

基于机器学习的三维矿产定量预测 ——以四川拉拉铜矿为例

向 杰¹, 陈建平², 肖克炎¹, 李 诗², 张志平³, 张 焯³

XIANG Jie¹, CHEN Jianping¹, XIAO Keyan¹, LI Shi², ZHANG Zhiping³, ZHANG Ye³

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

2. 中国地质大学(北京)地球科学与资源学院, 北京 100083;

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

2. School of Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China;

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

摘要:在大数据蓬勃发展的时代背景下,矿产资源定量预测作为地质大数据的核心部分,其综合分析挖掘多元信息的基本思路与大数据的理念不谋而合。以四川拉拉铜矿为例,开展基于机器学习的三维矿产资源定量预测。通过建立三维地质模型,提取成矿有利信息,构建研究区定量预测模型;基于“立方块预测模型”找矿方法,采用机器学习随机森林算法,计算出研究区成矿概率分布,以此圈定出 5 个找矿远景区。结果表明,随机森林具有更高的预测准确度与稳定性,且能够对控矿要素重要性做出定量评价。该研究成功地将机器学习应用于三维矿产定量预测,为今后的矿产资源预测评价做出了积极的探索。

关键词:矿产预测;三维建模;机器学习;随机森林算法;拉拉铜矿

中图分类号:P618.41;P612;P628 文献标志码:A 文章编号:1671-2552(2019)12-2010-12

Xiang J, Chen J P, Xiao K Y, Li S, Zhang Z P, Zhang Y. 3D metallogenic prediction based on machine learning: A case study of the Lala copper deposit in Sichuan Province. *Geological Bulletin of China*, 2019, 38(12):2010-2021

Abstract: Under the background of the vigorous development of big data, the quantitative prediction of mineral resources is the core part of geological big data. The basic idea of comprehensive analysis and mining of multi-information coincides with the concept of big data. With the Lala copper deposit as the study area, the authors carried out 3D mineral resources prediction based on machine learning. In this paper, 3D geological model was established to extract useful information of mineralization and build the quantitative prediction model of the study area. By using the "cube prediction model" prospecting method, the authors adopted the random forest algorithm of machine learning to calculate the probability distribution of mineralization in the study area. In this way, five prospecting prospective areas were delineated. The results show that the random forest has higher prediction accuracy and stability and can make quantitative evaluation on the importance of ore controlling factors. This study has successfully applied machine learning to the 3D mineral resources prediction and made a positive exploration for the prediction and evaluation of mineral resources in the future.

Key words: mineral prediction; 3D modelling; machine learning; random forest; Lala copper deposit

矿产资源预测评价是对一个国家、一个成矿区带或某个地区近地表和地下未发现矿产资源的矿种、类型、空间位置、数量和经济价值的估算与评价^①。21世纪以来,随着三维建模技术的日益发展,采用三维

收稿日期:2019-04-23;修订日期:2019-07-16

资助项目:国土资源部公益性行业专项《地质大数据技术研究与应用试点》(编号:201511079-02)、科技部深地资源勘查开采专项《深部成矿地质异常定量预测方法与模型》(编号:2017YFC0601502)、《深部矿产三维可视化预测评价软件系统研发》(编号:2017YFC0601501)和重庆科技创新专项《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)

作者简介:向杰(1990-),男,博士后,从事矿产资源定量预测与评价。E-mail: xiangjie@cugb.edu.cn

通讯作者:陈建平(1959-),男,博士,教授,从事矿产资源定量预测与评价。E-mail: 3s@cugb.edu.cn

立方体模型进行隐伏矿体(即第二找矿空间)的资源预测评价已经成为矿产勘查领域的热点。

赵鹏大等^[2]运用统计分析方法建立了矿床值与控矿因素之间定量联系的三维空间定位预测模型。陈建平等^[3-4]建立“立方块预测模型”找矿方法进行区域深部矿产资源三维预测评价。王功文等^[5-6]基于多源数据构建三维地质模型,应用证据权、分形等技术手段进行深部矿体预测和资源定量估计。肖克炎等^[7-9]构建了大比例尺三维预测工作流程,基于地质体三维模型和成矿预测理论建立立方体预测模型,开展三维矿体模拟与资源评估。毛先成等^[9-11]提出了“地质信息集成-成矿信息定量提取-立体定量预测”的深部矿产资源三维预测流程,研究了地质体三维形态定量分析、控矿地质因素定量提取及矿体立体定量预测方法。袁锋等^[12-13]运用三维综合信息成矿预测方法开展了中大比例尺的三维成矿预测实例研究。三维矿产定量预测的基本理论与技术框架已经基本成熟,并在深部矿产资源预测与评价工作中发挥了重要的作用。但三维建模技术仅作为一种数据集成手段来赋予地质要素空间属性,进行以数据驱动的矿产资源定量预测工作。从数学模型的角度,当前的三维预测还采用贝叶斯概率模型(如证据权法、找矿信息量法等)。许多研究表明,机器学习算法比贝叶斯概率算法更加准确,尤其是在特征空间很复杂的情况下,或输入的数据集具有不同的统计分布时,机器学习(Machine Learning)具有识别已知矿床与证据图层之间非线性关系的潜力^[14-18]。本次研究以四川拉拉铜矿为例,开展基于机器学习的三维矿产资源定量预测工作,为三维矿产资源评价的发展做出积极的探索。

1 地质矿产概况

1.1 区域地质背景

研究区位于扬子准地台康滇地轴中段。经历了漫长的地史演化:古元古代为弧后盆地环境,古一中元古代为拗拉槽环境,中一新元古代经历了由陆间裂谷至后造山裂谷环境,古生代一中生代历经了块断升降、地裂运动、前陆盆地、山间断陷等。复杂的构造环境形成了研究区良好的成矿背景,区内成矿元素丰富,是四川省重要的铜矿成矿带^[19]。

区域出露地层主要有前震旦系昆阳群、河口

群、三叠系、第四系,其中以古元古界河口群为主,中元古界昆阳群次之,再次为三叠系白果湾组及第四系。研究区在地史上主要经历了2期不同运动方式的构造活动。早期构造运动主要发生于会理运动造山期,是在南北向水平侧压力作用下形成的轴向近东西的河口复背斜,并伴随有超-基性岩的侵入和河口群通安地层的变质作用。晚期构造运动大致从晋宁期始,一直延续到喜马拉雅期,是在东西向水平侧压力作用下,利用和改造早期构造形成和发展起来的^[19-20]。研究区岩浆活动非常强烈,表现出多期次、多旋回的特点。其中,岩浆活动表现最强烈的是河口早期和中期,河口早期为强烈的火山喷发,河口中期则为基性岩侵入。虽有基性岩和酸性花岗岩的侵入,但海西期和印支期研究活动强度及规模较小^[21](图1)。

1.2 地质矿产特征

拉拉矿田地处于川滇有色金属成矿带北段,以“拉拉式”铜多金属矿产资源为主。拉拉铜矿田主要由5个矿区组成:落函铜矿区、落东铜矿区、天生坝铁矿区、红泥坡铜矿区及黎洪铁矿区,其中落函和红泥坡属于大型铜矿床。本次三维矿产定量预测研究区为最新发现的红泥坡矿区(图1-b)。由于缺少年代学数据,且康滇地区构造复杂,古一中元古代地层层序一直存在很大的争议,导致矿床成因仍然众说纷纭。如有顺层交代说、沉积-变质说、火山沉积-变质说、火山块状硫化物矿床、喷流沉积矿床等^[19-24]。本次通过详细的野外工作,并结合室内资料的整理与综合研究认为:矿床的形成主要经历了2个成矿期次,即火山喷发-沉积成矿期和变质热液改造成矿期。

红泥坡矿区圈定的5个矿体,厚度0.81~44.08m,南北长2900m、东西宽2600m,矿体埋深58~763m,分布标高范围1185~1825m。各个主要矿体一般分别赋存于一个较大的火山堆积旋回的中、上部,但个别矿体从中下部开始含矿,并延续到顶部。矿体的含矿类型可分为铜钴矿体、铜钼矿体、铜矿体和铁铜矿体。矿石以晶粒状变晶结构为主,按晶粒形态可分为自形粒状变晶结构、半自形粒状变晶结构、他形粒状变晶结构等。矿石的主要有用矿物有黄铜矿、斑铜矿、黄铁矿、磁铁矿、赤铁矿,主要脉石矿物有石英、钠长石、黑云母、白云母、方解石、白云石^[25]。

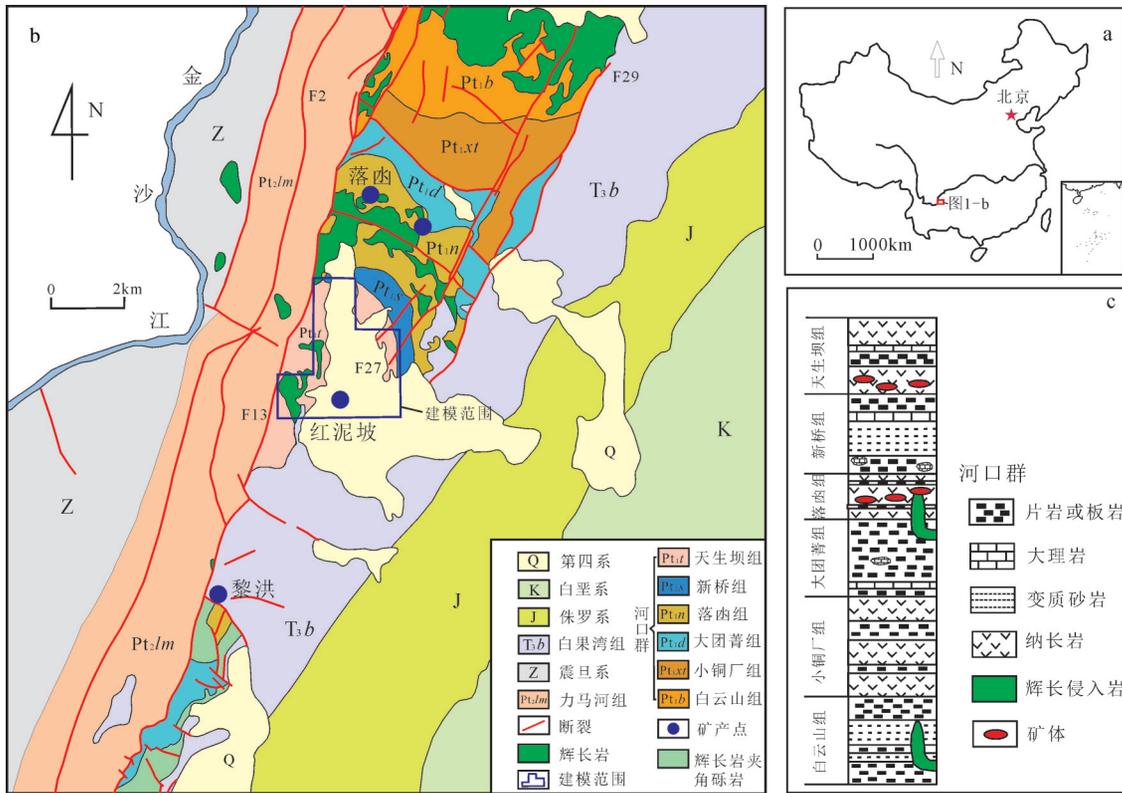


图1 拉拉铜矿田地质矿产简图

Fig. 1 Simplified geological map of the Lala orefield
a—研究区位置图; b—矿田地质图; c—河口群地层柱状图

2 机器学习矿产资源预测

机器学习的算法在矿产资源定量预测中的应用是当前数学地质与定量地学领域研究的热点, 因为其不仅能够处理大量的与矿产预测相关的证据特征图层, 还具有识别已知矿床与证据图层之间非线性关系的潜力^[14-15]。目前应用于矿产预测的机器学习算法主要包括人工神经网络、决策树、支持向量机及随机森林。对比研究表明, 随机森林相对于其他算法在不同训练样本下的预测能力具有更强的稳定性和鲁棒性, 此外, 随机森林能够对控矿要素的重要性进行排序, 对于训练研究区最佳的定量预测模型具有重要的意义^[14]。

2.1 基本流程

基于机器学习的三维矿产资源定量预测基本流程如图2所示。首先, 基于区内钻孔、剖面、地质图、物化探数据等建模数据, 以三维建模软件Surpac构建研究区的三维地质实体模型。然后, 将三维实

体模型立方块化, 通过与已知矿体进行统计分析, 提取成矿有利信息, 进而构建研究区的定量预测模型。最后, 采用机器学习算法——随机森林计算每个“立方块”的成矿概率, 通过含矿性分析确定找矿远景区的阈值并圈定找矿远景区, 并对找矿远景区做出评述。

2.2 随机森林算法

随机森林(Random Forest)是机器学习模型中一种重要的集成算法, 它是2001年Breiman在决策树^[26]和套袋法^[27]的基础上提出的一种由决策树分类器集合构成的组合分类器^[28]。随机森林基本思想是将多个没有关联的决策树合并在一起, 每棵树具有相同的分布。模型的分类误差取决于决策树的分类能力和树之间的相关性。对于分类算法, 新的输入样本根据森林中的决策树进行分类和预测。

2.2.1 决策树

决策树也是机器学习中的一种预测模型, 其特征着对象属性和对象值之间的映射关系。它利用

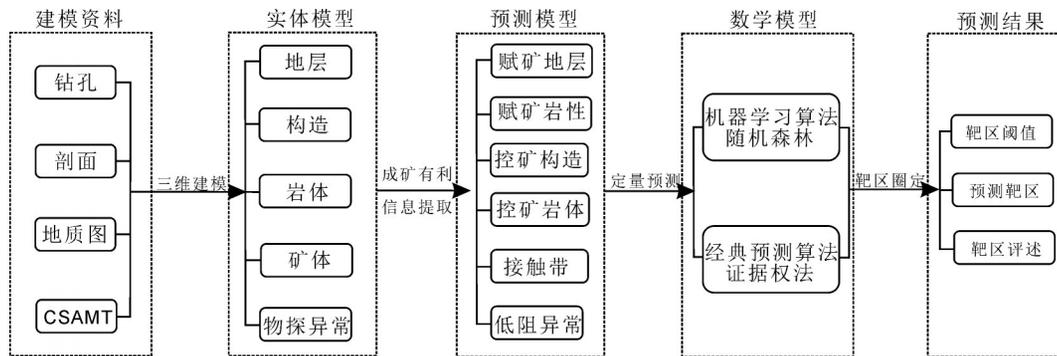


图2 机器学习三维定量预测流程图

Fig. 2 Flow chart of the 3D quantitative prediction based on machine learning

树的结构将数据记录分成3类节点:根节点、中间节点和叶节点,树的一个叶节点代表一定条件下的记录集对应着决策树的叶节点,树的分支根据记录字段属性建立;通过在分支子集中重复建立下层节点生成决策树。CART是一种典型的二叉决策树,既可以做分类也可以做回归。CART算法使用基尼系数(Gini Index)作为不纯度来衡量一个集合样本的杂质,对于一个特征变量为 m 的分类问题,其训练样本集 T 中有 n 个样本,则该训练样本集在的Gini系数为:

$$G(T)=1-\sum_{i=1}^n p_i^2 \quad (1)$$

式中 p_i 表示分类结果中第 i 个类别出现的条件概率。而按照某划分标准 S (如属性 A 的地 j 个属性值)将样本划分为 k 个子集,分别为 T_1, T_2, \dots, T_k ,则该划分的Gini系数为

$$G(T,S)=\sum_{i=1}^k \frac{n_i}{n} * G(T_i) \quad (2)$$

Gini系数越小,证明二分之一之后子样本的纯净度越高,该法分裂产生的决策树分枝越能代表不同类别直接的差异性,因为选择具有最小Gini系数的划分标准 S 作为最优分裂属性及最优分裂节点。

2.2.2 套袋法

袋装法采用等概率随机有放回地重复抽样技术选择训练数据构造分类器,最后进行组合,也称为自助法。随机有放回采样,即每次从原始训练集里随机采集固定个数的样本后,都将样本放回,又重新进行下一轮随机采样。由于袋外数据没有参与模型拟合,可以用来无偏估计分类的准确率,其实现过程:①每轮从原始样本集 D 中使用有放回随

机抽样的方法抽取 T 个训练样本,共进行 n 轮抽取,得到 n 个相互独立的训练集;②每次使用一个训练集 D_n 进行训练,得到第 n 个弱学习器;③针对分类问题可将上步得到的 n 个弱学习器采用多数投票的方式得到最终的强学习器;④针对回归问题,则将回归结果的算术平均值作为最终的强学习器。

随机森林应用于矿产预测,其特征变量即为定量预测模型中的控矿要素,利用其分类能力判断其含矿属性特征。其含矿属性主要分为3种情况:①钻孔控制的已知矿体;②钻孔控制的明确不含矿的块体;③未控制区域,不明确含矿与否。训练样本中,情况①为正样本,情况②为负样本,二者共同作为因变量的观察值;预测样本中,情况③作为因变量的观测值。本次研究基于KNIME构建随机森林的算法模型(图3),KNIME是基于Eclipse的开源数据挖掘软件,它通过工作流的方式完成数据仓库及数据挖掘中数据的抽取-转换-加载操作。

3 三维地质建模

三维地质建模是三维地质填图、深部地质调查、重点成矿区带大比例尺三维成矿预测等前沿性工作的重要基础,也是解决某些深部地质问题,乃至地质规律研究的重要途径。本次三维地质建模采用GEMCOM国际矿业软件公司的三维建模软件Surpac 6.3,基于平面地质图、钻孔数据、勘探线剖面、地球物理综合解译图等开展三维建模工作,建模数据详见表1。采用剖面建模法,构建了研究区的地层、岩体、断裂、低阻异常、矿体三维实体模型(图4)。

为了进行三维定量预测,需要将三维实体模型

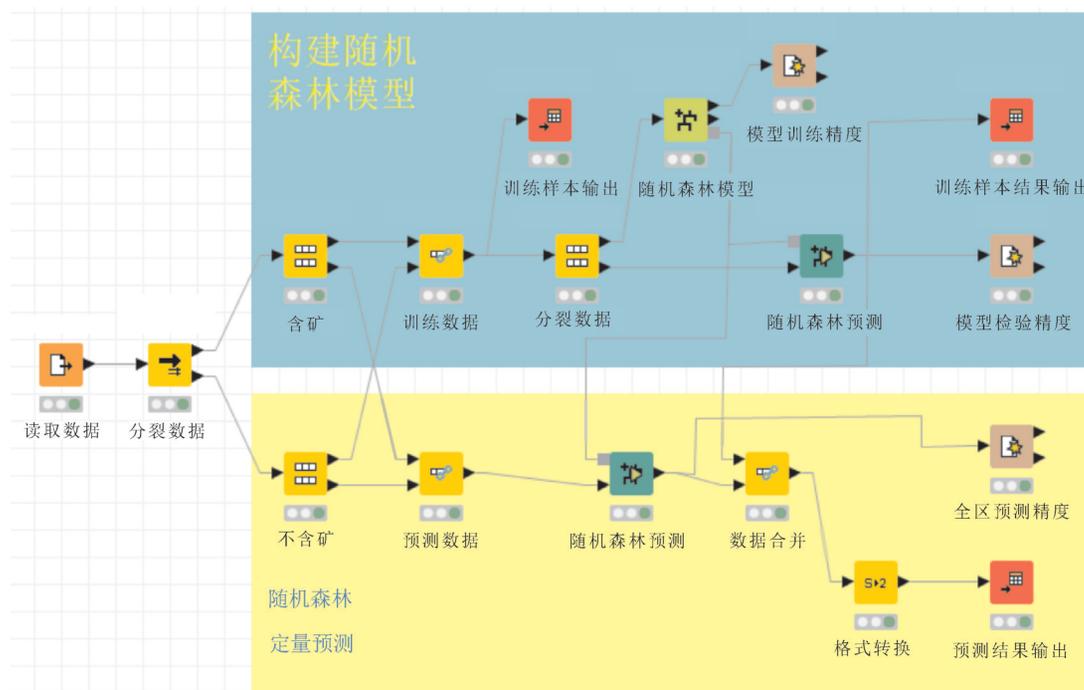


图3 基于KNIME的随机森林矿产资源预测模型构建

Fig. 3 Construction of random forest mineral resource prediction model based on the KNIME

“立方体化”。本次研究建模坐标范围南北 2899615~2903025m, 东西 792980~796250m, 海拔 1000~2050m。考虑到钻孔间距、地质剖面勘探线的勘探间距、对三维地质实体模型块体剖分后的数据量,以及计算机计算能力等,在保证块体单元的尺寸能够满足三维成矿预测计算精度的前提下尽可能提高计算机的运行速度,因此按照 25m×25m×25m 的块体单元尺寸对整个实体模型区进行块体划分,整个研究区共划分单元块体总数为 3099906 块。研究区存在大量的细脉矿体(图 4-d),采用上述单元块体尺寸剖分整个研究区时,只要包含了矿体

即赋值为矿体立方体,已知矿体共 14064 块。需要指出的是,采用该尺度剖分矿体用于储量估算极不准确,但是构建的块体模型是为了作为机器学习中正样本标记。

4 三维定量预测

4.1 成矿有利信息提取

在定量预测过程中,最终决定何种 GIS 图层参与模型运算的是预测模型。预测模型是在找矿模型的基础上,通过立方体(或者网格单元)进行赋值建立不同部位的特征变量,并结合已知矿体(或矿点)进行成矿有利条件的分析和提取,最终将各控矿要素及其成矿有利范围进行组合形成预测模型^[27]。本次从地层、构造、岩体、地球物理等方面开展成矿有利信息提取。

4.1.1 地层有利信息

本次主要对 5 个地层单元进行了建模,不同地层块体包含的矿体数目如图 5-a 所示。如果考虑不同地层分布范围的差别,可以用含矿率的概念对比分析。含矿率可以通过不同地层中含矿数目/地层块体数目得到(图 5-a),含矿率最高的为天生坝组

表 1 三维地质建模数据基础

Table 1 The database of 3D geological modelling

资料名称	比例尺	数量	探测深度/m
地形地质图	1:5000	1	地表
基岩地质图	1:5000	1	地表
矿层柱状对比图	1:5000	2	1000
钻孔柱状图	1:200	147	70~1030
地质勘探剖面	1:2000	21	1000~1200
CSAMT 综合解释剖面图	1:1 万	12	2000
CSAMT 综合解释中段图	1:1 万	3	500, 1000, 1500

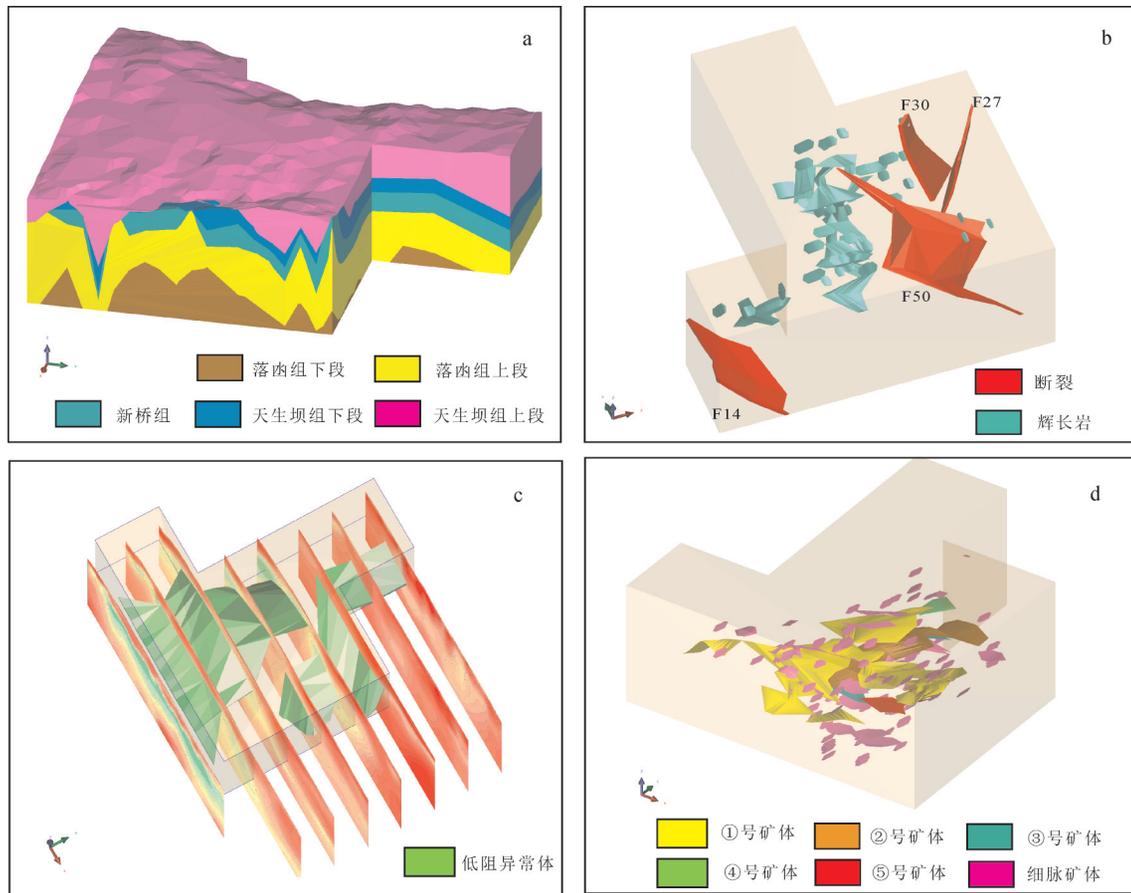


图 4 拉拉铜矿三维地质实体模型

Fig. 4 3D geological entity model of the Lala copper deposit
a—地层模型;b—断裂岩体模型;c—低阻异常模型;d—矿体模型

下段(Pt_{t1}),其次为落鹵组上段(Pt_{n2})、新桥组($Pt_{t,s}$)、天生坝组上段(Pt_{t2})、落鹵组下段(Pt_{n1})。通过以上 2 个方面的统计分析可以发现,红泥坡矿区的主要赋矿地层为天生坝组下段和落鹵组上段。

4.1.2 构造有利信息

红泥坡铜矿的形成与区域基底断裂有密切的联系,从成矿有利信息定量提取角度,研究区断裂块体数目为 41462,与矿体叠加的数目为 110 块,则断裂含矿率为 0.0025,达到了一定的含矿率水平,因此断裂为定量预测的控矿要素。此外,成矿运移通道往往通过断裂的缓冲区表征。本次通过“膨胀缓冲法”快速构建不同距离的断裂缓冲区^[8],进而通过含矿性分析确定最佳的断裂缓冲距离。“膨胀缓冲法”是采用形态学分析中的膨胀算法,将其应用于三维块体模型中。本次研究的立方块模型大小为 25m,每次缓冲一个立方格,即 25m 距离。通过 N 次

缓冲便可以得到 0~N*25m 的缓冲区,将 N-1 次缓冲得到的模型去除,即可达到不同等级的断裂缓冲区块体模型。本次研究中断裂缓冲区统计结果如图 5-b 所示,当断裂缓冲区为 100m 时为最佳的断裂缓冲距离。

除直接分析断裂及其缓冲区外,在定量预测中往往可以定量分析断裂的特征部位来挖掘有利的成矿信息。本次研究通过构造频数、构造等密度、方位异常度及中心对称度表征断裂构造的特征部位,董庆吉^[28]对具体算法做了详尽阐述,不在此赘述。本文采用 3DMP 软件系统实现空间计算:①构造频数,即截面网格中构造构造产出的条数,直接反映了区域构造的复杂程度,体现了区域构造格架的主体特征。与已知矿体进行叠加分析含矿率,进而确定有利的成矿区间为(0, 1.375),其空间分布如图 6-a 所示。②构造等密度,反映了线性构造的复杂

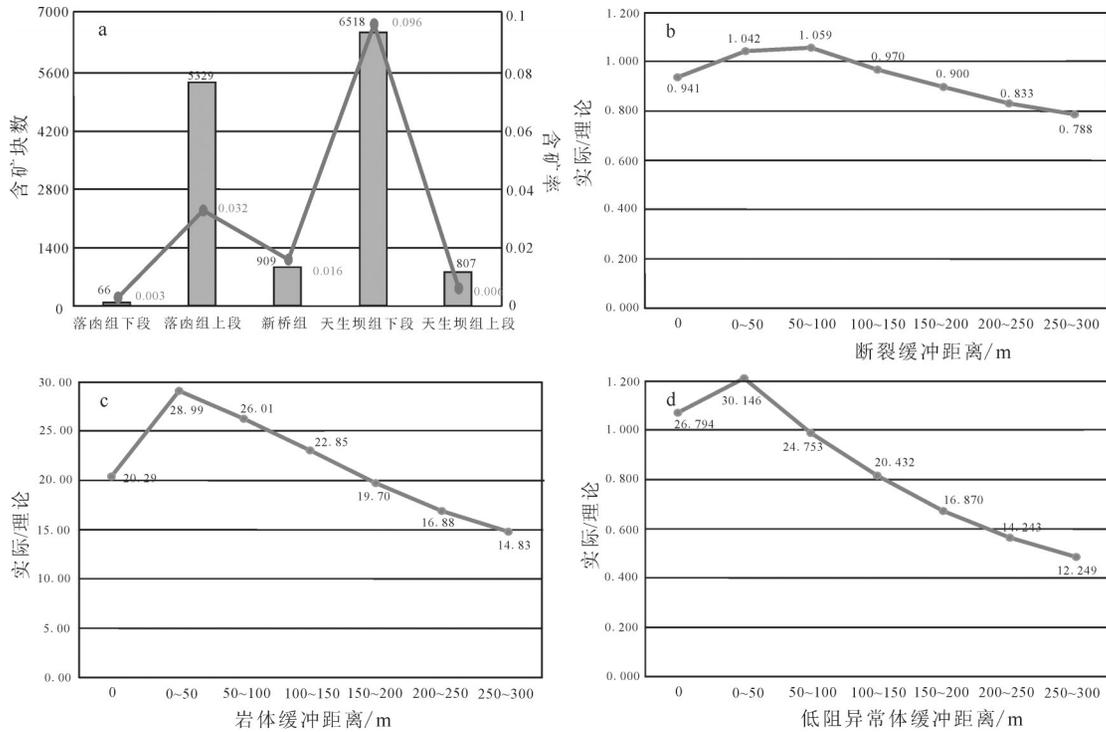


图5 成矿有利信息提取统计分析

Fig. 5 Statistical analyses of the useful information of mineralization

a—地层含矿性统计;b—断裂缓冲含矿率统计;c—岩体缓冲含矿率统计;d—CSAMT低阻异常缓冲含矿率统计

程度,其高值区展布一般与研究区构造在空间展布一致,体现了区域构造格架的主体特征。构造等密度与矿点的叠加分析,确定成矿有利区间为(1.045, 1.492),其空间分布如图6-b所示。③方位异常度,主要用来刻画区域构造的空间展布特征,定量表示出次级构造与主干构造的分布关系。本次研究中构造主干断裂方位为 $30^{\circ} \pm 10^{\circ}$ (图6-c),断裂缓冲距离设置为100m,确定其成矿有利区间为(0,0.1),空间分布特征见图6-d。④中心对称度,代表了构造对称的特征,主要用于揭示这些具有放射状断裂体系的环形构造,其高值区的分布一般与侵入岩体的分布相一致,如图6-e统计分析其有利成矿区间为(0.001,0.210),有利区间内的地质异常体分布见图6-f。

4.1.3 岩体有利信息

本次岩体三维地质建模中,只构建了辉长岩与辉长岩夹侵入角砾岩,在定量预测中将其合并统称为“岩体”。块体模型中岩体总块数为35812,其中含矿块数为584,含矿率为0.016,达到了一定的含矿率水平,因此岩体为定量预测的控矿要素。同样,本次研究通过“膨胀缓冲法”统计分析,当岩体缓冲

区范围为50m时,实际的缓冲区含矿率与理论含矿率的比值达到28.993,因此确定岩体缓冲距离为50m(图5-c)。

4.1.4 地球物理有利信息

本次研究采用的CSAMT方法在地层划分、断裂位置、岩体推断中起到了重要的作用。此外,通过将视电阻率剖面与已知矿体进行叠加分析,结合研究区的控矿环境及对各测线的解释推断结果,在12条综合解译剖面上圈定了12个中低阻异常区域,进而进行三维空间重构,形成矿化异常体。矿化异常体立方块总数为32033块,包含的矿块为1407块,含矿率为0.044。由于视电阻率是连续的,而中低阻的定义为相对区域,通过做缓冲区可以尽可能的减少信息的疏漏。采用“膨胀缓冲法”分析不同缓冲区情况下的含矿率情况,确定最佳缓冲距离为50m(图5-d)。

通过上述的成矿信息提取,结合研究区的区域矿产特征,形成了如表2所示的四川拉拉铜矿三维定量预测模型(表2)。

4.2 随机森林矿产预测

在矿产资源定量预测中,随机森林的应用主要

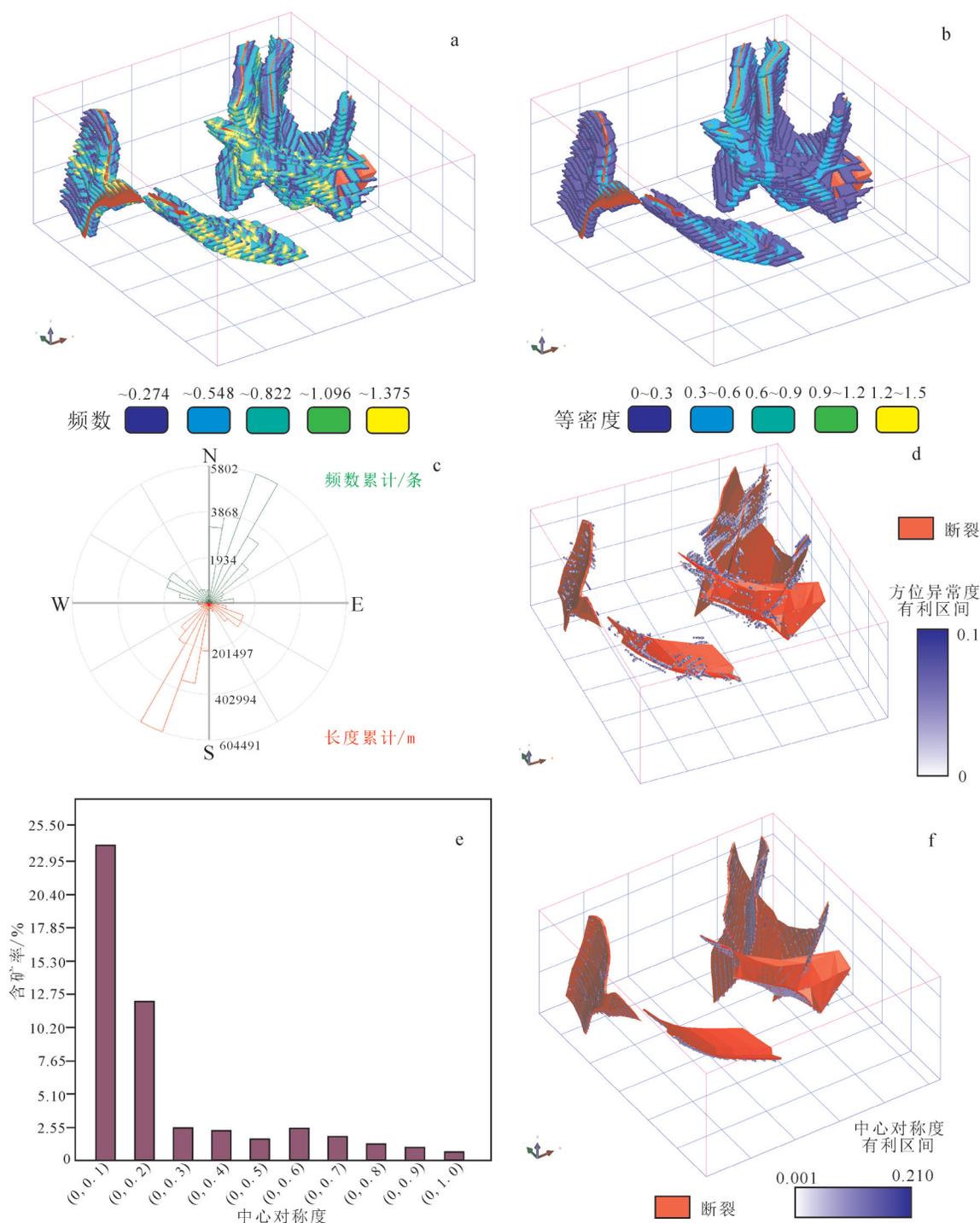


图 6 断裂构造特征定量分析与成矿有利信息分布

Fig. 6 Quantitative analyses of fault structural characteristics and the distribution of favorable mineralization information

a—频数; b—等密度; c—断裂玫瑰花图; d—方位异常度; e, f—中心对称度

分为 2 个方面:一方面利用其分类能力,预测找矿远景区,圈定成矿有利区;另一方面利用其对因子的重要性评价,对控矿要素进行重要性排序,优化找矿模型。再进行以上工作之前,对其精度进行评

价,确定模型的有效性^[29-30]。

4.2.1 精度检验

本次采用随机森林进行矿产预测是基于其分类能力,而对分类结果进行评价,确定分类结果的

表2 拉拉铜矿定量预测模型

Table 2 Quantitative prediction model of the Lala copper deposit

矿床类型	控矿要素	要素类型	特征变量	特征值	
火山沉积变质 型铜(铁)矿	地层	赋矿地层	天生坝组	天生坝组	
			新桥组	新桥组	
			落函组	落函组	
	构造	控矿断裂	断裂破碎带	基底断裂	断裂
				断裂缓冲区	100m缓冲区
		构造发育部位	方位异常度	(0,0.1)	
			构造等密度	(1.045,1.492)	
			中心对称度	(0.001,0.210)	
			构造频数	(0,1.375)	
	岩体	有利成矿岩体	辉长岩	辉长岩	
		岩体接触带	岩体缓冲区	缓冲区50m	
		有利岩体特征	岩体分异度	(0.07,3.04)	
	地球物理	视电阻率异常	CSAMT异常	低阻异常体	
			CSAMT异常缓冲	缓冲区50m	

精度和可靠性,常见的方法有ROC曲线。本次研究全区3099906个块体,选择14064个块体作为原始训练样本。研究表明,含矿单元和非含矿单元的比值为1:1,得到的随机森林训练模型最优。该训练样本集中随机选取70%作为训练集,剩下的30%作为测试集,用于构建随机森林模型。计算得到模型训练精度为96.7%,模型检验精度为96.5%(图7)。无论是模型训练精度还是模型检验精度都非常高,证明了RF模型构建的准确性,取得了令人满意的效果。

4.2.2 因子重要性分析

研究表明,组合分类器比单一分类器的分类效果好,随机森林是一种利用多个分类树对数据进行

判别与分类的方法,它在对数据进行分类的同时,还可以给出各个变量(基因)的重要性评分,评估各个变量在分类中所起的作用。随机森林算法计算每个特征变量在每棵数上所做的贡献,通过求贡献评价的方式定量评价特别变量的重要性。因此,在矿产资源定量预测过程中,可以利用该特性对控矿要素进行评价,对各控矿要素进行重要性排序,指导形成最佳定量预测模型^[31]。

本次研究通过设置不同的训练样本比例获取多组测试数据,从而多次评价控矿要素的得分,最后综合多次得分,对控矿要素进行排序。测试1是将训练样本中训练集设置为20%,测试2是将训练样本训练集设置为30%,以此类推,测试8的训练样

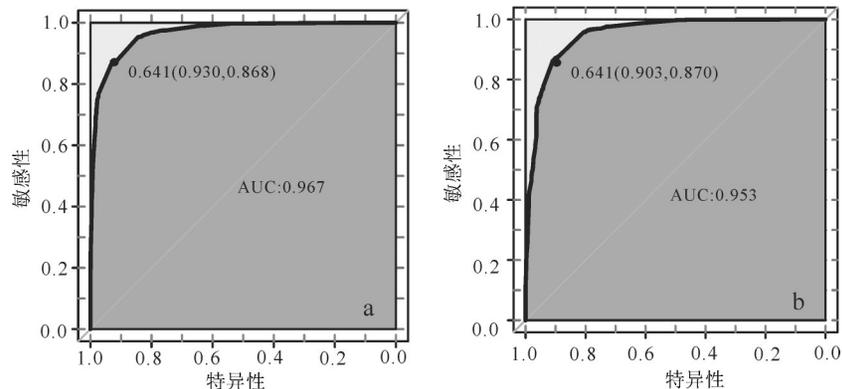


图7 随机森林定量预测精度检验

Fig. 7 Precision test of random forest quantitative prediction

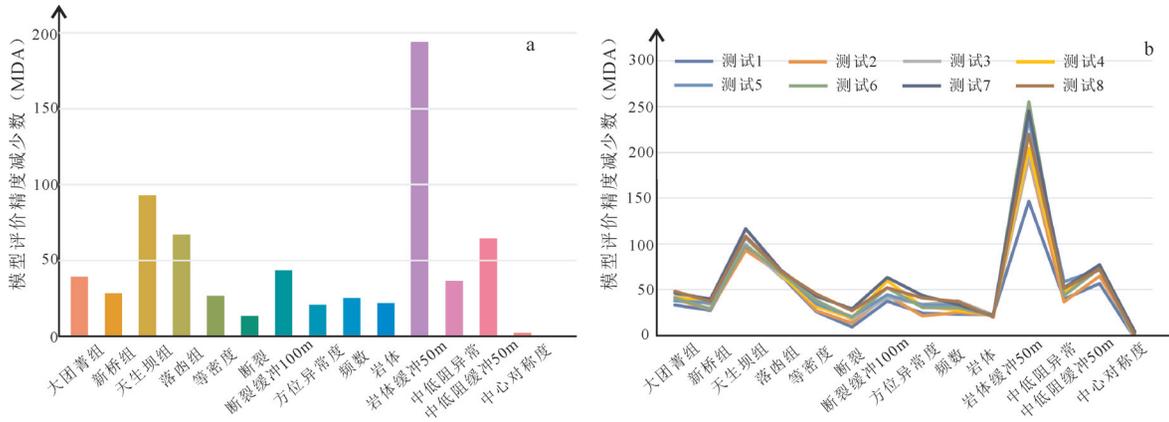


图8 拉拉铜矿定量预测模型控矿要素分析统计图

Fig. 8 Statistical chart of mineral control factors analyses of Lala quantitative prediction model

a—测试数据2;b—全部测试数据

本中训练集设置为90%。各控矿要素在不同的测试集中总体稳定性较好(图8),岩体缓冲区50m在众多控矿因子中重要性最强,天生坝组次之,最弱的为构造中心对称度。从图8可知,断裂及与之相关的等密度、方位异常度、频数、中心对称度都表现出较弱的重要性,证明该区断裂构造为次要控矿要素。此外,岩体虽然重要性不强,但是岩体缓冲区表现出强烈的控矿特征。从图4可以观察到,研究区辉长岩主要以岩脉产出,其与矿体模型直接重叠部分较少,而往往在矿体附近产出,这也印证了重要性排序结果,证明辉长岩是必要的控矿要素。中低阻异常及其缓冲区表现出与地层要素相近的重要性,其可以归为该区重要的控矿要素。

4.2.3 找矿远景区圈定

随机森林算法在矿产资源定量预测评价中的应用,除对控矿要素给出重要性评价外,还能够对其他块体属于“矿”还是“非矿”进行分类,并给出可能性概率(RF概率)。在本次研究中,由于矿产预测的特殊性(即定量预测并非对矿体直接进行预测,而是预测成矿的有利部位),不直接采信随机森林的分类结果,而是通过统计RF概率与所含矿体直接的关系确定阈值,进而确定成矿有利区(图9-a)。

图9-b为随机森林计算RF概率的含矿性统计图。矿数比为所含矿体数与总矿体数的比值,随着RF概率阈值不断增大,大于阈值的块体数目越来越少,所包含的矿体数目必然减少,因此矿数比不断减小。然而,含矿率不断增大,因为含矿表示的

是大于RF概率阈值的矿体数目与大于RF阈值的块体数,两者都在不断减小,但是其比值是不断增加的,表明RF概率与含矿性是正相关的。根据其统计规律,选定RF概率为0.95为成矿有利区的下限值,概率大于0.95的立方块体即为成矿有利区块(图9-c)。钻孔分布区域即为勘探程度较高区域,且控制已知矿体,其与成矿概率高值区分布重合性较好。结合矿区地质特征及已有的找矿事实,在研究区的深边部圈定了5个找矿远景区(图9-d),并根据其找矿概率划分为A级(3个)和B级(2个)。

5 结论

本次研究通过三维地质建模技术,构建了四川拉拉铜矿红泥坡矿田的三维地质实体模型,将实体模型“立方体化”之后,通过成矿信息提取,确定了该研究的三维定量预测模型。在定量预测模型的指导下,采用机器学习算法——随机森林对红泥坡矿区开展定量预测。

(1)基于机器学习算法开展三维定量预测,模型训练精度与模型检验精度均达到96%以上,证明了该方法的有效性。此外,在已知矿区的深边部圈定出找矿远景区5个,其中A级远景区3个,B级远景区2个。

(2)随机森林算法可计算控矿要素重要性,据此确定了辉长岩为必要的控矿要素,赋矿地层与中低阻异常为重要要素,断裂及其特征标志为次要要素。这对于进一步确定研究区的找矿标志,以及成

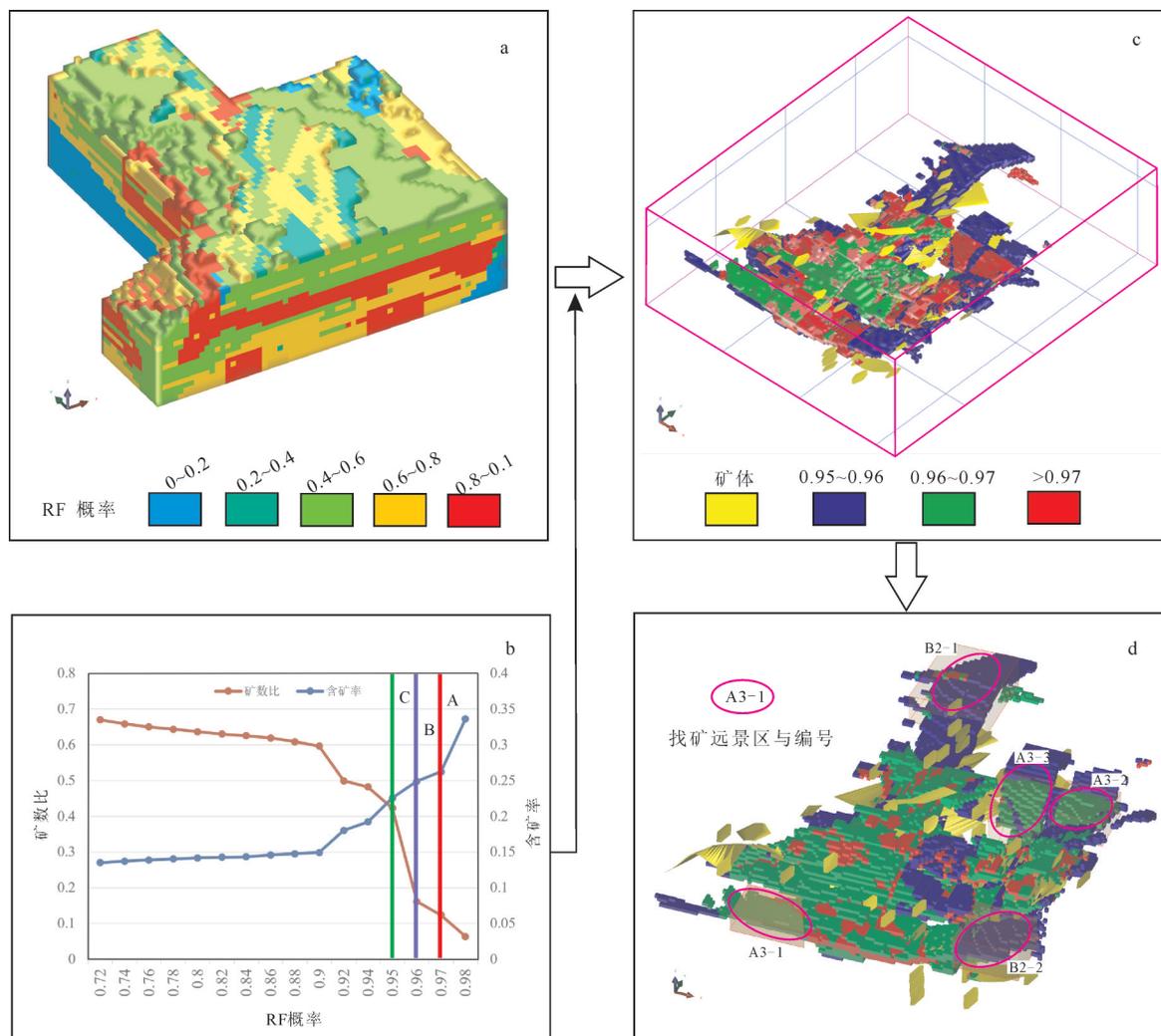


图9 随机森林圈定找矿远景区流程

Fig. 9 The process for delineating prospective area by random forest

a—全区RF概率分布;b—含矿率分析统计;c—成矿有利块体;d—找矿远景区分布

矿地质特征具有重要的意义。

(3)机器学习应用于三维矿产预测还处于起步阶段,合理构建找矿模型,通过机器学习的方式完成成矿信息的自动提取和找矿远景区的自动圈定,是未来的发展方向。

致谢:感谢中国地质调查局发展研究中心陈辉博士的指导,以及四川地质矿产勘查开发局403地质队龚灵明工程师在资料收集方面给予的帮助,同时感谢中国地质大学(北京)胡彬在三维建模中的帮助。

参考文献

[1]肖克炎,张晓华,李景朝,等.全国重要矿产总量预测方法[J].地学前缘,2007,14(5):20-26.

[2]赵鹏大,李紫金,胡光道.重点成矿区三维立体矿床统计预测——以安徽月山地区为例[M].武汉:中国地质大学出版社,1992.

[3]陈建平,吕鹏,吴文,等.基于三维可视化技术的隐伏矿体预测[J].地学前缘,2007,4(5):54-61.

[4]陈建平,于萍萍,史蕊,等.区域隐伏矿体三维定量预测评价方法研究[J].地学前缘,2014,21(5):211-220.

[5]Wang G W, Li R X, Emmanuel J M, et al. 3D geological modeling for prediction of subsurface Motargets in the Luanchuan district, China[J]. Ore Geology Reviews, 2015, 71: 592-610.

[6]Wang G W, Zhang S, Yan C, et al. Mineral potential targeting and resource assessment based on 3D geological modeling in Luanchuan region, China[J]. Computers & Geosciences, 2011, 37(12): 1976-1988.

[7]肖克炎,李楠,孙莉,等.基于三维信息技术大比例尺三维立体矿

- 产预测方法及途径[J]. 地质学刊, 2012, 36(3): 229-236.
- [8]Xiao K Y, Li N, Porwal A, et al. GIS-based 3D prospectivity mapping: A case study of Jiama copper-polymetallic deposit in Tibet, China [J]. *Ore Geology Reviews*, 2015, 71: 611-632.
- [9]毛先成, 戴塔根, 吴湘滨, 等. 危机矿山深边部隐伏矿体立体定量预测研究——以广西大厂锡多金属矿床为例[J]. *中国地质*, 2009, 36(2): 424-435.
- [10]毛先成, 邹艳红, 陈进, 等. 隐伏矿体三维可视化预测[M]. 长沙: 中南大学出版社, 2011.
- [11]Mao X C, Chen J, Deng H, et al. 3D Quantitative Predictivity of Concealed Ore Bodies in Fenghuangshan Copper Deposit, Tongling District, China[J]. *Acta Geologica Sinica*, 2015, 88(2s): 454-456.
- [12]袁峰, 李晓辉, 张明明, 等. 隐伏矿体三维综合信息成矿预测方法[J]. *地质学报*, 2014, 88(4): 630-643.
- [13]Li X H, Yuan F, Zhang M M, et al. Three-dimensional mineral prospectivity modeling for targeting of concealed mineralization within the Zhonggu iron orefield, Ningwu Basin, China [J]. *Ore Geology Reviews*, 2015, 71: 633-654.
- [14]Rodriguez G V, Sanchez C M, Chica O M, et al. Machine learning predictive models for mineral prospectively: An evaluation of neural networks, random forest, regression trees and support vector machines[J]. *Ore Geology Reviews*, 2015, 71: 804-818.
- [15]Zhang S, Xiao K, Carranza E J M, et al. Maximum entropy and random forest modeling of mineral potential: Analysis of gold prospectivity in the Hezuo-Meiwu district, west Qinling Orogen, China[J]. *Natural Resources Research*, 2019, 28(3):645-664.
- [16]Zuo R, Xiong Y. Big Data Analytics of Identifying Geochemical Anomalies Supported by Machine Learning Methods[J]. *Natural Resources Research*, 2017: 1-9.
- [17]O'Brien J J, Spry P G, Dan N, 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(149): 74-86.
- [18]Carranza E J M, Laborte A G. Random forest predictive modeling of mineral prospectivity with small number of prospects and data with missing values in Abra (Philippines) [J]. *Computers & Geosciences*, 2015, 74: 60-70.
- [19]周家云, 陈家彪, 沈冰, 等. 四川拉拉铜矿构造成矿动力学机制[J]. *大地构造与成矿学*, 2008, 32(1): 98-104.
- [20]孙燕, 李承德. 四川拉拉铜矿床成矿机制研究[J]. *成都地质学院学报*, 1990, 4: 1-9.
- [21]Zhou J Y, Zheng R C, Zhu Z M, et al. Geochemistry and Sm-Nd Dating of the Gabbro in the Lala Copper Ore District, Sichuan Province, China[J]. *Bulletin of Mineralogy Petrology & Geochemistry*, 2009, 28(2): 111-122.
- [22]申屠保涌. 四川会理拉拉厂铜矿床地质地球化学特征及成矿模式[J]. *沉积与特提斯地质*, 1997, (21): 115-130.
- [23]陈根文, 夏斌. 四川拉拉铜矿床成因研究[J]. *矿物岩石地球化学通报*, 2001, 20(1): 42-44.
- [24]朱志敏, 曾令熙, 周家云, 等. 四川拉拉铁氧化物铜金矿床 (IOCG)形成的矿相学证据[J]. *高校地质学报*, 2009, 15(4): 485-495.
- [25]张武鹏, 余臻, 李峰, 等. 四川拉拉会理红泥坡铜矿地质特征及找矿标志[J]. *有色金属工程*, 2016, 6(2): 80-84.
- [26]Breiman L. *Classification and Regression Trees*[M]. New York: Routledge, 1984.
- [27]Breiman L. Bagging predictors[J]. *Machine Learning*, 1996, 24: 123-140.
- [28]Breiman L. Random forests[J]. *Machine Learning*, 2001, 45: 5-32.
- [29]向杰, 陈建平, 胡桥, 等. 基于矿床成矿系列的三维成矿预测——以安徽铜陵矿集区为例[J]. *现代地质*, 2016, 30(1): 230-238.
- [30]董庆吉, 肖克炎, 陈建平, 等. 西南“三江”成矿带北段区域成矿断裂信息定量化分析[J]. *地质通报*, 2010, 29(10): 1479-1485.
- [31]向杰, 陈建平, 胡彬, 等. 基于三维地质-地球物理模型的三维成矿预测——以安徽铜陵矿集区为例[J]. *地球科学进展*, 2016, 31: 603-614.