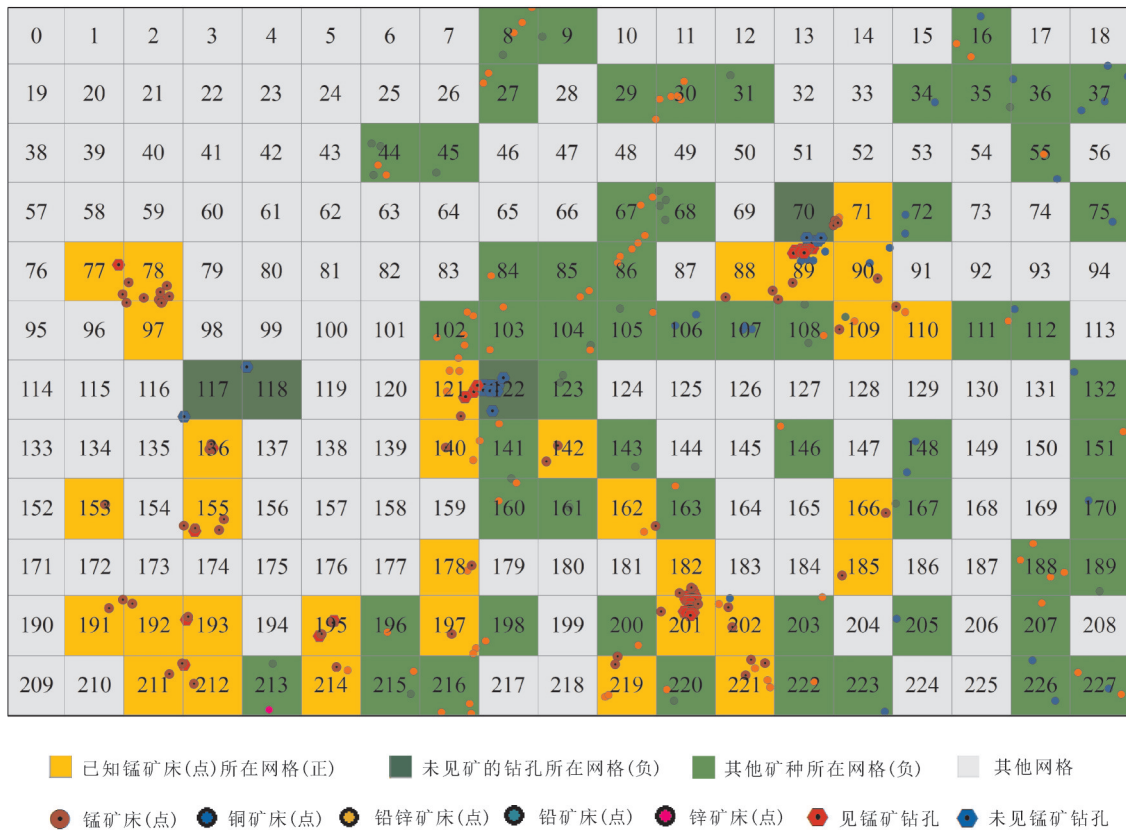


图 5 正负样本分布与区域编码

Fig. 5 Distribution and regional code of positive and negative samples

110、142、153、162、178、193、197、211 与 221 号区域为已知矿区,由于是验证样本的一部分,最终的预测结果为含矿概率 0.6~0.8,其余的验证样本均预测概率为 1,模型的验证准确率达 88.89%,召回率达 66.67%。其余的新预测高概率区还包括 79、94、96、137、169、172、189、203 及 204 号区域,绝大多数为有大塘坡组出露的区域,可见模型已将大塘坡组的空间分布特征提取出来并训练到模型中,并作为找矿预测的重要依据,进一步证实了该模型的可靠性。图 6-d 更加清楚地显示出新的预测区沿断裂走向分布的特征,且多处于 Mn 含量高值区。

基于 AlexNet 模型得分预测区域与 Softmax 函数归一化后的预测区域进行对比(表 2),发现 91、96、154 及 184 号四个区域不论何种方法都被确定为含矿概率较大的预测区,其中 91 号和 154 号的区域含矿概率为 1,96 号含矿概率为 0.5,表明该区有很大的可能含有锰矿。这 4 个预测区不但处于裂隙盆地内且都有大塘坡组出露的痕迹,这与松桃-花垣

大塘坡式沉积型锰矿的成矿类型相呼应。综合松桃-花垣地区的地质背景分析,认为该 4 个区域最有可能含有未发现的沉积型锰矿。

目前深度卷积神经网络模型 AlexNet 在实际的地质找矿预测应用中仍然面对一些问题。

(1) 针对研究区的尺度确定与网格划分大小的问题有待于进一步研究。若研究范围过大,则样本不够精确;若研究区范围过小,则样本数据范围局限,对于预测精度要求极大,地表的特征是否能反映出地下的成矿位置还未可知。网格划分的大小也间接影响成矿预测效果。若网格划分过细,虽然能最大程度地利用有限的样本,但样本数还是较少,且带来成倍的计算量,会大大降低模型效率;若网格划分过大,则不能有效利用已有的样本点,且预测的精度会降低。因此需要选择合适的研究尺度,在保证样本量的同时,也要尽可能地提高预测精度。

(2) 在样本的选取问题中,本次研究已经尽可能收集到已知的含锰矿或不含锰矿的样本点。将

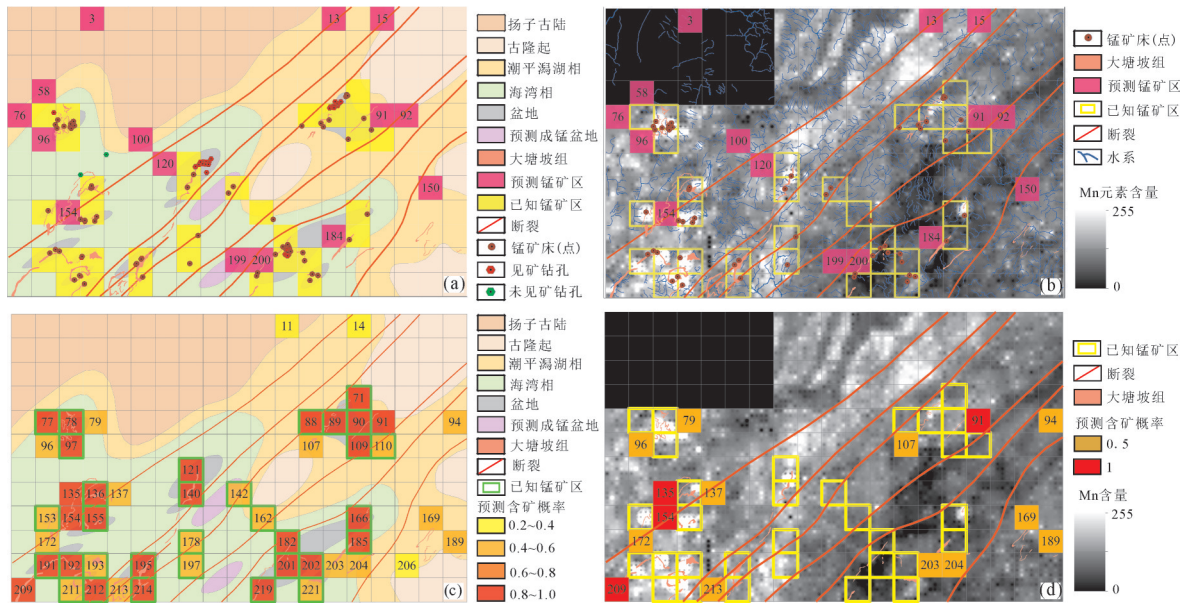


图6 预测结果与各个控矿要素图层叠加

Fig. 6 Superposition of prediction result and each ore-controlling element layer

a—模型直接打分预测结果与沉积相、断裂、大塘坡组出露叠加;b—模型直接打分预测结果与Mn含量分布与水系叠加;
c—Softmax 含矿概率预测法与沉积相、断裂叠加;d—Softmax 的预测结果与Mn含量分布叠加

已知的矿床(点)作为正样本(32个),将没有打到矿的钻孔(4个)和其他矿种(58个)当作负样本,但样本量对于深度学习是远远不够的。目前常用的通过图像的几何变换^[35]或使用极坐标空间中的径向变换^[36](radial transform)对原始图像进行扩充的方法并不适用于地质样本,因此急需找到合适的方法解决样本量的问题,这样训练出的模型效果会更好。

(3)在大数据的时代背景下,应当寻求能广泛适用于大量要素图层的模型。若样本量足够大,则所有在与矿床模型库匹配之后产生的有可能成为控矿要素对应的图层都可以对 AlexNet 模型进行训练,从而挖掘更潜在的控矿因素,真正做到让大数据“发声”,逐步实现“智能”找矿。

4 结 论

文本应用大数据挖掘下的预测思维,借助深度

卷积神经网络算法刻画复杂的、非线性的地学空间模式,提出了一种既能提取空间分布特征,又能够挖掘不同成矿条件下多种控矿要素图层关联性的二维找矿预测的方法。以松桃—花垣地区为研究区,得到以下结论。

(1)采用深度卷积神经网络模型 AlexNet,通过训练松桃—花垣地区已知锰矿的分布与 Mn 元素、沉积相、大塘坡组出露、断裂、水系之间的关系,得到基于深度卷积神经网络的分类模型。该模型可以有效地挖掘控矿要素地表分布特征与地下矿床就位空间的关系。

(2)经过训练后,该深度卷积神经网络分类模型的验证准确率为 88.89%,召回率为 66.67%,损失值为 0.08,可以认为该模型的精度是可靠的。

(3)应用该模型对未知区进行二维成矿预测,共圈定 91、96、154、184 号 4 个成矿远景区,其中 91

表 3 基于 AlexNet 模型得分预测结果与 Softmax 函数归一化后预测结果对比

Table 3 Contrast between the prediction result based on the score of AlexNet model and the prediction result normalized by Softmax function

AlexNet 模型得分	Softmax 计算含矿概率
第一次实验 3、58、92、150、184、199	含矿概率为 1 91、135、154、209 含矿概率>0.5 79、94、96、137、169、172、189、203、204
第二次实验 13、15、76、91、92、96、100、120、154、199、200	含矿概率>0 11、14、80、98、101、114、115、134、158、184、187、206

号和154号的区域含矿概率为1,96号含矿概率为0.5,认为预测区具有很大概率存在尚未发现的矿床。建议重新在该区选定地点布孔验证。

参考文献

- [1]赵鹏大. 大数据时代数字找矿与定量评价[J]. 地质通报, 2015, 34(7): 1255-1259.
- [2]肖克炎, 孙莉, 李楠, 等. 大数据思维下的矿产资源评价[J]. 地质通报, 2015, 34(7): 1266-1272.
- [3]陈建平, 李婧, 崔宁, 等. 大数据背景下地质云的构建与应用[J]. 地质通报, 2015, 34(7): 1260-1265.
- [4]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.
- [5]郑啸, 李景朝, 王翔, 等. 大数据背景下的国家地质信息服务系统建设[J]. 地质通报, 2015, 34(7): 1316-1322.
- [6]Zuo R G, Xiong Y H. Big data analytics of identifying geochemical anomalies supported by machine learning methods[J]. Natural Resources Research, 2018, 27(1): 5-13.
- [7]Twarakavi N K C, Misra D, Bandopadhyay S. Prediction of arsenic in bedrock derived stream sediments at a gold mine site under conditions of sparse data[J]. Natural Resources Research, 2006, 15(1): 15-26.
- [8]Chen M, Mao S W, Liu Y H. Big data: A survey[J]. Mobile Networks and Applications, 2014, 19(2): 171-209.
- [9]O'Brien J J, Spry P G, Nettleton D, et al. Using Random Forests to distinguish gahnite compositions as an exploration guide to Broken Hill-type Pb-Zn-Ag deposits in the Broken Hill domain, Australia[J]. Journal of Geochemical Exploration, 2015, 149: 74-86.
- [10]Gonbadi A M, Tabatabaei S H, Carranza E J M. Supervised geochemical anomaly detection by pattern recognition[J]. Journal of Geochemical Exploration, 2015, 157: 81-91.
- [11]Kirkwood C, Cave M, Beamish D, et al. A machine learning approach to geochemical mapping[J]. Journal of Geochemical Exploration, 2016, 167: 49-61.
- [12]Zhao J N, Chen S Y, Zuo R G. Identifying geochemical anomalies associated with Au-Cu mineralization using multifractal and artificial neural network models in the Ningqiang district, Shaanxi, China[J]. Journal of Geochemical Exploration, 2016, 164: 54-64.
- [13]Xiong Y H, Zuo R G. Recognition of geochemical anomalies using a deep autoencoder network[J]. Computers & Geosciences, 2016, 86: 75-82.
- [14]Chen Y L, Wu W. Application of one-class support vector machine to quickly identify multivariate anomalies from geochemical exploration data[J]. Geochemistry: Exploration, Environment, Analysis, 2017, 17(3): 231-238.
- [15]陈建平, 李婧, 谢帅, 等. 中国地质大数据研究现状[J]. 地质学刊, 2017, 41(3): 353-366.
- [16]陈三明. 锡矿山锡矿田多元地学综合信息成矿预测研究[D]. 中国地质大学(北京)博士学位论文, 2012: 1-290.
- [17]陈剑平. 基于MATLAB的神经网络模式识别技术在油气化探中的研究及应用[D]. 中国地质大学(北京)硕士学位论文, 2008: 1-49.
- [18]杨浩. 深度学习与主成分分析融合的研究与应用[D]. 成都理工大学硕士学位论文, 2016: 1-41.
- [19]Albora A M, Ucan O N, Ozmen A, et al. Separation of Bouguer anomaly map using cellular neural network[J]. Journal of Applied Geophysics, 2001, 46(2): 129-142.
- [20]刘展, 刘茂诚, 魏巍, 等. 基于细胞神经网络方法的重力异常分离[J]. 中国石油大学学报(自然科学版), 2010, 34(4): 57-61.
- [21]李超, 江玉乐, 胡明科, 等. 细胞神经网络在重力异常分异中的研究及应用[J]. 物探化探计算技术, 2015, 37(1): 16-21.
- [22]刘艳鹏, 朱立新, 周永章. 卷积神经网络及其在矿床找矿预测中的应用——以安徽省兆吉口铅锌矿床为例[J]. 岩石学报, 2018, 34(11): 3217-3224.
- [23]Krizhevsky A, Sutskever I, Hinton G E. Image Net classification with deep convolutional neural networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc., 2012: 1097-1105.
- [24]周琦, 杜远生, 袁良军, 等. 古天然气渗漏沉积型锰矿床找矿模型——以黔湘渝毗邻区南华纪“大塘坡式”锰矿为例[J]. 地质学报, 2017, 91(10): 2285-2298.
- [25]刘翼锋, 王庆生, 高兴基, 等. 贵州锰矿地质[M]. 贵阳: 贵州人民出版社, 1989: 1-191.
- [26]王砚耕, 王兴来, 朱顺才. 贵州东部大塘坡组地层沉积环境和成锰作用[M]. 贵阳: 贵州人民出版社, 1985.
- [27]王砚耕. 一个浅海裂谷盆地的古老热水沉积锰矿——以武陵山震旦纪锰矿为例[J]. 岩相古地理, 1990, (1): 38-45.
- [28]赵东旭. 震旦纪大塘坡期锰矿的内碎屑结构和重力流沉积[J]. 地质科学, 1990, (2): 149-157.
- [29]刘宝珊, 余光明, 陈成生. 雅鲁藏布江缝合带日喀则群的蛇绿岩质海底扇及其板块构造意义[J]. 岩相古地理, 1993, 13(2): 13-24.
- [30]何明华. 松桃及邻区早震旦世大塘坡早期岩相古地理及成锰条件[J]. 贵州地质, 1993, 10(1): 62-67.
- [31]何明华. 黔东北及邻区早震旦世成锰期岩相古地理及菱锰矿矿床[J]. 沉积与特提斯地质, 2001, 21(3): 39-47.
- [32]向文勤, 肖永开. 铜仁—松桃地区南华系大塘坡式锰矿地质特征及成矿规律探讨[J]. 西南科技大学学报, 2013, 28(4): 31-38.
- [33]牟军, 王安华, 黄道光. 贵州松桃—印江地区“含锰岩系”沉积微相特征与远景预测[J]. 贵州地质, 2014, 31(2): 99-104.
- [34]匡文龙, 李雪宇, 杨绍祥. 湘西北地区民乐式锰矿成矿地质特征及矿床成因[J]. 地质科学, 2014, 49(1): 305-323.
- [35]Hobert J P. The data augmentation algorithm: Theory and methodology[C]//Brooks S, Gelman A, Jones G, et al. Handbook of markov chain monte carlo. Boca Raton: CRC Press, 2011: 253-293.
- [36]Salehinejad H, Valae S, Dowdell T, et al. Image augmentation using radial transform for training deep neural networks[C]//IEEE. Proceedings of the 2018 IEEE international conference on acoustics, speech and signal processing. Piscataway: IEEE Press, 2018: 3016-3020.