

基于大数据挖掘的露天非金属矿产 资源定量预测模型构建

Construction of Quantitative Prediction Model for Open-pit Non-metallic Mineral Resources
Based on Big Data Mining

刘春生(中国建筑材料工业地质勘查中心黑龙江总队,黑龙江 哈尔滨 150040)

摘要:为实现露天非金属矿产资源定量预测,以地质理论为指导,在计算机技术基础上,将理论分析与大数据挖掘算法相结合,建立三维非金属矿产资源定量预测模型。收集地质资料,通过二项分布模型模拟资源分布状态,统计期望矿点数量及对应函数关系式,提取数据特征,完成数据关联分析,构建空间属性数据库,分幅平差处理数据,消除信息误差、减少数据丢失,基于训练样本数据计算定量预测模型先验概率,获得控矿要素及重要性评估指标,得到最终的定量预测结果。实验部分证明了所提模型预测精度高,三维可视化结果能够清晰呈现存在矿产资源的位置,可应用价值高。

关键词:地质大数据;大数据挖掘;非金属矿产;定量预测;预测模型构建

中图分类号: P628

文献标志码: A

文章编号: 1672-609X(2023)01-0029-06

Abstract: In order to realize the quantitative prediction of nonmetallic mineral resources in the open air, a three-dimensional quantitative prediction model of nonmetallic mineral resources is established by combining theoretical analysis with big data mining algorithm on the basis of computer technology under the guidance of geological theory. Collect geological data, simulate resource distribution state through binomial distribution model, count the expected number of ore spots and corresponding function relationship, extract data characteristics, complete data association analysis, build spatial attribute database, process data by amplitude adjustment, eliminate information error, reduce data loss, train sample data to calculate the prior probability of quantitative prediction model, and obtain ore control factors and importance evaluation indicators. The final quantitative prediction results are obtained. The experiment part proves that the proposed model has high prediction accuracy, and the 3D visualization results can clearly show the location of mineral resources, which is of high application value.

Key words: geological big data; big data mining; non metallic mineral resources; quantitative prediction; prediction model construction

1 前言

在矿产资源被不断发现和挖掘的过程中,更深层次的矿产资源寻找和勘探难度在不断增加,针对矿产的预测和分析技术也不断更新换代,在大数据背景下,更需要重视并有效地推动对矿产资源的寻找,目前最常用的方法是根据成矿规律进行定量预测的,但不能深度的预测出成矿的体量。

左仁广^[1]根据相关性理论,采集区域大数据信息,结合地质异常数据,完善数据种类多样化,通过

机器学习实现数据的特征提取和定量预测,但方法受环境影响过大,结果不确定性较强;马瑶等人^[2]通过提取和挖掘地质复杂数据,识别其中的矿化信息,融合异常数据并实现信息关联转化,在逻辑回归的基础上建立人工网络,完成定量预测和成矿发现,但方法存在一定的数据错分问题。

因此,本文在遵循成矿规律的基础上,结合了针对露天非金属矿产的大数据挖掘方法进行信息提取,从地质背景环境研究出发,在成矿规律中挖掘更丰富的内涵因素。通过成矿的规律和模式数据完成对应的定量预测模型构建,对露天非金属采矿区进行对应的地质、物探、化探等相关信息的采集挖掘,在实际勘探的基础上实现数据和地质的对应结合,主要解决地质理论中无法深入的实际问题,并利用

[作者简介] 刘春生(1981—),男,高级工程师,从事矿产地质勘探工作。

[引用格式] 刘春生. 基于大数据挖掘的露天非金属矿产资源定量预测模型构建[J]. 中国矿山工程,2023,52(1):29-34.

多个数学模型拓展找矿和成量预测结果,重视数据和实质,最终构架出依托于计算机技术的定量预测评估模型,系统、科学的完成相关的地质研究的预测问题。

2 露天非金属矿产数据采集

在露天非金属矿产资源的定量预测模和评估中,主要以地质理论为指导,以计算机技术为基础进行大数据挖掘,将理论分析与数据挖掘结合起来,对地质信息进行数据挖掘,建立三维实体地质模型,针对目标预测地带设定多源数据单元,实现数据准备和选择,根据地质图件的实际尺寸确定评价单元大小,将非金属矿产研究区域划分成等面积等形状的小区域,方便实现统一观测和取值,利用概率统计算法^[3],由于矿产资源分布状态表现为随机,所以建立二项分布模型为

$$P_n(x) = C_n^x p^x q^{n-x} \quad (1)$$

式中, $P_n(x)$ 表示 x 矿脉的矿产资源在 n 个网格下的分布状态; C_n^x 表示 x 矿脉在 n 个网格下矿点间隔; p^x 表示矿脉中包含矿产资源的实际体量; q^{n-x} 表示矿产资源预测实际体量^[4]。

统计被归纳在网格单元内的期望矿点数量,计算与期望矿点间的函数关系为

$$\sigma(x)/E(x) = \sqrt{(1-S)/(n \times S)} = 1/3 \quad (2)$$

式中, S 表示预测网格单元的实际面积; $\sigma(x)$ 表示网格单元内的预测矿点数量; $E(x)$ 表示实际落入网格单元内的矿点数量。

在完成网格划分的基础上,抽取和采集研究区域地质矿体的数据,将矿脉的地质信息通过地质、构造、矿产、物探、化探等不同的图件表示出来,并根据矿点、构造、重力空间数据等不同的信息要素,对划分出的各网格单元进行对应的信息匹配^[5]。

根据多边形要素可以对矿山所在地质环境中的多元化信息进行相应的信息熵计算,对数据进行分块化处理,单独分析各区属性。网格单元内地质环境的信息熵主要描述地质结构的相对变异程度,具体可表示为

$$E_{ij} = - \frac{\sum_{k=1}^n x_k \ln x_k}{\ln n} \quad (3)$$

式中, E_{ij} 表示在 i 行 E_{ij} 列对应的网格单元内矿山地质中的信息熵; x_k 表示网格单元中包含的第 k 类面

元地质体所占面积与总单元面积的比例。

针对矿脉的断裂、密度及构造等复杂情况,需要利用线要素完成数据抽取,通过GIS空间叠加法分析对应的单元构造复杂度,这有利于后续对矿体结构和量体数据的挖掘和预测,综合反映出研究区域成矿因素,具体的计算过程为

$$C_i = \sum_{k=1}^p q_{ik} \quad (4)$$

式中, C_i 表示 i 行网格单元对应的构造复杂度, p 表示构造线性条件的变量总数量, q_{ik} 表示 i 行网格单元中的实际变量取值。

完成相应的地质信息抽取后,受到信息类型、量纲不同等的限制,数据之间没有一定的衡量条件,因此原始数据无法满足数据模型的建立要求,此时需要通过数据转换将其呈现出正态分布状态,在此基础上统一其量纲和线性变换条件,通过观察值转换数据:

$$U = F(X) \quad (5)$$

其中, X 表示原始的地质采集观测数据, F 表示数据转换函数, U 表示经过转换后的新观测数据。统一数据量纲的过程中,进行了有效地数据归一化变换,使数据变化在 $0 \sim 1$,实现评价单元的观测现象的统一化预处理。最终输出的数据采集结果能完整的呈现地质环境描述,为后续基于大数据挖掘的预测提供图像和文字两种不同形式的信息基础。

3 大数据挖掘下露天非金属矿产资源定量预测模型

针对上述地质环境数据采集得到的结果,为使其在矿产预测工作中发挥更好的作用,研究中针对地质数据中包含的地质、物探、化探等相关信息进行综合挖掘,根据物探和化探数据结果,首先从中找到地质异常点,通过理论分析和实地勘探的结合,将各类型采集数据相融合,地质环境的航磁数据和化探数据进行格式统一和处理计算,完成数据的特征提取^[6]。

利用多源多类的数据进行关联分析,挖掘出这些数据之间的深层潜在相关性,分析地质要素之间的相互关系,矿产资源的空间属性及时间等特征。不同地质空间资料从不同角度反映地质目标的性质,而同一空间位置或呈现相同频次的地质数据,其空间属性之间存在一定的关联性,根据不同类型的

地质空间采集大数据,建立基于位置属性的数据强关联^[7]。从而,能够根据不同的空间资料,建立基于位置的数据间的强烈相关性。在同一坐标系下,对不同主题类型的空间数据进行整合,并对进行空间属性提取,构建空间属性数据库,关联度分析基于空间数据库,完成各项目集集合的关联性挖掘,统计不同地质现象、地质体、化学元素等地质体的发生频率,将频率最高的特征属性或超出某一数值的特征属性转化为相关性的判断规则。

在分析和研究露天非金属材料矿产资源的定量预测时,主要以化探数据作为大数据信息,结合地、物、遥相关数据,提高大数据采集和挖掘精度,同时在有限的原始数据基础上,对数据进行分幅平差处理^[8],消除原始数据中存在的信息误差,从中提取地质环境的低缓异常信息。

在实际的应用分形过程中,首先要确定各地质图幅和元素之间正负异常的上下限,针对性的补偿图幅之间的系统定量,在地质图幅的相邻边缘进行分幅平均值补偿计算,降低低缓假异常的产生和真异常的丢失情况,线性拟合方案在这个过程中起到异常重要的作用。

为保证地质环境的物探和化探能同时完成相关性的定量分析和处理,通常情况下将地质化探数据作为坐标中心点,以固定直径完成平差化处理^[9],计算求得航磁数据的最大最小值,分别用 (ΔTd) 和 (ΔTx) 表示,在此基础上考虑地质航磁对应的正负异常数值,挖掘和提取航磁梯度带^[10]的数据特征,得到大数据挖掘下的地质环境及成矿条件定量分析数据集。

对于露天非金属材料矿产资源的定量预测而言,在构建预测模型时,主要依靠对采集数据的挖掘分类,通过对分类数据特征的深度挖掘提取,实现矿产资源定量预测模型的最优匹配。

利用针对地质挖掘数据进行样本训练,实现模型构建前的数据分类,主要用于研究控矿分析要素,并计算所研究区域控矿分析要素的对应条件要素,分类概率模型。

假设在建立非金属材料矿产资源找矿和定量预测概念模型时,假设有 m 个模型 y_1, y_2, \dots, y_m 总体标记为 Y ,在研究过程中每个概念模型对应的控矿实际要素分别为 F_1, F_2, \dots, F_m 。

$$Y = \{y_1, y_2, \dots, y_m\} \quad (6)$$

针对所研究区域内的矿山进行控矿要素收集^[11],假设共有 n 个,将 n 个控矿要素对应的属性标记为向量 $X, \{y_1, y_2, \dots, y_m\}$ 对应的生成概率表示为 $\{p_1, p_2, \dots, p_m\}$, P_i 对应的是 X 属于 y_i 的概率。在控矿模型中假设第 i 个模型中包含 k_i 个要素,并将所有要素统一标记为 F_i ,具体表示为

$$F_i = \{f_i, f_i, \dots, f_{k_i}\} \quad (7)$$

此时 m 个模型中实际包含的控矿要素为 H 个:

$$H = \sum_{i=1}^m k_i \quad (8)$$

针对非金属材料矿产资源的找矿定量预测模型对应的先验概率 $P(y_i)$ 为

$$P(y_i) = \frac{k_i}{H} \quad (9)$$

假设矿产资源的定量研究区域中第 $j(1 \leq j \leq n)$ 个控矿要素在第 $i(1 \leq i \leq m)$ 个定量预测模型 Y_i 中出现的概率可表示为 $p(x_j | y_i)$,在定量预测模型中对应的各个控矿要素是相对独立的,利用贝叶斯算法可以计算得出包含在 m 个矿产资源定量模型中的概率为

$$P(y_i | X) = \frac{P(X | y_i) P(y_i)}{P(X)} \quad (10)$$

此时 $P(X)$ 作为分母需要保持常数状态,为了得定量预测的稳定性,需要将分子尽量调大,此时调整为

$$P(X | y_i) p(y_i) = P(x_1 | y_i) P(x_2 | y_i) \dots$$

$$P(x_n | y_i) = P(y_i) \prod_{j=1}^n p(x_j | y_i) \quad (11)$$

最终对矿产资源的分类结果为

$$P = \{p_1, p_2, \dots, p_m\} \quad (12)$$

式中, P_j 表示的是矿产预测区域中的主要控矿要素归属于定量预测模型 Y_j 的现实概率^[12]。

利用预测定量模型对非金属材料矿产资源进行预测分配时,主要通过预测模型中输入关键词,与模型之间完成自动匹配,在此基础上将控矿因素与定量模型之间完成要素匹配,针对性研究出控矿要素对矿产资源定量预测的重要性,首先根据 m 个定量预测模型筛选出对应的 M_1, M_2, \dots, M_m ,此时筛选出的模型相对应的控矿要素为 F_1, F_2, \dots, F_m ,对于第 i 个定量预测模型而言,在空旷要素的相关挖掘数据在经历清洗筛选时,主要按照矿产资源所在的地质条件进行分类,不同种类 c_i 下包含的空旷要素数量

可通过 $Num_{i1}, Num_{i2}, \dots, Num_{ic_i}$ ($1 \leq i \leq m$) 表示, 在第 i 个定量预测模型中的第 j 类地质条件类型, 每个对应类型的控矿要素在实际应用中的重要性可表示为

$$p_{ij} = \frac{1}{Num_{ij}} \quad (13)$$

式中, i 的取值范围保持在 $[1, m]$, j 的取值保持在 $[0, c_i]$ 。

控矿要素在多个定量预测模型中出现时, 最终的重要性评估指标为

$$P = \sum p_{ij} \quad (14)$$

式中, p_{ij} 表示每个定量预测模型中包含的控矿要素对应重要性。

根据筛选出的定量预测模型完成对应的要素使用率计算, 并得到最终的定量预测结果。

4 模型功能实验

针对露天非金属矿产资源的定量预测模型, 在全面的数据库支持下, 需要在其功能实现状态下验证具体的有效性, 利用地理信息查询软件, 根据所研究区域的钻孔数据, 完成露天非金属矿产资源的矿床三维空间建模, 实现空间数据分析。

利用 XJN3Ds 分析系统针对建模过程中的不同模块进行分析, 对研究区域实施地质勘探, 获取大量的地表及地下的相关地质信息和成矿条件, 通过量化描述可以更直观的表现出研究区地质体的三维形态模型, 本文选择的研究区域属于石灰岩矿区, 矿界中心位置在南北方向上存在跨级, 其中包含大断裂, 并以此为戒边界线, 明确非金属矿产资源特征之间明显的差异化。

首先通过在研究区域进行钻孔, 得到相应的地质成分数据, 将数据输入模型内完成非金属矿产资源三维模型, 为了保证钻孔数据有效且通用, 在钻孔时主要考虑钻孔的分布区域和规整程度, 共钻取 17 个孔洞, 根据钻孔所遇岩石层数可以基本确定出研究区域的地层总数量, 通过地层之间的上下关系确定具体的排列顺序, 为了方便解读分析, 在统计钻孔数据时将处于同一地层的数据分配到一个钻孔数据表格中, 钻孔数据见表 1。

根据上述统计出的钻孔数据生成对应的地层面, 为了能更真实的反应出地层面特征情况, 更好的利用钻孔数据间的空间相关性, 利用 ArcMap 软件

表 1 地层钻孔数据

钻孔编号	X 坐标	Y 坐标	深度/m
11	88.207 1	43.328 0	1957.42
12	88.306 2	43.316 5	1979.11
13	88.405 5	43.319 4	1978.87
14	88.506 1	43.306 2	1918.87
21	88.524 8	43.314 6	1932.41
31	88.538 7	43.317 6	1890.88
32	88.538 5	43.315 4	1890.88
33	88.619 8	43.316 8	1833.89
42	88.722 1	43.316 3	1823.89
43	88.780 7	43.318 5	1833.88
51	88.811 4	43.306 6	1833.88
52	88.811 3	43.309 1	1795.55
53	88.880 5	43.308 9	1828.87
54	88.900 6	43.310 2	1828.87
61	88.905 4	43.315 8	1967.17
71	88.914 6	43.326 2	1770.38
72	88.923 4	43.313 1	1770.37

加载钻孔数据并生成钻孔数据点图如图 1 所示。

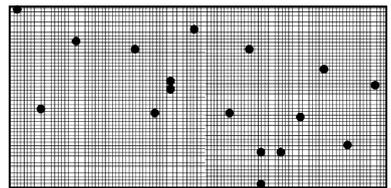


图 1 钻孔数据点图

根据钻孔点生成第零面地层图如图 2 所示。

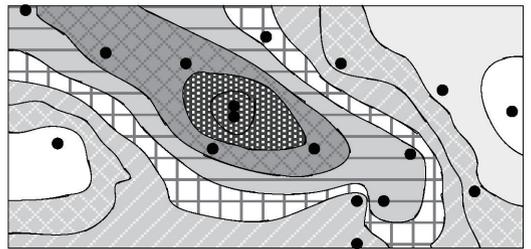


图 2 地表面 kriging 插值图

在此基础上重复图 1 和图 2 两步的操作, 依次生成四个地层对应的图像, 但此时生成得到的四个地层呈现状态是重叠在一起的, 为了能完整的展现出三维地层模型, 需要利用到 ArcScene 三维模块完成数据加载, 此时得到的三维显示图如图 3 所示。

通过图 3 能够清晰看到非金属矿产的预测三维边缘线, 为了使结果更加清晰, 此时根据三维显示输

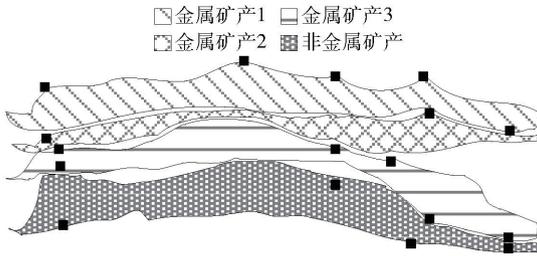


图3 四个地层三维显示图

出后的四个地层状态数据,从中提取地层面内边界,并换成多边形图形,具体如图4所示。

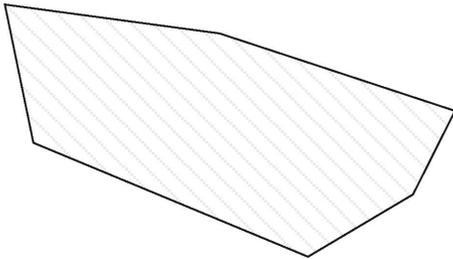


图4 地表面内边界提取图

此时将地表面内边界的提取图完成竖立转换,获得地质体状态图样,具体如图5所示。

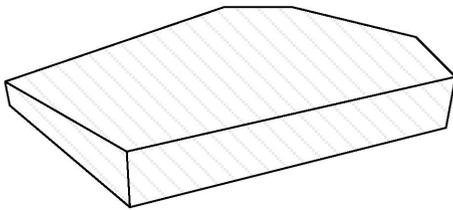


图5 地质体状态图样

重复多次平面的竖立转换后,最终生成完整的地质体如图6所示。

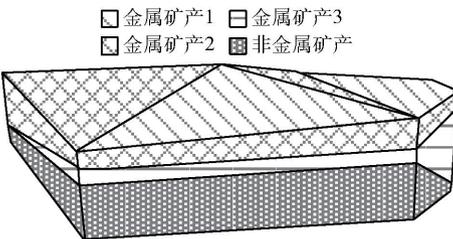


图6 最终生成区域非金属地质体图样

根据上述的地质地形图样建模输出,可以清晰的呈现出对非金属矿产资源所在地理空间,获得明确的定量预测结果,得到选定地质环境的坡度单位及输出大小,完成系数参考转换并构建出坡度栅格图,并且工作人员能够从中获得非金属矿产的路径走向分析,包括地质高距及高程的变化系数关系,获得的等高线图如图7所示。

基于图7能够得出,所研究区域内非金属矿产

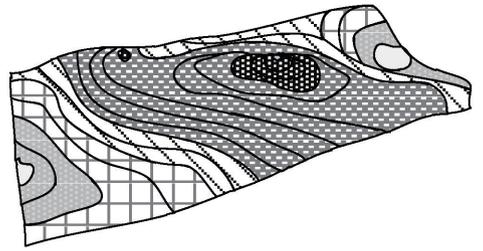


图7 等值线图

资源,矿体总体呈现U型发展趋势,自北向南完成阶梯态的平缓延展过渡,地质山体的走势从北向南逐渐上升,呈现中间高两边低的地形状态。

为了验证在构建模型下定量预测数据的准确度,利用正太分布图来统计分析预测结果,分布如图8所示。

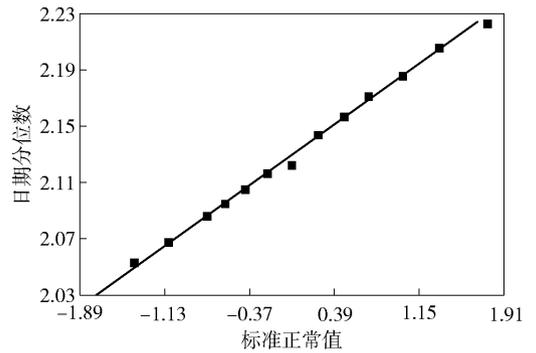


图8 统计数据绘制分布图

从图8中可以看出,针对研究区域的定量预测结果近似符合正态分布条件,预测的数据值分布均处于真实值附近,只存在个别的离散点,就此能够证明所研究方法的预测准确性。

5 结论

针对非金属矿产资源的定量预测和模型构建,通过对地理环境及信息的三维分析,使用大数据挖掘算法获得钻孔获取矿床数据中的矿产特征,在内边界基础上完成对应插值下的地质体构建,实现定量预测。通过功能实现也证明所研究的建模方法具有一定的稳定性,预测结果简单明了的同时定量预测结果也足够准确。

[参考文献]

[1] 左仁广. 基于数据科学的矿产资源定量预测的理论与方法探索[J]. 地学前缘, 2021, 28(3): 49-55.

[2] 马瑶, 赵江南. 机器学习方法在矿产资源定量预测应用研究进展[J]. 地质科技通报, 2021, 40(1): 132-141.

[3] 向杰, 肖克炎, 陈建平, 等. 基于成矿系统的三维定量

预测研究——以四川拉拉铜矿为例[J]. 地球学报, 2020,41(2):135-143.

[4] 屈云燕,田升平,商朋强,等. 中国重晶石矿资源预测模型及潜力分析[J]. 化工矿物与加工,2020,49(5):42-47.

[5] 胡鹏,张海坤,曹亮,等. 印度尼西亚苏门答腊岛斑岩型铜(钼)矿资源定量评价[J]. 矿床地质,2020,39(5):867-878.

[6] 成秋明. 极端地质事件定量模拟与预测[J]. 中国科学:地球科学,2022,52(6):975-991.

[7] 赵风顺,华杉,吴昊,等. 印度尼西亚苏门答腊岛浅成低温热液型金(银)矿资源总量预测[J]. 地质科技通报,2021,40(1):119-131.

[8] 于晓飞,公凡影,李永胜,等. 中国典型钴矿床地质特

征及重点地区矿产资源预测[J]. 吉林大学学报(地球科学版),2022,52(5):1377-1418.

[9] 余石勇,赵元艺. 美国阿拉斯加金矿资源特征及远景区预测[J]. 地质与勘探,2020,56(6):1296-1304.

[10] 赵汀,王安建,刘超. 基于国情调查大数据的矿产资源保障程度动态分析系统设计及实现[J]. 地质通报,2020,39(Z1):400-405.

[11] 马瑶,赵江南. 机器学习方法在矿产资源定量预测应用研究进展[J]. 地质科技通报,2021,40(1):132-141.

[12] 张振杰,成秋明,杨玠,等. 机器学习与成矿预测:以闽西南铁多金属矿预测为例[J]. 地学前缘,2021,28(3):221-235.

(上接第15页)

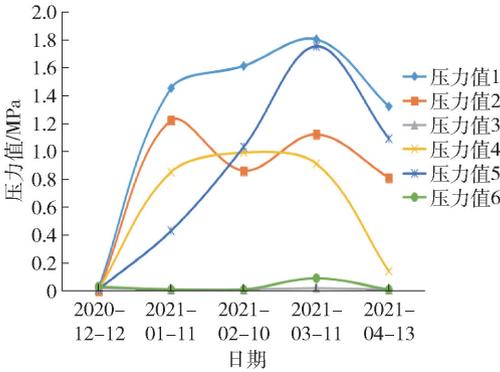


图12 应力观测图

但聚矿道发生的塑性变形较明显,主要发生在巷道顶底板。

(4)聚矿道与眉线分别采用增加砂浆锚杆+长锚索、钢拱架+长锚索支护方式效果显著,能保证聚矿道的行程。

(5)在断层位置进行锚杆支护时,由于围岩破碎,支护体能立即产生作用,在第三个月时达到应力最大值,此后围岩压力减小;距离采场较远时,断层应力不大。

(6)现场实际地质条件比模拟条件更复杂,多条断层带相互交织,模拟环境需要再进一步优化。

只有保证出矿进路、聚矿道稳定的情况下,才能进行下一步形成“爆破漏斗”施工,因此,该研究至关重要。

[参考文献]

[1] 艾敏,杨玉婷,许志发,等. 国内外自然崩落法研究与应用现状[J]. 中国水运(下半月),2018,18(11):241-243.

[2] 常晋元. 论铜矿峪矿自然崩落法拉底、崩落、出矿之关系[J]. 有色金属(矿山部分),2000(6):2-8+15.

[3] 刘知仁. 隧道工程中穿越断层破碎带开挖施工技术的应用探究[J]. 工程建设与设计,2022,(6):119-121.

[4] 周游,杨竹胜,周成嵩,等. 隧道断层破碎带坍塌机理及处理措施研究[J]. 能源与环保,2022,44(2):150-156.

[5] 谭文兵,冯钰炜,杜侠飞,等. 近断层深部巷道变形破坏分析及支护研究[J]. 有色金属(矿山部分),2022,74(2):46-52.

[6] 郭艳飞. 深井开采高应力巷道支护技术的研究应用[J]. 山西化工,2021,41(6):153-155.

[7] 刘斯忠,王洪江,吴爱祥,等. 破碎软岩中长锚索支护参数设计及其应用[J]. 铜业工程,2013(2):29-32.

[8] 李争荣,侯克鹏,冯兴隆,等. 复杂构造条件下自然崩落底柱巷道变形控制[J]. 中国矿业,2021,30(1):182-187.