



(12)发明专利申请

(10)申请公布号 CN 111709189 A

(43)申请公布日 2020.09.25

(21)申请号 202010573091.3

(22)申请日 2020.06.22

(71)申请人 中国地质科学院矿产资源研究所  
地址 100037 北京市西城区百万庄大街26号

(72)发明人 陈其慎 王琨 张艳飞 邢佳韵  
胡乔青 郑国栋 龙涛 魏然

(74)专利代理机构 北京高沃律师事务所 11569  
代理人 崔玥

(51)Int.Cl.  
G06F 30/27(2020.01)  
G06K 9/62(2006.01)  
G06N 20/00(2019.01)

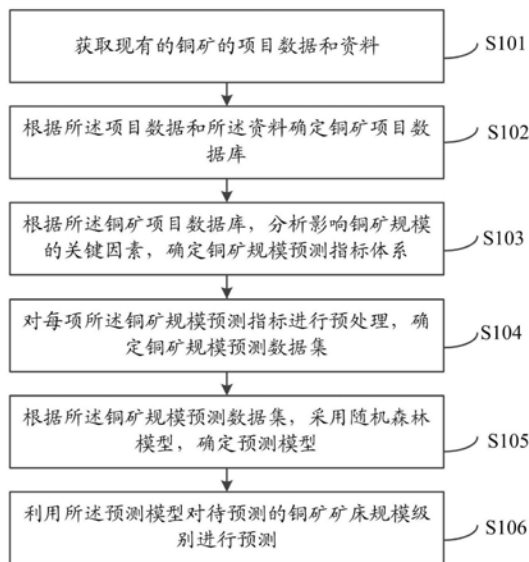
权利要求书2页 说明书8页 附图2页

(54)发明名称

一种铜矿矿床规模预测方法及系统

(57)摘要

本发明涉及一种铜矿矿床规模预测方法及系统。所述方法包括获取现有的铜矿的项目数据和资料;根据所述项目数据和所述资料确定铜矿项目数据库;根据所述铜矿项目数据库,分析影响铜矿规模的关键因素,确定铜矿规模预测指标体系;对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集;根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型;利用所述预测模型对待预测的铜矿矿床规模级别进行预测。本发明所提供的一种铜矿矿床规模预测方法及系统,实现对铜矿矿床规模的快速预测。



1. 一种铜矿矿床规模预测方法,其特征在于,包括:

获取现有的铜矿的项目数据和资料;所述项目数据包括项目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、累计查明资源储量、平均品位、成矿温度、成矿盐度和勘查程度;所述资料包括:矿区地质详查报告和钻探勘查报告;

根据所述项目数据和所述资料确定铜矿项目数据库;

根据所述铜矿项目数据库,分析影响铜矿规模的关键因素,确定铜矿规模预测指标体系;所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积;所述铜矿规模预测指标体系包括多项铜矿规模预测指标;所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深;

对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集;所述预处理包括分类、标准化处理和离散化处理;

根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型;所述预测模型以所述铜矿规模预测数据集为输入,以矿床规模级别为输出;

利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

2. 根据权利要求1所述的一种铜矿矿床规模预测方法,其特征在于,所述对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集,具体包括:

对所述铜矿规模预测指标进行分类,得到连续数据和离散数据;

对所述连续数据进行标准化处理;

对所述离散数据进行分类赋值;

根据标准化处理后的数据,利用ChiMerge算法进行离散化处理;

根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。

3. 根据权利要求2所述的一种铜矿矿床规模预测方法,其特征在于,所述对所述连续数据进行标准化处理,具体包括:

利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化处理;其中,  $X_s$  为标准化变量,  $X$

为连续数据,  $X_{min}$  为连续数据的最小值;  $X_{max}$  为连续数据的最大值。

4. 根据权利要求1所述的一种铜矿矿床规模预测方法,其特征在于,所述根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型,之前包括:

以所述铜矿规模预测数据集为输入,采用10折交叉验证方法确定初始随机森林模型;

根据所述铜矿规模预测数据集通过迭代确定随机森林模型。

5. 一种铜矿矿床规模预测系统,其特征在于,包括:

数据获取模块,用于获取现有的铜矿的项目数据和资料;所述项目数据包括项目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、累计查明资源储量、平均品位、成矿温度、成矿盐度和勘查程度;所述资料包括:矿区地质详查报告和钻探勘查报告;

铜矿项目数据库确定模块,用于根据所述项目数据和所述资料确定铜矿项目数据库;

铜矿规模预测指标体系确定模块,用于根据所述铜矿项目数据库,分析影响铜矿规模的关键因素,确定铜矿规模预测指标体系;所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积;所述铜矿规模预测指标体系包括多项铜矿规模预测指标;所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深;

铜矿规模预测数据集确定模块,用于对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集;所述预处理包括分类、标准化处理和离散化处理;

预测模型确定模块,用于根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型;所述预测模型以所述铜矿规模预测数据集为输入,以矿床规模级别为输出;

预测模块,用于利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

6. 根据权利要求5所述的一种铜矿矿床规模预测系统,其特征在于,所述铜矿规模预测数据集确定模块具体包括:

分类单元,用于对所述铜矿规模预测指标进行分类,得到连续数据和离散数据;

标准化处理单元,用于对所述连续数据进行标准化处理;

分类赋值单元,用于对所述离散数据进行分类赋值;

离散化处理单元,用于根据标准化处理后的数据,利用ChiMerge算法进行离散化处理;

铜矿规模预测数据集确定单元,用于根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。

7. 根据权利要求6所述的一种铜矿矿床规模预测系统,其特征在于,所述标准化处理单元具体包括:

标准化处理子单元,用于利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化处理;

其中, $X_s$ 为标准化变量, $X$ 为连续数据, $X_{min}$ 为连续数据的最小值; $X_{max}$ 为连续数据的最大值。

8. 根据权利要求5所述的一种铜矿矿床规模预测系统,其特征在于,还包括:

初始随机森林模型确定模块,用于以所述铜矿规模预测数据集为输入,采用10折交叉验证方法确定初始随机森林模型;

随机森林模型确定模块,用于根据所述铜矿规模预测数据集通过迭代确定随机森林模型。

## 一种铜矿矿床规模预测方法及系统

### 技术领域

[0001] 本发明涉及矿床规模预测技术领域,特别是涉及一种铜矿矿床规模预测方法及系统。

### 背景技术

[0002] 铜作为工业化建设的重要原料,是支撑电力、通讯、建筑、交通运输、机械制造、国防军工等产业发展的重要矿产资源。中国铜资源贫乏,国内铜原料供应远不能满足经济和社会发展的需要。如何在有限的资料条件下,快速对项目的优劣做出客观评价成为矿业公司面临的首要问题。

[0003] 由于境外矿业项目存在公开信息资料少、距离远、工作程度低等问题,国内矿业公司很难出国收集资料,可获取的国外项目信息非常有限,数据非常凌乱,无法客观评估这些项目的好坏,增加了海外投资项目的风险。

[0004] 另一方面,传统地质专家对铜矿矿床资源储量的评价需要系统的开展实地勘查工作,依赖钻探等勘查工程和长期实践经验,需要高昂的人力、经济和时间成本,无法在筛选项目初期,尚未开展系统勘查工作时,对目标矿床资源储量进行初步判断,预测目标项目是否可能成大矿的潜力和概率。

### 发明内容

[0005] 本发明的目的是提供一种铜矿矿床规模预测方法及系统,实现对铜矿矿床规模的快速预测。

[0006] 为实现上述目的,本发明提供了如下方案:

[0007] 一种铜矿矿床规模预测方法,包括:

[0008] 获取现有的铜矿的项目数据和资料;所述项目数据包括项目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、累计查明资源储量、平均品位、成矿温度、成矿盐度和勘查程度;所述资料包括:矿区地质详查报告和钻探勘查报告;

[0009] 根据所述项目数据和所述资料确定铜矿项目数据库;

[0010] 根据所述铜矿项目数据库,分析影响铜矿规模的关键因素,确定铜矿规模预测指标体系;所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积;所述铜矿规模预测指标体系包括多项铜矿规模预测指标;所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深;

[0011] 对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集;所述预处理包括分类、标准化处理和离散化处理;

[0012] 根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型;所述预测模型

以所述铜矿规模预测数据集为输入,以矿床规模级别为输出;

[0013] 利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

[0014] 可选的,所述对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集,具体包括:

[0015] 对所述铜矿规模预测指标进行分类,得到连续数据和离散数据;

[0016] 对所述连续数据进行标准化处理;

[0017] 对所述离散数据进行分类赋值;

[0018] 根据标准化处理后的数据,利用ChiMerge算法进行离散化处理;

[0019] 根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。

[0020] 可选的,所述对所述连续数据进行标准化处理,具体包括:

[0021] 利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化处理;其中, $X_s$ 为标准化变量, $X$ 为连续数据, $X_{min}$ 为连续数据的最小值; $X_{max}$ 为连续数据的最大值。

[0022] 可选的,所述根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型,之前包括:

[0023] 以所述铜矿规模预测数据集为输入,采用10折交叉验证方法确定初始随机森林模型;

[0024] 根据所述铜矿规模预测数据集通过迭代确定随机森林模型。

[0025] 一种铜矿矿床规模预测系统,包括:

[0026] 一种铜矿矿床规模预测系统,包括:

[0026] 数据获取模块,用于获取现有的铜矿的项目数据和资料;所述项目数据包括项目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、矿床资源储量、平均品位、成矿温度、成矿盐度和勘查程度;所述资料包括:矿区地质详查报告和钻探勘查报告;

[0027] 铜矿项目数据库确定模块,用于根据所述项目数据和所述资料确定铜矿项目数据库;

[0028] 铜矿规模预测指标体系确定模块,用于根据所述铜矿项目数据库,分析影响铜矿规模的关键因素,确定铜矿规模预测指标体系;所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积;所述铜矿规模预测指标体系包括多项铜矿规模预测指标;所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深;

[0029] 铜矿规模预测数据集确定模块,用于对每项所述铜矿规模预测指标进行预处理,确定铜矿规模预测数据集;所述预处理包括分类、标准化处理和离散化处理;

[0030] 预测模型确定模块,用于根据所述铜矿规模预测数据集,采用随机森林模型,确定预测模型;所述预测模型以所述铜矿规模预测数据集为输入,以矿床规模级别为输出;

[0031] 预测模块,用于利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

[0032] 可选的,所述铜矿规模预测数据集确定模块具体包括:

- [0033] 分类单元,用于对所述铜矿规模预测指标进行分类,得到连续数据和离散数据;
- [0034] 标准化处理单元,用于对所述连续数据进行标准化处理;
- [0035] 分类赋值单元,用于对所述离散数据进行分类赋值;
- [0036] 离散化处理单元,用于根据标准化处理后的数据,利用ChiMerge算法进行离散化处理;
- [0037] 铜矿规模预测数据集确定单元,用于根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。
- [0038] 可选的,所述标准化处理单元具体包括:

[0039] 标准化处理子单元,用于利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化处理;其中, $X_s$ 为标准化变量, $X$ 为连续数据, $X_{min}$ 为连续数据的最小值; $X_{max}$ 为连续数据的最大值。

[0040] 可选的,还包括:

- [0041] 初始随机森林模型确定模块,用于以所述铜矿规模预测数据集为输入,采用10折交叉验证方法确定初始随机森林模型;
- [0042] 随机森林模型确定模块,用于根据所述铜矿规模预测数据集通过迭代确定随机森林模型。

[0043] 根据本发明提供的具体实施例,本发明公开了以下技术效果:

[0044] 本发明所提供的一种铜矿矿床规模预测方法及系统,通过根据目前已有的铜矿项目数据信息,将矿床成矿理论与机器学习方法相结合,通过预测模型,即建立预测指标与矿床规模之间的关系模型,在大数据驱动下对目标铜矿矿床规模进行快速预测,对大量铜矿项目进行初步筛选,能大幅度节约时间成本,降低矿业投资风险,控制矿业投资规模,提高优质项目投资命中率。以解决矿业公司信息不对称、先验经验不足,面对无系统勘查数据的项目,无法对铜矿矿床规模进行预判的问题。

## 附图说明

[0045] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0046] 图1为本发明所提供的一种铜矿矿床规模预测方法流程示意图;

[0047] 图2为本发明所提供的一种铜矿矿床规模预测系统结构示意图。

## 具体实施方式

[0048] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0049] 本发明的目的是提供一种铜矿矿床规模预测方法及系统,实现对铜矿矿床规模级

别的预测。

[0050] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0051] 图1为本发明所提供的一种铜矿矿床规模预测方法流程示意图，如图1所示，本发明所提供的一种铜矿矿床规模预测方法，包括：

[0052] S101，获取现有的铜矿的项目数据和资料；所述项目数据包括项目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、矿床资源储量、平均品位、成矿温度、成矿盐度和勘查程度；所述资料包括：矿区地质详查报告和钻探勘查报告；

[0053] S102，根据所述项目数据和所述资料确定铜矿项目数据库。

[0054] S103，根据所述铜矿项目数据库，分析影响铜矿规模的关键因素，确定铜矿规模预测指标体系；所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积；所述铜矿规模预测指标体系包括多项铜矿规模预测指标；所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深。

[0055] 铜矿规模预测指标体系针对铜矿不同的成因类型会有一些的差异，所述指标体系主要针对斑岩型、矽卡岩型铜矿。

[0056] S104，对每项所述铜矿规模预测指标进行预处理，确定铜矿规模预测数据集；所述预处理包括分类、标准化处理和离散化处理。

[0057] S104具体包括：

[0058] 对所述铜矿规模预测指标进行分类，得到连续数据和离散数据。

[0059] 对所述连续数据进行标准化处理。利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化处理；其中， $X_s$ 为标准化变量， $X$ 为连续数据， $X_{min}$ 为连续数据的最小值； $X_{max}$ 为连续数据的最大值。

[0060] 对所述离散数据进行分类赋值。

[0061] 根据标准化处理后的数据，利用ChiMerge算法进行离散化处理。

[0062] 根据资源储量数值大小进行排序，每个数值属于一个区间。计算每一对相邻区间的卡方值，计算方法如下：

$$[0063] \quad \chi^2 = \sum_{i=1}^m \sum_{j=1}^k \frac{(A_{ij} - E_{ij})^2}{E_{ij}};$$

$$[0064] \quad E_{ij} = \frac{R_i * C_j}{N};$$

$$[0065] \quad R_i = \sum_{j=1}^k A_{ij};$$

$$[0066] \quad N = \sum_{j=1}^k C_j ;$$

$$[0067] \quad C_j = \sum_{i=1}^m A_{ij} ;$$

[0068] 其中,  $m=2$ ,  $k$ 为类的数量,  $A_{ij}$ 为第*i*区间第*j*类的实例数量,  $E_{ij}$ 为 $A_{ij}$ 的期望频率,  $R_i$ 为第*i*区间的实例数量,  $N$ 为总实例数量,  $C_j$ 为第*j*类的实例数量。

[0069] 根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。

[0070] S105, 根据所述铜矿规模预测数据集, 采用随机森林模型, 确定预测模型。所述预测模型以所述铜矿规模预测数据集为输入, 以矿床规模级别为输出。

[0071] 在S105之前还包括:

[0072] 以所述铜矿规模预测数据集为输入, 采用10折交叉验证方法确定初始随机森林模型。

[0073] 计算袋外数据(OOB)错误率作为随机森林模型的误差率, 误差率最小值对应的参数即为模型的最优参数。根据自变量(除矿床资源储量之外的数据)的Gini指数对自变量重要性进行排序, 保留80%数量的自变量, 重新构建随机森林模型, 重复这一过程直到自变量数量少于自变量总数的平方根。获得不同指标项组合和对应的随机森林模型误差率。

[0074] 根据所述铜矿规模预测数据集通过迭代确定随机森林模型。所述随机森林模型的误差率最小。

[0075] 所述随机森林模型具体为:

[0076] 将原始训练集记为:  $S = \{(x_i, y_i), i=1, 2, \dots, n\}$ ,  $(X, Y) \in R^d \times R$ , 采用bootstrap方法从原始样本集*S*中随机抽样, 样本集 $S_r = \{(x_r, y_r), r=1, 2, \dots, B\}$ 。

[0077] 使用 $S_r$ 生成一颗不剪枝的树 $h_i$ , 对*B*棵树循环进行如下步骤生成一系列决策树 $h_1(x), h_2(x), h_B(x)$ 。

[0078] 从*p*个自变量中随机选择 $M_{try}$ 个自变量。其中*p*为自变量个数,  $M_{try} = \sqrt{p}$ 。

[0079] 在每个节点根据Gini指数选取最优的分裂变量和分裂点。

[0080] 考虑分裂变量*j*和分裂点*s*, 定义一对半平面如下:

[0081]  $R_1(j, s) = \{X | X_j \leq s\}$ ,  $R_2(j, s) = \{X | X_j > s\}$ 。

[0082] 按下式求出分类变量*j*和分裂点*s*。

$$[0083] \quad \min_{x_i \in R_1(j, s)} \sum_{k=1}^K p_{m_k} (1 - p_{m_k}) + \min_{x_i \in R_2(j, s)} \sum_{k=1}^K p_{m_k} (1 - p_{m_k})。$$

[0084] 其中, 目标变量分为*K*类,  $m_1, m_k, p_{m_k}$ 代表属于*m<sub>k</sub>*类的样本占总样本的概率。

[0085] 分裂直到树的叶子节点中包含的样本量为5。

[0086] 得到*B*种分类结果, 对每个记录进行投票表决决定最终分类。

[0087]  $f(x_t) = \text{majorityvote} \{h_i(x_t)\}_{i=1}^B$

[0088] 其中,  $f(x_t)$ 为最终分类,  $h_i(x_t)$ 为单个决策树分类模型, majorityvote为多数票。



[0089] 最终分类 $f(x_t)$ 的误差率为误分个数占样本总数的比率。

[0090] S106,利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

[0091] 作为一个具体的实施例,获取了研究区265个斑岩型、矽卡岩型铜矿项目资料,由于本发明对铜矿矿床规模的预测基于不同成因类型铜矿大数据的综合分析,数据的数据和质量是影响预测准确度的重要因素。

[0092] 建立铜矿规模预测指标体系,包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变种类、硅化、矽卡岩化、绿泥石化/绿帘石化、钾长石化、绢云母化、碳酸盐化、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、是否处在汇聚型板块边缘、是否在俯冲带上、矿体总数、构造复杂程度、主矿体形态、矿体平均埋深、累计查明资源储量等。

[0093] 对铜矿规模预测指标进行预处理;对连续数据进行标准化处理,对离散数据进行分类赋值。如指标项是否在俯冲带上,在俯冲带上记为1,不在俯冲带上记为0。蚀变类型转化为硅化、矽卡岩化、绿泥石化/绿帘石化、钾长石化、绢云母化、碳酸盐化等指标,出现该蚀变特征记为1,未出现记为0。

[0094] 将矿床资源储量确定为因变量(预测目标),其他数据项为自变量(影响因素)。矿床资源储量具体指矿床的累计资源储量。利用ChiMerge算法对连续变量进行离散化处理,将连续变量转化为不同值域区间的离散变量。建立铜矿规模预测数据集;

[0095] 通过反复迭代构建随机森林模型,建立铜矿规模预测指标项和矿床规模之间的关系模型。

[0096] 构建初始随机森林模型,根据袋外数据(OOB)错误率对自变量重要性进行排序,保留80%数量的自变量,重新构建随机森林模型,重复这一过程直到自变量数量少于自变量总数的平方根;例如初始参加计算的一共有24个自变量,构建初始随机森林模型,通过bootstrap抽样方式产生500个样本子集,每次从24个自变量中随机选择5个自变量进行训练,输出袋外数据错误率,对24个自变量的重要性进行排序,保留前19个自变量,构建新的随机森林模型,重复这一过程,直到自变量个数少于5个时,停止计算。

[0097] 根据每个随机森林模型训练数据的错误率,筛选出最优的随机森林模型,利用随机森林模型构建预测模型对目标铜矿数据进行预测,得到目标铜矿项目的矿床规模级别和概率。

[0098] 最优的随机森林模型包含12个自变量组合,具体为主岩、是否在俯冲带上、构造复杂程度、蚀变种类、赋矿地层、成矿温度、品位、勘查程度、裂隙发育程度、绿泥石化/绿帘石化、平均埋深和矿体形态。

[0099] 本发明将矿床成矿理论与机器学习方法相结合,能够基于现有不同类型铜矿数据,对目标铜矿项目的矿床规模进行预判,通过验证数据证明本发明有较高的准确率和参考价值。

[0100] 图2为本发明所提供的一种铜矿矿床规模预测系统结构示意图,如图2所示,本发明所提供的一种铜矿矿床规模预测系统,包括:数据获取模块201、铜矿项目数据库确定模块202、铜矿规模预测指标体系确定模块203、铜矿规模预测数据集确定模块204、预测模型确定模块205和预测模块206。

[0101] 数据获取模块201用于获取现有的铜矿的项目数据和资料;所述项目数据包括项

目名称、所属国家、所属地区、经度、纬度、大地构造位置、成矿区带、矿床成因类型、成矿时代、成矿作用、赋矿地层、主岩、蚀变类型、矿石类型、矿体特征、累计查明资源储量、平均品位、成矿温度、成矿盐度和勘查程度；所述资料包括：矿区地质详查报告和钻探勘查报告。

[0102] 铜矿项目数据库确定模块202用于根据所述项目数据和所述资料确定铜矿项目数据库。

[0103] 铜矿规模预测指标体系确定模块203用于根据所述铜矿项目数据库，分析影响铜矿规模的关键因素，确定铜矿规模预测指标体系；所述关键因素包括勘查程度、成矿区带、成矿时代、赋矿地层、主岩、蚀变类型和岩体出露面积；所述铜矿规模预测指标体系包括多项铜矿规模预测指标；所述铜矿规模预测指标为是否在俯冲带上、是否在汇聚型板块边缘、成矿时代、赋矿地层、主岩、蚀变种类、蚀变分带性、矿石类型、裂隙发育程度、岩体出露面积、成矿温度、成矿盐度、平均品位、勘查程度、构造复杂程度、矿体总数、主矿体形态或矿体平均埋深。

[0104] 铜矿规模预测数据集确定模块204用于对每项所述铜矿规模预测指标进行预处理，确定铜矿规模预测数据集；所述预处理包括分类、标准化处理和离散化处理。

[0105] 预测模型确定模块205用于根据所述铜矿规模预测数据集，采用随机森林模型，确定预测模型；所述预测模型以所述铜矿规模预测数据集为输入，以矿床规模级别为输出。

[0106] 预测模块206用于利用所述预测模型对待预测的铜矿矿床规模级别进行预测。

[0107] 所述铜矿规模预测数据集确定模块具体包括：分类单元、标准化处理单元、分类赋值单元、离散化处理单元和铜矿规模预测数据集确定单元。

[0108] 分类单元用于对所述铜矿规模预测指标进行分类，得到连续数据和离散数据。

[0109] 标准化处理单元用于对所述连续数据进行标准化处理。

[0110] 分类赋值单元用于对所述离散数据进行分类赋值。

[0111] 离散化处理单元用于根据标准化处理后的数据，利用ChiMerge算法进行离散化处理。

[0112] 铜矿规模预测数据集确定单元用于根据所述离散化处理后的数据和分类赋值后的数据确定所述铜矿规模预测数据集。

[0113] 所述标准化处理单元具体包括：标准化处理子单元

[0114] 标准化处理子单元用于利用公式  $X_s = \frac{X - X_{min}}{X_{max} - X_{min}}$  对所述连续数据进行标准化

处理；其中， $X_s$ 为标准化变量， $X$ 为连续数据， $X_{min}$ 为连续数据的最小值； $X_{max}$ 为连续数据的最大值。

[0115] 本发明所提供的一种铜矿矿床规模预测系统还包括：初始随机森林模型确定模块和随机森林模型确定模块。

[0116] 初始随机森林模型确定模块用于以所述铜矿规模预测数据集为输入，采用10折交叉验证方法确定初始随机森林模型。

[0117] 随机森林模型确定模块用于根据所述铜矿规模预测数据集通过迭代确定随机森林模型。

[0118] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统

而言,由于其与实施例公开的方法相对应,所以描述的比较简单,相关之处参见方法部分说明即可。

[0119] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本发明的方法及其核心思想;同时,对于本领域的一般技术人员,依据本发明的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本发明的限制。

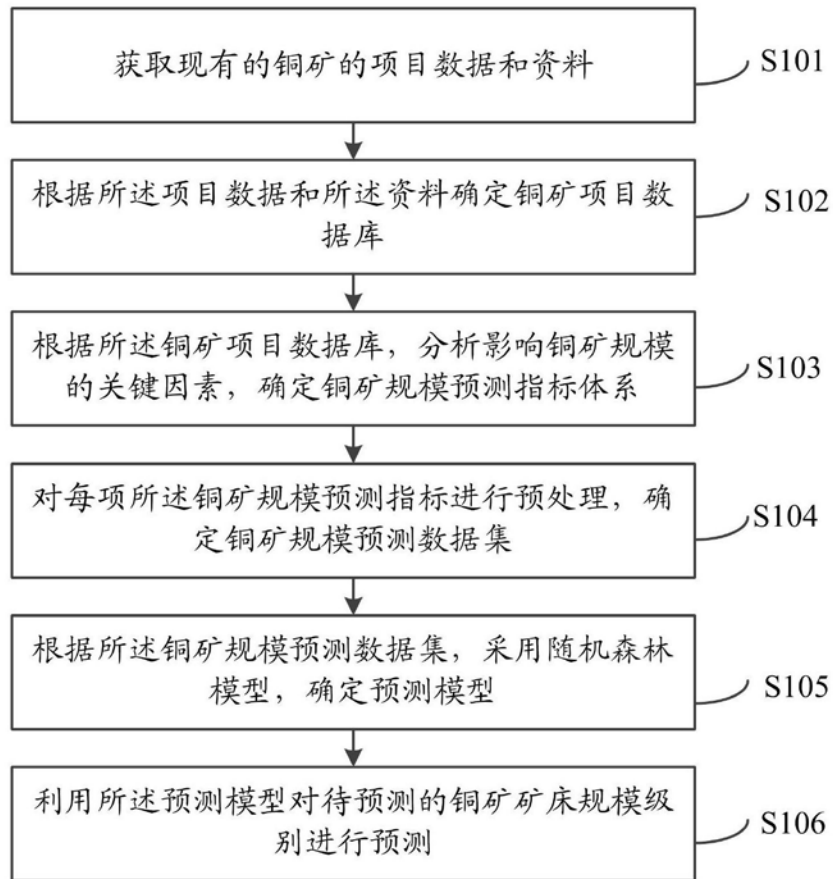


图1

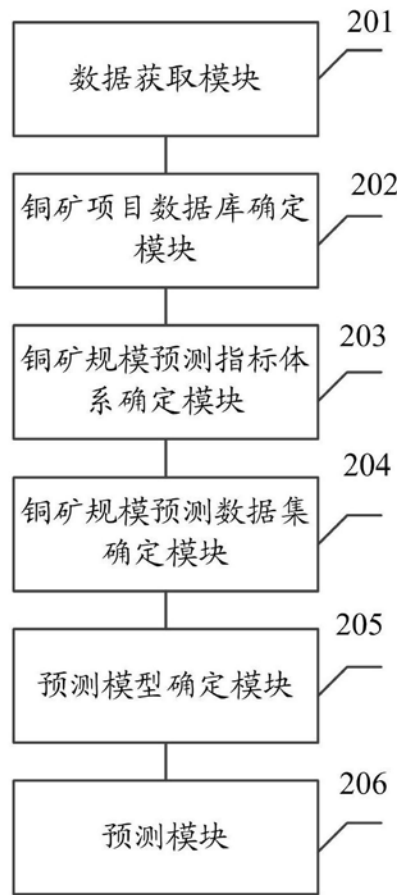


图2