



# (12) 发明专利申请

(10) 申请公布号 CN 113155809 A

(43) 申请公布日 2021.07.23

(21) 申请号 202110267879.6

(22) 申请日 2021.03.12

(71) 申请人 北京理工大学

地址 100081 北京市海淀区中关村南大街5号

(72) 发明人 刘瑞斌 邱苏玲 李安 殷允嵩

(51) Int. Cl.

G01N 21/71 (2006.01)

G06K 9/62 (2006.01)

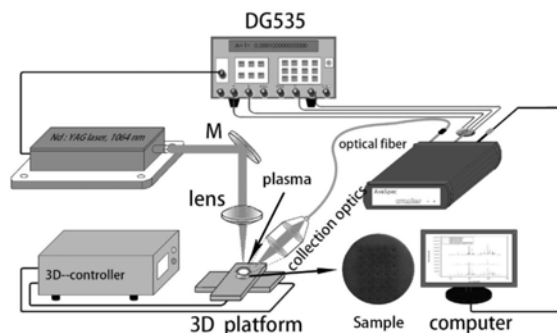
权利要求书2页 说明书7页 附图2页

## (54) 发明名称

一种矿石分类与实时定量分析的光谱检测新方法

## (57) 摘要

本发明公开了一种基于激光诱导击穿光谱对矿石分类与实时定量分析的光谱检测新方法,该方法包括:获取若干样品被激光激发后的光谱数据,首先进行光谱预处理来降低光谱的波动性,其次通过支持向量机对全部的矿石进行准确分类,并利用主成分分析法降维,提升元素定量分析的精度,之后通过相关性变量筛选偏最小二乘回归分析(R-PLS)改进算法,极大的降低了矿石的预测集的均方根误差以及平均相对误差。该技术去除了冗余的数据,使建模变量数大大减少,提高了运算速度。本发明基于激光诱导击穿光谱技术实现了对矿石中铁含量的定量分析,构建出一种利用激光诱导击穿光谱技术的对矿石分类与实时定量分析的光谱检测新方法。



1. 一种基于激光诱导击穿光谱对矿石分类与实时定量分析的检测新方法,其特征在于:包括如下步骤:

步骤1:用15Mpa的机械粉末压饼机将矿石样品压制成厚2mm、直径为13mm的小圆饼。获取同一类样品不同位置被激光激发后等离子体的光谱数据,每个位置重复测量5次,每种样品采集100组光谱数据。对同一样品分别重复测试N组,平均样品元素分布不均匀的影响,进而获取每个样品中元素成分的真实含量;

步骤2:利用拉依达准则对各个样品的异常光谱数据进行剔除,拉依达准则一般假设数据具有征态分布,计算数据的标准偏差并根据要求划定概率区间,最后对超出概率区间的粗大误差进行剔除、对峰位漂移修正和缺失峰补全。

异常光谱剔除算法实现:

(1) 采集到的LIBS光谱用 $X_{ij}$ 表示, $X_{ij}$ 指编号为i的样品第j个样本点的光谱数据,将每个通道光谱强度求和,各通道总光强中位数对应的光谱作为该样本中心点 $\bar{X}_i$ ;

(2) 求出每一个样本点离中心点的距离(即欧式距离  $D_j = \sqrt{\sum_{i=1}^n (X_{ij} - \bar{X}_i)^2}$ ),之后将其归一化(0-1);

(3) 计算剩余误差  $v_j = D_j - \bar{X}_i$ ,并按贝塞尔公式算出标准偏差 $\sigma$ ,若某个测量值 $D_j$ 的剩余误差 $v_j$ 满足  $|v_j| = |D_j - \bar{X}_i| > 3\sigma$  则认为含有粗大误差值的坏值,应予剔除。

步骤3:采用主成分分析结合支持向量机的分类方法。利用支持向量机对模型分类,将预处理后的全部光谱数据采用主成分分析法降维,在降维之后,随机选择了训练集和预测集,提取训练集的前10个主成分构造特征空间,在特征空间下采用小样本的5-折交叉验证来对训练集训练,然后对建模集和预测集进行分类正确率达到100%,可以对矿石进行准确分类。

步骤4:利用步骤(3)所述的方法进行分类后,对光谱数据进行相关性变量筛选偏最小二乘回归法(R-PLS)。将全部光谱强度作为输入数据,Fe的含量作为回归目标变量。模型的回归效果、定标精度、预测精度、预测误差分别用决定系数(R<sup>2</sup>)、校正集的均方根误差(RMSEC)、预测集的均方根误差(RMSEP)、平均相对误差(ARE)来综合衡量。

步骤5:对矿石分类后在进行R-PLS,最终得到待测物中Fe的含量,其正确率相比未分类+PLS和分类+PLS有了很大的提高。

2. 根据权利要求1所述的方法,其特征在于:步骤2中从待测物质的全部光谱数据中:首先对其进行异常光谱数据的筛选与剔除降低谱线的波动性,然后将光谱背景积分强度归一化、峰位漂移修正和缺失峰补全。

3. 根据权利要求1所述的方法,其特征在于:步骤3中采用主成分分析结合支持向量机的分类方法,先进行分类,在进行降维,在特征空间下采用小样本的5-折交叉验证来对训练集训练。

4. 根据权利要求1所述的方法,其特征在于:先用支持向量机进行准确分类后,再对光谱数据进行相关性变量筛选偏最小二乘回归法(R-PLS)进行分析,将全部光谱强度作为输入数据,Fe的含量作为回归目标变量。

5. 一种测试系统,其特征在于:包括:主动调Q的Nd:YAG 1064nm脉冲激光器、三维电动平台、四通道光纤光谱仪(Avantes Avsdesktop USB2,Netherlands)。

其中,所述Nd:YAG纳秒激光器均作为激发光源,激光光束聚焦于物体表面;

所述四通道光纤光谱仪用于采集物体被激发产生的等离子体的光谱数据,所述光谱数据被用于按照权利要求1-4任一项所述方法来获取样品中Fe元素的含量。

## 一种矿石分类与实时定量分析的光谱检测新方法

### 技术领域

[0001] 本发明属于全新的矿石分类与实时定量定性分析的光谱检测新方法,具体涉及一种利用激光诱导击穿光谱的成分定量分析方法、测试系统及处理算法,用于实现在低材料损耗的情况下,快速便捷地对矿石进行分类和定量分析。

### 背景技术

[0002] 激光诱导击穿光谱技术是一种利用脉冲激光对样品表面的小部分材料进行击穿并产生等离子体,等离子体冷却发光,随后对该等离子体所发射出的光谱进行分析,即是分析样品材料的组成成分的全元素检测手段。该测试手段具有诸多特点及优势。首先,激光诱导击穿光谱技术的检测过程对目标样品损伤小,仅需要对目标样品表面微克级别的材料进行击穿,可以认为是无损检测;其次,激光诱导击穿光谱技术的测试过程不需要对目标样品进行复杂的预处理,处理过程不需要使用化学试剂,对环境无污染。在测试过程中仅需要直接将脉冲激光聚焦于待测样品表面,即可得到其元素组成的相关信息;同时激光诱导击穿光谱技术对目标样品的形态没有特定的要求,对于气态,液态和固态的样品均可以进行测试;最后,由于激光诱导击穿光谱技术的测试过程是通过光谱信息进行分析从而得到测试结果,因此其测试过程也相对传统的检测方法比较快速。

[0003] 现如今,在科研以及工业应用领域,如生物医学,环境检测,港口海关,材料分析等诸多领域,对于目标样品的元素成分检测的需求都非常迫切。与化学分析方法或者高能射线探测法等传统元素测试方法相比较,激光诱导击穿光谱技术在测试过程中具备对于样品的损伤更小,样品制备等测试前准备过程更加便捷,测试过程耗时更短等特点。因此更加适合于各种需样品进行元素测试的场合。例如在生物医学相关等场合,要求测试手段能够在显微镜下操作。在港口海关,为了应对大量的货物吞吐量,在对货物进行检测时,则需要测试手段方便快捷,且对于货物的损伤较小。在材料分析领域,则需要其测试手段的测试范围足够宽,并且能够尽可能地对目标样品中的各种成分进行测试和分析。正是因为激光诱导击穿光谱技术的各种特点和优势,其能够很好的满足上述各种测试场合的要求。

[0004] 目前大多数研究多集中在对矿物成分、地理来源进行分类与鉴别上,以及LIBS结合化学计量学方法提升矿石分析精度。将分类和定量分析方法结合,通过支持向量机对光谱进行分类,然后采用相关性变量筛选偏最小二乘回归进行定量分析;在光谱的选择方面,依托主成分分析方法,建立光谱数据和典型元素含量线性关系,依次取出相关系数 $R$ 大于0.9,0.85,⋯,0的光谱进行偏最小二乘回归,分别计算出决定系数 $R^2$ ,预测集的均方根误差(RMSEP),平均相对误差(ARE)等信息。未经过分类的矿石经过全谱偏最小二乘回归分析后, $RMSEP=3.227\%$ , $ARE=31.75\%$ 。而将矿石利用我们提出的这种新方法即进行分类后再进行相关性变量筛选偏最小二乘回归分析(R-PLS),铁矿石、锰矿石、铬矿石的RMSEP分别降至0.975%、0.418%、0.123%,ARE分别降至1.46%、6.72%、1.09%。表明将矿石分类后再进行相关性变量筛选偏最小二乘回归分析大大提升了定量分析的准确性。因此,这种对矿石分类与实时定量分析的光谱检测新方法具有很大的优势。

## 发明内容

[0005] 本发明的目的是提供一种矿石分类和实时定量分析方法、测试系统,其利用激光诱导击穿光谱的技术来实现矿石分类和实时定量分析检测。

[0006] 一方面,本发明提供了一种元素定量分析方法,包括如下步骤:

[0007] 步骤1:获取同一类样品不同位置被激光激发后等离子体的光谱数据,每个位置重复测量5次,每种样品采集100组光谱数据。对同一样品分别重复测试N组,平均样品元素分布不均匀的影响,进而获取每个样品中元素成分的真实含量;

[0008] 所述同一类样品是指成分相同,含量不同的同类样品,譬如其含量是由国标测定;

[0009] 步骤2:采用主成分分析结合支持向量机的分类方法。利用支持向量机对模型分类,将预处理后的全部光谱数据采用主成分分析法降维,在降维之后,随机选择了训练集和预测集,提取训练集的前10个主成分构造特征空间,在特征空间下采用小样本的5-折交叉验证来对训练集训练,之后对矿石进行准确分类。

[0010] 步骤3:利用步骤(2)所述的方法进行分类后,对光谱数据进行相关性变量筛选偏最小二乘回归法(R-PLS)。将全部光谱强度作为输入数据,Fe的含量作为回归目标变量。模型的回归效果、定标精度、预测精度、预测误差分别用决定系数( $R^2$ )、校正集的均方根误差(RMSEC)、预测集的均方根误差(RMSEP)、平均相对误差(ARE)来综合衡量。

[0011] 步骤4:对矿石分类后在进行相关性变量筛选偏最小二乘回归分析算法(R-PLS),最终得到待测物中Fe的含量,其正确率相比未分类+偏最小二乘(PLS)和支持向量机分类+PLS有了很大的提高

[0012] 使用传统未分类+全谱PLS和分类+全谱PLS的方法计算所带来的误差比较大,其计算结果的不准确性主要来自于以下几个方面:(一)样品中冷却原子的自吸收效应,导致光谱信息中特征峰的强度与样品中实际的成分浓度出现一定程度的非线性依赖关系。(二)不同成分互相干涉,各个成分的特征峰之间可能存在一定程度的互相重叠并因此影响到定量分析结果的准确性。(三)样品在被激发后所产生的等离子气体在空间上和时间内存在一定程度的不均匀性,因此也会使得传统“偏最小二乘法回归”的计算结果的误差增大。本装置所采用的定量分析算法中,加入了支持向量机来分类,即:支持向量机分类+相关性变量筛选偏最小二乘回归分析改进算法进行定量分析时,其误差大大降低,提高了准确分类的精确度。

[0013] 进一步优选,步骤3的执行过程如下:

[0014] 从待测物质中的光谱数据中获取所述贡献度满足预设标准时所在波长点对应的的光谱数据,并输入训练后的所述偏最小二乘法回归模型得到待测物质中各个成分的预测含量;

[0015] 将待测物质中的光谱数据输入训练后的所述支持向量机回归模型得到待测物质中各个成分对应的残差;

[0016] 最后,将待测物质中各个成分的预测含量减去对应残差得到各个成分的含量。

[0017] 进一步优选,步骤2中贡献度满足预设标准的光谱数据是指贡献度最大所在波长点对应的的光谱数据。

[0018] 进一步优选,步骤1中获取光谱数据后,再对每个样品的所述光谱数据进行去噪处理,所述去噪处理包括:采用数据窗口移动的局部加权回归散点平滑法对光谱数据进行背

景连续光谱去除处理,其中,数据窗口中包含了n个波长点的光谱数据,n的取值范围为 $1-j_m*c_1$ , $j_m$ 表示一个扫描通道中得到波长点总数, $c_1$ 为扫描通道个数。

[0019] 进一步优选,步骤1中获取光谱数据后,再对每个样品的所述光谱数据进行去噪处理,所述去噪处理包括:对光谱数据进行归一化处理,所述归一化计算公式如下:

$$[0020] \quad I_{ij}^c = \frac{P_{ij}^c}{\sum_{j=(c-1)*j_m+1}^{c*j_m} P_{ij}^c}$$

[0021] 式中, $I'_{ij}$ 表示样品在第c个扫描通道进行的第i次扫描中第j个波长点对应的归一化后的光谱数据, $P_{ij}^c$ 表示样品在第c个扫描通道进行的第i次扫描中第j个波长点对应的归一化前的光谱数据, $j_m$ 表示一个扫描通道中得到波长点总数。

[0022] 进一步优选,利用粒子群算法对所述支持向量机模型的惩罚参数c和核参数g进行寻优。

[0023] 第二方面,本发明提供一种测试系统,包括:主动调Q的Nd:YAG 1064nm脉冲激光器、三维电动平台、四通道光纤光谱仪(Avantes Avsdesktop USB2,Netherlands);

[0024] 其中,所述Nd:YAG 1064nm脉冲激光器作为激发源,激光光束经过透镜组之后聚焦于物体表面;所述三维电动平台用来以一定的速率调节样品,使每一次激光都能作用于新的样品区域;所述四通道光纤光谱仪用在收集处,通过透镜将光耦合进光纤中传输到光谱仪。

[0025] 第三方面,本发明提供一种计算机,所述计算机上安装了可以运行的程序,所述处理器执行程序时实现上述方法的步骤。

[0026] 有益效果

[0027] 1、本发明构建的支持向量机分类+相关性变量筛选偏最小二乘回归分析改进算法,该方法首先采用拉依达准则对异常光谱进行剔除,其次采用主成分分析结合支持向量机的方法构建分类模型,对矿石快速、高准确度分类。随后通过相关性变量筛选偏最小二乘回归法,提高了模型的回归效果、定标精度、预测精度、预测误差分别用决定系数( $R^2$ )、校正集的均方根误差(RMSEC)、预测集的均方根误差(RMSEP)、平均相对误差(ARE)。实现了成分的定量分析,是一种矿石分类和实时定量分析的光谱检测新方法。

[0028] 2、本发明构建的测试系统通过使用纳秒激光器作为激发光源,使得激光诱导击穿光谱在实现全成分元素激发的前提下维持了较小的体积,利用本发明构建的测试系统,被测物体的前置准备工作需求少,测量方便,快捷,且该系统的适用范围广,可以研究任何形态,任何形貌的材料,提高了激光诱导击穿光谱的测量的准确性。

## 附图说明

[0029] 图1是本发明实施例提供的测试系统的结构示意图;

[0030] 图2是分类前后对35种矿石采用全谱偏最小二乘回归法得到的Fe含量的定标曲线(a)全部矿石;(b)铁矿石;(c)锰矿石;(d)铬矿石;

[0031] 图3是对35种矿石分类后再进相关性变量筛选偏最小二乘回归分析得到的Fe含量的定标曲线。(a)铁矿石;(b)锰矿石;(c)铬矿石;(d)全部矿石。

## 具体实施方式

[0032] 下面将结合实施例对本发明做进一步的说明。如图1所示,本发明实施例提供的是一种基于纳秒激光诱导击穿光谱的矿石分类和定时定量分析的测试系统,下述以其为实例进行具体的阐述,但是应当理解,本发明所述测试系统不限于本实施例中说限定的。

[0033] 本实施例中,测试系统包括:1064nm波长的Nd:YAG纳秒激光器、三维电动平台,四通道光纤光谱仪。

[0034] 激光器的能量、频率均为可调的,经过透镜聚焦后,激光聚焦于样品的表面。其中移动显微系统可以作为承载原件,在三轴上调节样品位置,辅助对焦,以及按照一定的步进位移样品,采集尽可能多的新的点。收集系统将等离子体发出的光耦合进入光纤中,进入光谱仪中,与PC相连,进行下一步的处理。

[0035] 基于本实施例提供上述测试系统,操作人员仅需要将待测样品放置于激光定位指示位置,即可对样品进行元素成分含量测量,然后对矿石进行分类以及实时定量分析。

[0036] 应当理解,其他可行的实施例中,Nd:YAG纳秒激光器、三维电动位移台、和收集光纤的安装位置可以进行相应调整,但是应当至少保证各个部件实现其基本功能,譬如,Nd:YAG皮秒激光器的激光束聚焦于待测物质的表面,光纤光谱仪可以采集到光谱数据,三位电动位移台有足够的移动空间。

[0037] 本应用案例所用样品为35种矿石,其中铁矿石14种,锰矿石12种和铬矿石9种列于表1。用15Mpa的机械粉末压片机将样品压制为厚为2mm、直径为13mm的小圆饼。

[0038] 表1铁矿石、锰矿石和铬矿石的样品型号及Fe含量

[0039] Table 1 Sample type and Fe content of the iron ore, manganese ore and chromium ore

Number	Sample name	Fe /%	category	Number	Sample name	Fe /%	category	
1	ZBK323	62.63		19	GBW07266	2.07		
2	YSBC28767-2008	63.07		20	ZBK333	8.05		
3	ZBK322	66.52		21	GBW07265	1.4		
4	YSBC28766-2008	62.65		22	YSBC26704-2013	6.03	manganese	
5	ZBK393	68.29		23	GSD03-2590-2010	2.75	ore	
6	ZBK392	65.71		24	QD 09-96	12		
7	ZBK321	46.93	iron ore	25	YSBC26701-2013	10.5		
8	YSBC28768-2008	66.18		26	Mn-k2-16	9.66		
[0040]	9	ZBK391	64.42	27	GBW07819	8.2809		
	10	GBW07829	64.49	28	ZBK441	12.9		
	11	YSBC28769-2008	67.84	29	ZBK440	9.76		
	12	GBW(E)070087	64.82	30	GBW07820	10.7288		
	13	GBW(E)070085	63.93	31	k3-4	11.19	chromium	
	14	YSB46701a	64.37	32	k3-1	9.71	ore	
	15	YSBC16701-2007	3.65	33	GSD333001-94	9.53		
	16	GBW07262	2.24	manganese	34	GBW07818	7.39266	
	17	ZBK334	8.1	ore	35	GSD333001.1-94	9.71	
	18	ZBK332	6.71					

[0041] 为了实现成分含量的测量,本发明基于激光诱导击穿光谱提供了一种成分定量分

析方法,包括如下步骤:

[0042] 步骤1:获取若干同一类样品被激光激发后等离子体的光谱数据并分别重复N组,以及获取每个样品中成分的真实含量。

[0043] 本实施例中,光谱信号通过光纤光谱仪进行收集。光纤光谱仪将样品被激发后所产生的等离子体发光的光谱信息进行收集。每次收集将得到一个完整的光谱数据,该光谱数据包含了各个波长上样品的等离子体发光的强度数值,按照波长的顺序依次排列,每一次扫描得到了 $j_m * c_1$ 个数据点的光谱数据, $j_m$ 表示一个扫描通道中得到波长点总数, $c_1$ 为扫描通道个数,本实施例中共有4个扫描通道, $j_m$ 等于2048,一共得到8192个数据点。本实施例中,每个位置重复测量5次,针对每个样品重复 $N=20$ 次,进而将得到100个完整的光谱信息,将这100个光谱信息排列成为一个矩阵,即为本装置扫描测试一个样品所得到的原始光谱数据矩阵 $X$ , $X_{i,j}$ 为其矩阵元素, $i$ 代表扫描测试的次数, $j$ 代表光纤光谱仪所取得的光谱数据中的强度数值按照波长顺序排列的次序,该强度数值的次序与其相应波长的数值一一对应,因此 $j$ 的数值事实上也代表了该强度数值所代表的波长的信息。因此,矩阵 $X$ 中的每一个成分即代表了光纤光谱仪100次测试中的第 $i$ 次测试所取得的第 $j$ 个强度数值。其他可行的实施例中, $N$ 为正整数。

[0044] 本实施例中优选采集到光谱数据后对其进行去噪处理,包括去除背景连续光谱以及归一化处理。

[0045] 去除杂乱光谱:

[0046] 为避免其对测量结果的影响,采用拉依达准则对异常光谱进行剔除。拉依达准则一般假设测量数据具有正态分布,计算数据的标准偏差并根据要求划定概率区间,最后对超出概率区间的粗大误差进行剔除。

[0047] 算法实现:

[0048] (1) 采集到的LIBS光谱用 $X_{i,j}$ 表示, $X_{i,j}$ 指编号为 $i$ 的样品第 $j$ 个样本点的光谱数据,将每个通道光谱强度求和,各通道总光强中位数对应的光谱作为该样本中心点 $\bar{X}_i$ 。

[0049] (2) 求出每一个样本点离中心点的距离(即欧式距离 $D_j = \sqrt{\sum_{i=1}^n (X_{ij} - \bar{X}_i)^2}$ ),如图2(a)所示为样品的光谱数据对应的欧式距离分布图。之后将其归一化(0-1)。

[0050] (3) 为了更直观地看出距离的概率密度分布规律,可画出分段距离与频数的关系。

[0051] (4) 计算剩余误差 $v_j = D_j - \bar{X}_i$ ,并按贝塞尔公式算出标准偏差 $\sigma$ ,若某个测量值 $D_j$ 的剩余误差 $v_j$ 满足 $|v_j| = |D_j - \bar{X}_i| > 3\sigma$ 则认为 $v_j$ 是含有粗大误差值的坏值,应予剔除。

[0052] 去除异常光谱后,大部分谱线的波动性均有所下降。然后采用光谱背景积分强度归一化和峰位漂移修正和缺失峰补全对去除异常光谱后的数据进行处理。

[0053] 归一化处理:

[0054] 修正后的线状光谱数据将通过“光谱背景积分强度归一化法”进行光谱不确定性的处理。即对每一个通道的背景光谱强度进行单独的归一化,以此减少光谱的不确定性,方便后续步骤分析各成分的浓度差异。线状光谱中的光谱异常,如光谱漂移,偶尔出现的噪音光谱等也将在这一步中被去除。计算公式如下:



$$[0055] \quad I_{ij}^c = \frac{P_{ij}^c}{\sum_{j=(c-1) \times j_m + 1}^{c \times j_m} P_{ij}^c}$$

[0056] 式中,  $I_{ij}^c$  表示样品在第  $c$  个扫描通道进行的第  $i$  次扫描中第  $j$  个波长点对应的归一化后的光谱数据,  $P_{ij}^c$  表示样品在第  $c$  个扫描通道进行的第  $i$  次扫描中第  $j$  个波长点对应的归一化前的光谱数据, 即本实施例中对应为光谱数据  $S_{ik}$  展开后的数据,  $N \times K$  个光谱数据  $S_{ik}$  展开后又是  $N \times (j_m \times c_1)$  大小的矩阵,  $P_{ij}^c$  为矩阵元素, 此时, 再以各个扫描通道的光谱数据为单元分别进行归一化处理。

[0057] 步骤2: 利用支持向量机结合主成分分析进行分类: 由于矿石中光谱特征峰数量非常多, 光谱特征峰数量相近, 但谱线强度及峰位置存在明显差异。将预处理后的全部光谱数据 (3500\*8192) 即3500组光谱, 一个谱有8192个点, 采用主成分分析降维, 假设前  $M$  个主成分累计解释率达到92%以上, 可以认为前  $M$  个主成分已经涵盖了矿石光谱数据的绝大部分信息。在降维之后, 随机选择了训练集和预测集的比例关系为:25:10, 提取训练集的前10个主成分构建特征空间, 在特征空间下采用小样本的5-折交叉验证法来对训练集进行训练, 其建模集和预测集的分类正确率均为100%。采用主成分分析结合支持向量机的分类方法, 可以实现矿石快速、高准确度分类鉴别。

[0058] 步骤3: 相关性变量筛选偏最小二乘回归法

[0059] 由于支持向量机可以实现对全部矿石准确的分类, 因此对经过分类后的光谱数据进行相关性变量筛选偏最小二乘回归法 (R-PLS)。该方法去除了冗余的数据, 使建模变量数大大减少, 提高了运算速度。模型的回归效果、定标精度、预测精度、预测误差分别用决定系数 ( $R^2$ )、校正集的均方根误差 (RMSEC)、预测集的均方根误差 (RMSEP)、平均相对误差 (ARE) 来综合衡量。

[0060] 具体如下:

[0061] 基于训练后的支持向量机+相关性变量筛选偏最小二乘回归法, 利用待测物质的光谱数据得到待测元素的含量。具体如下:

[0062] 1、利用支持向量机结合主成分分析进行分类, 具体如下:

[0063] 主成分分析主成分个数选取准则: 随着主成分个数的增多, 建模效果越好。当主成分的个数较少时, 预测精度不够, 所以随着主成分个数的增加, RMSEP值变小, 预测精度上升, 但主成分的个数太多时又会导致过拟合, RMSEP值变大, 预测精度降低。因此主成分个数的选取通过RMSEP最小方法确定。

[0064] 2、进行相关性变量筛选, 具体如下:

[0065] 首先建立光谱数据和Fe元素含量线性关系, 分别选出相关系数  $R$  大于0.9、0.85、...、0.05、0的光谱数据, 依次进行偏最小二乘回归分析。

[0066] 当选择的相关系数较小 ( $R \in [0-0.5]$ ) 时, 预测集的均方根误差 (RMSEP) 随  $R$  的变化, 其变化规律不明显, 但整体上RMSEP偏大。当  $R$  大于0.5时, RMSEP随  $R$  的变化呈现出先减小后增大的趋势, 这是因为当  $R$  太小时, 与Fe元素含量相关性较弱峰被选择的多, 使得分析精度不够。而当  $R$  取的太大时, 由于选择的变量太少, 容易造成过拟合且容差能力弱, 同样影响

了分析精度。相关性变量筛选偏最小二乘回归分析达到最佳预测效果的选择依据是：选择 RMSEP 最小对应的相关系数。

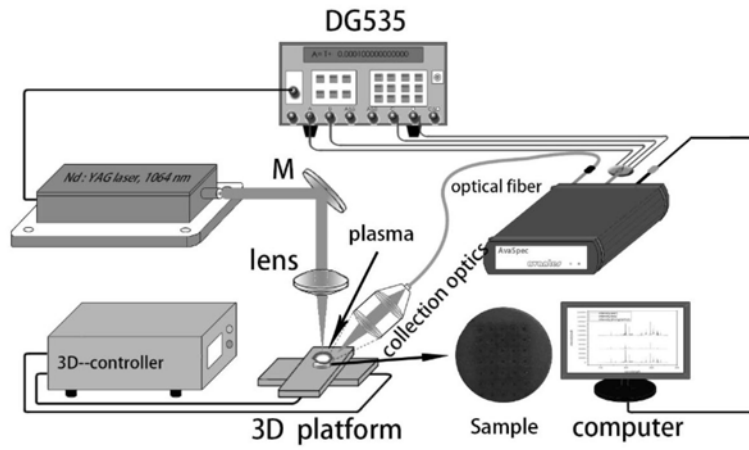


图1

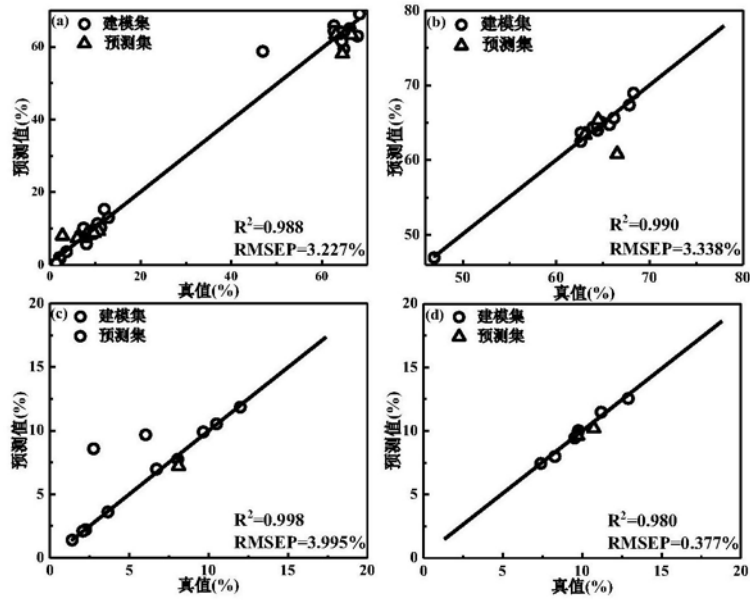


图2

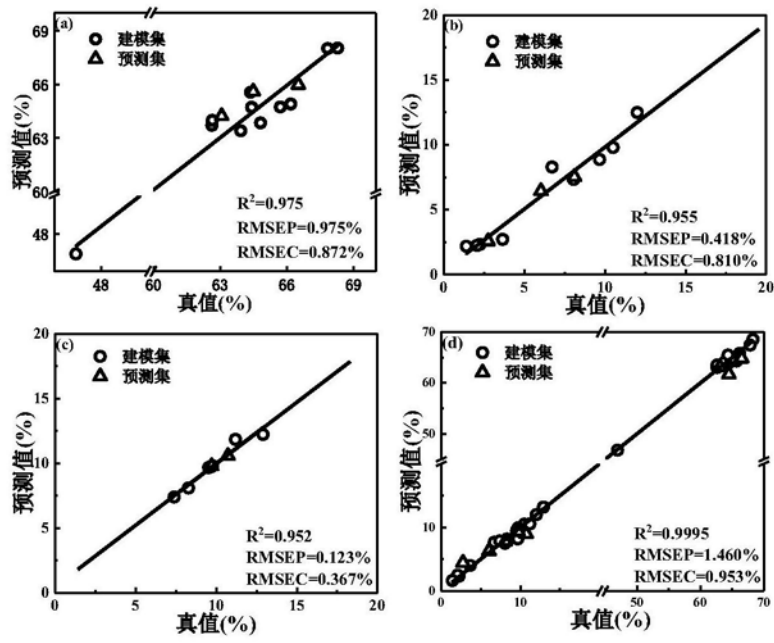


图3