



(12) 发明专利

(10) 授权公告号 CN 101915744 B

(45) 授权公告日 2012. 11. 07

(21) 申请号 201010218291. 3

(22) 申请日 2010. 07. 05

(73) 专利权人 北京航空航天大学

地址 100083 北京市海淀区学院路 37 号

(72) 发明人 张广军 李庆波 李丽娜

(74) 专利代理机构 北京汇泽知识产权代理有限公司 11228

代理人 程殿军

(51) Int. Cl.

G01N 21/35 (2006. 01)

(56) 对比文件

李响等. 一种血糖红外无创检测的改进小波预处理方法. 《光学学报》. 2010, 第 30 卷 (第 2 期), 579-584.

李丽娜等. 基于交互式自模型混合物分析的近红外光谱波长变量优选方法. 《分析化学》. 2009, 第 37 卷 (第 6 期), 823-827.

崔广. 基于近红外光谱分析的纺织品中羊毛

含量检测技术研究. 《中国优秀硕士学位论文全文数据库》. 2007, (第 5 期), 37-52.

Roman Rosipal. Kernel Partial Least Squares for Nonlinear Regression and Discrimination. 《Neural Network World》. 2003, 第 13 卷 (第 3 期), 291-300.

Roman Rosipal. Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Space. 《Journal of Machine Learning Research》. 2001, (第 2 期), 97-123.

审查员 张小丽

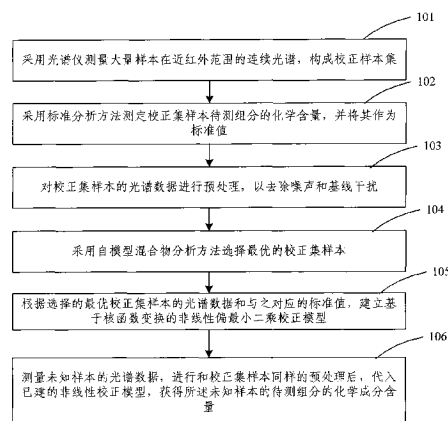
权利要求书 4 页 说明书 12 页 附图 10 页

(54) 发明名称

物质成分含量的近红外光谱无损检测方法
及装置

(57) 摘要

本发明公开一种物质成分含量的近红外光谱无损检测方法及装置, 利用其装置, 通过采用光谱仪采集校正集样本、对光谱进行预处理、通过样本优选选择最优的校正样本集、和利用最优的校正样本集建立非线性校正模型等步骤, 然后利用光谱仪采集未知成分含量的光谱样本, 经与校正集样本相同的光谱预处理, 通过已建非线性校正模型检测该未知样本的成分含量。采用本发明提供的方法和装置, 能够有效解决现有物质成分含量的近红外无损检测方法中校正模型复杂、训练速度慢、不易于硬件实现的问题, 并能显著提高物质成分含量无损检测结果的精度及稳定性。



1. 一种物质成分含量的近红外光谱无损检测方法,其特征在于,该检测方法包括:

A、利用近红外光谱仪采集大量被测样品的近红外光谱数据,构成校正样本集;

B、采用标准分析方法测定校正样本集中样本待测组分的化学含量的真值;

C、对所述近红外光谱数据进行预处理,以去除噪声、基线或其他干扰待测物质成分信息的无用信号;

D、对上述经预处理后的校正集样本的光谱数据进行基于自模型混合物分析的校正集样本优选,具体为:

D1、将光谱矩阵 X_{new} 进行转置得到 X'_{new} ,则 X'_{new} 矩阵每一列各元素代表一个样本在各个波长下的吸光度值;

D2、计算光谱矩阵 X'_{new} 中各列向量即样本 i 的纯度值 $p_{i,1}$,其计算公式为: $p_{i,1} = \sigma_i / (\mu_i + \alpha)$

其中, μ_i 为均值、 σ_i 为标准差、 α 为补偿因子;

D3、根据步骤 D2 求得各列向量 i 的 $p_{i,1}$ 值,判断 $p_{i,1}$ 值的大小,将具有最大 $p_{i,1}$ 值的第 i 个列向量作为选出的第一个样本;

D4、选择第 k ($k \geq 2$) 个样本,具体为:根据下列公式计算矩阵 X'_{new} 中列向量 i 的长度

$$l_i, l_i = \sqrt{\sum_{j=1}^n (d_{i,j})^2 / n};$$

其中, $d_{i,j}$ 为光谱矩阵 X'_{new} 中第 i 行第 j 列元素,由 $d(l)_{i,j} = d_{i,j} / \sqrt{\mu_i^2 + (\sigma_i + \alpha)^2}$,得到关系矩阵 $C = D(l)D(l)^T/n$,根据下列公式计算关系权函数 $\rho_{i,k}$;

$$\rho_{i,k} = \begin{vmatrix} c_{i,i} & c_{i,p_1} & \dots & c_{i,p_{k-1}} \\ c_{p_1,i} & c_{p_1,p_1} & \dots & c_{p_1,p_{k-1}} \\ \dots & \dots & \dots & \dots \\ c_{p_{k-1},i} & \dots & \dots & c_{p_{k-1},p_{k-1}} \end{vmatrix}$$

其中, k 表示待确定的第 k 个样本, p_{k-1} 表示目前已经选定了的第 $(k-1)$ 个样本在 C 矩阵中所在列向量的标号, p_1 表示通过步骤 D3 已选择的第一个样本在 C 矩阵中所在列向量的标号;则纯度值 $p_{i,k}$ 为:

$p_{i,k} = \rho_{i,k} (\sigma_i / (\mu_i + \alpha))$;将具有最大 $p_{i,k}$ 值的第 i 个列向量作为选出的第 k 个样本;

D5、重复上述步骤 D4,并通过依次迭代选择的样本建立多元校正模型,交互验证评价模型的性能,将具有最小预测均方根误差 RMSEP 时选取的样本个数作为最优的样本个数,所述 RMSEP 计算公式为:

$$\text{RMSEP} = \sqrt{\frac{\sum_i^n (\hat{y} - y)^2}{n}}; \text{其中, } \hat{y} \text{ 为预测值, } y \text{ 为参考值, } n \text{ 为样本个数};$$

D6、用优选出的一定个数的样本组成新的最优的校正样本集;

E、用优选出的校正样本集建立基于核函数变换的非线性偏最小二乘校正模型;具体

为：

E1、确定初始参数，包括核函数及其参数、主成分个数；

E2、对校正集的自变量进行核函数变换，然后进行中心化处理；其中，核函数变

换公式为：
$$\mathbf{K}_{train} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix}$$
；n 表示校正集样本的个数， $k(\cdot, \cdot)$

表示两个向量的内积； x_i ($i = 1, \dots, n$) 表示校正集样本的光谱；核函数变换过程即是非线性处理过程，核函数变换之后的矩阵包含了非线性信息；中心化的公式为

$\tilde{\mathbf{K}}_{train} = (\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T) \mathbf{K}_{train} (\mathbf{I} - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T)$ ； $\mathbf{I}$ 表示单位矩阵， $\mathbf{1}_n$ 表示各元素都为 1 的 n 维向量；

E3、对经核函数变换后得到的矩阵 $\tilde{\mathbf{K}}_{train}$ ，建立偏最小二乘校正模型，并根据交互验证方法确定最优的核函数、核参数以及主成分个数；所述最终建立的基于核函数变换的非线性偏最小二乘校正模型为 $\hat{\mathbf{y}} = \tilde{\mathbf{K}}_{train} \mathbf{b} + \mathbf{f}$ ；其中， $\mathbf{b}$ 为 $\tilde{\mathbf{K}}_{train}$ 矩阵原始变量的偏最小二乘回归系数向量， $\mathbf{f}$ 为残差向量；

F、用所构造的基于核函数变换的非线性偏最小二乘校正模型，检测未知样本的物质成分含量。

2. 根据权利要求 1 所述的物质成分含量的近红外光谱无损检测方法，其特征在于，步骤 C 所述对所述近红外光谱数据进行预处理，进一步包括小波变换过程，包括：

C1、选择小波基和小波分解层数，将光谱矩阵 X 的每条信号进行小波分解；其中光谱矩阵 X 的每一行中各元素代表一个样本在各个波长下的吸光度值；

C2、对光谱信号在小波域内的低频段小波系数置零来实现基线校正，利用较高频段小波系数阈值处理来实现噪声去除；

C3、用分解后去噪和基线校正后的第 N 层低频和高频系数进行信号重构，重构的各条光谱信号组成新的光谱矩阵 X_{new} 。

3. 根据权利要求 1 所述的物质成分含量的近红外光谱无损检测方法，其特征在于，步骤 C 所述对所述近红外光谱数据进行预处理，进一步包括微分处理过程，具体为：选择微分窗口宽度 g 及微分阶次；并对所述光谱数据进行一阶微分去除与波长无关的漂移，或进一步进行二阶微分处理去除与波长相关的漂移。

4. 根据权利要求 1 所述的物质成分含量的近红外光谱无损检测方法，其特征在于，所述核函数主要有协方差核、多项式核和高斯核；所述协方差核 $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ 、多项式核 $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + p)^q$ 、高斯核 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2 / (2 \times \sigma^2))$ ；

其中，i、j 分别表示第 i、j 个样本的光谱向量；p、q、 σ 分别为需根据实际情况确定的核参数。

5. 根据权利要求 1 所述的物质成分含量的近红外光谱无损检测方法，其特征在于，步骤 F 所述检测未知样本的物质成分含量的过程，包括：

F1、用光谱仪采集未知样本的光谱；

F2、对于未知样本的光谱，采用与校正集样本相同的预处理方法进行光谱处理，去除噪声、基线以及其他干扰物质成分信息的无用信息；

F3、对于经预处理后的未知样本的光谱进行核函数变换,然后进行中心化处理;未知样本的核函数变换公式为:

$$\mathbf{K}_{test} = \begin{bmatrix} k(x_{t_1}, x_1) & k(x_{t_1}, x_2) & \dots & k(x_{t_1}, x_n) \\ k(x_{t_2}, x_1) & k(x_{t_2}, x_2) & \dots & k(x_{t_2}, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_{t_l}, x_1) & k(x_{t_l}, x_2) & \dots & k(x_{t_l}, x_n) \end{bmatrix};$$

其中, x_{t_i} ($i = 1, \dots, n$) 表示未知的样本的光谱; x_i ($i = 1, \dots, n$) 表示校正集样本的光谱;中心化的公式为 $\tilde{\mathbf{K}}_{test} = (\mathbf{K}_{test} - \frac{1}{n} \mathbf{I}_l \mathbf{I}_n^T \mathbf{K}_{train})(\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T)$;

F4、通过已建好的模型获得未知样本的物质成分含量的预测值;计算公式为 $\hat{\mathbf{y}}_{test} = \tilde{\mathbf{K}}_{test} \mathbf{b}$; $\hat{\mathbf{y}}_{test}$ 为最终通过上述近红外光谱无损检测方法获得的检测结果的值。

6. 一种实现权利要求 1 所述的物质成分含量的近红外光谱无损检测方法的物质成分含量的近红外光谱无损检测装置,其特征在于,该装置主要包括校正集样本优选单元(100)、模型建立单元(200)和预测值获取单元(300);其中,

校正集样本优选单元(100),用于获取用于建立基于核函数变换的非线性偏最小二乘校正模型的最优的校正集样本;该校正集样本优选单元(100)进一步包括:校正集样本数据获取子单元(110),用于获取大量已知样本的近红外连续光谱数据,以及与之相应的采用标准分析方法测得的校正样本集中样本待测组分的化学含量的真值,构成校正集样本;校正集样本数据预处理子单元(120),用于对所述校正集样本数据获取子单元(110)获取的所述校正集样本光谱数据进行预处理;最优校正集样本获取子单元(130),用于根据所述校正集样本数据预处理子单元(120)获取的预处理后的校正集样本光谱数据和所述校正集样本数据获取子单元(110)中获取的相应的真值,确定最优的一定数量的用于建立校正模型的样本并提供给所述模型建立单元(200)和预测值获取单元(300);

模型建立单元(200),用于建立基于核函数变换的非线性偏最小二乘校正模型;所述模型建立单元(200)包括:校正集样本核函数变换子单元(210),用于对所述最优校正集样本获取子单元(130)得到的一定数量的最优校正集样本进行核函数变换,构成校正集样本的核函数变换后的矩阵;基于核函数变换的非线性偏最小二乘校正模型获取子单元(220),用于对所述校正集样本核函数变换子单元(210)中得到的核函数变换后的矩阵建立非线性偏最小二乘校正模型,并将获得的回归系数提供给预测值获取单元(300);以及

预测值获取单元(300),用于得到所述的物质成分含量近红外光谱无损检测的结果。

7. 根据权利要求 6 所述的物质成分含量的近红外光谱无损检测装置,其特征在于,所述预测值获取单元(300)包括:

未知样本光谱数据获取子单元(310),用于获取待测的未知成分含量的样本的近红外波段的连续光谱;

未知样本光谱数据预处理子单元(320),用于对所述未知样本光谱数据获取子单元(310)获取的未知样本数据进行预处理,所采用的预处理方法及其参数与所述校正集样本数据预处理子单元(120)所采用的相同;

未知样本核函数变换子单元(330),利用所述最优校正集样本获取子单元(130)获得的最优的校正集样本,用于对所述未知样本光谱数据预处理子单元(320)获取的预处理后的未知样本光谱数据进行核函数变换,所采用的核函数及其参数与所述校正集样本核函数

变换子单元 (210) 所采用的相同, 构成未知样本的核函数变换后的矩阵;

预测值计算子单元 (340), 用于根据所述基于核函数变换的非线性偏最小二乘校正模型获取子单元 (220) 得到的回归系数和所述未知样本核函数变换子单元 (330) 得到的未知样本的核变换后的矩阵, 通过计算得到所述未知样本的被测物质成分含量的检测值。

物质成分含量的近红外光谱无损检测方法及装置

技术领域

[0001] 本发明涉及近红外光谱分析技术,尤其涉及一种物质成分含量的近红外光谱无损检测方法及装置。

背景技术

[0002] 近红外光谱检测技术具有无损伤、无污染、实时快速测定物质成分浓度(或者性质参数)的特点,因此被广泛应用于农业、生物医学、化工等领域。虽然近红外光谱能携带丰富的物质分子结构信息,且容易获取和被处理,可用来分析绝大多数种类的化合物及其混合物的成份浓度(或者性质参数),但是,近红外光谱信息强度比中红外谱区低、谱峰宽,样品近红外光谱的有效信息率低。对复杂样品进行近红外光谱分析即是要从复杂、重叠、变动的光谱中提取微弱信息。而随着高性能光谱仪器及基础测试技术的不断完善,基于化学计量学的近红外光谱微弱信息提取方法研究,则成为了近红外光谱无损检测物质成分浓度(或者性质参数)的关键点之一,它决定着近红外光谱检测技术的最终应用与前景。

[0003] 多元校正是一种重要的化学计量学方法,多元校正是近红外光谱检测技术中的一个关键步骤,用于建立在已知物质成分浓度(或性质参数)与光谱之间的关联关系,利用该关系,可用于后续预测未知样品的物质成分浓度(或性质参数),高精度多元校正模型的建立可使得物质成分含量的近红外光谱无损检测的结果更精确。

[0004] 多元校正方法可分为线性和非线性两大类。线性校正方法如多元线性回归、主成分回归、偏最小二乘回归等,其理论性质易于讨论,常常能比较好的解决问题。然而,随着人们对事物认识能力的提高,对于复杂样品,多组分混合物分析体系中各组分的相互作用、仪器的基线漂移等,致使吸收光谱偏离朗伯-比耳定律,呈现非线性的特质,为了建立预测准确性好和稳健性强的近红外光谱分析多元校正模型,非线性建模方法的研究越来越引起人们的关注。目前常见的非线性校正方法有非线性偏最小二乘法、人工神经网络等。基于样条变换、多项式拟合等的非线性偏最小二乘法,选择恰当的参数能很好的拟合非线性关系,但是仍然不能解决模型的复杂性问题;人工神经网络是一种智能的非线性多元校正方法,具有很强的非线性建模能力,但也存在一些局限性,如模型和参数选择复杂、训练速度慢、容易陷入极小点、过拟合等等。

[0005] 另外,校正集样本是建立多元校正模型的基础,多元校正即是根据已知的一定数量的校正集样本的光谱和标准方法测得的该校正集样本的物质成分浓度(或性质参数)建立关联关系。由于样本化学组成复杂,且样本信息包含在相似性很强的高维光谱中,校正集样本的合理筛选对提高模型预测精度至关重要。通常要求校正集样本应具有范围宽、分布均匀、精度高、典型性的特点,选择特征信息丰富、尽量少干扰的足够数量的校正集样本是模型准确的前提,决定了模型的适应性和可靠性。

[0006] 目前常用的校正集样本优选方法有:随机法;基于同类就近选择样本的光谱欧氏距离算法;根据样本光谱之间距离差异进行样本挑选的算法,如 Kennard-Stone 法等。随机法选取样本完全是随意的,没有任何规律,或者仅遵循简单的规则,通过该方法每次组成校

正集的样本可能差异很大,不能保证所选样本代表性及模型的外推能力。同类就近选择样本的建模方法只能降低非线性的校正误差,无法发现除去光谱特征异常样本干扰,往往不够可靠和精确。Kennard-Stone法的优点是能保证训练库中样本按照空间距离分布均匀,但是需要进行数据转换和计算样本两两空间距离,计算量大。

[0007] 综上所述,利用常用的非线性多元校正方法和校正集样本优选方法的近红外光谱检测技术,均存在模型复杂、计算量大的问题,以及还存在物质成分含量近红外光谱无损检测中校正模型的训练速度慢、不易于硬件实现等缺陷;且采用现有非线性校正模型不适用于小样本建模、容易产生过拟合,影响了物质成分含量近红外光谱无损检测结果的精度。另外,因离线训练的校正模型复杂,通常的物质成分含量的近红外光谱无损检测装置硬件中写入的校正模型固定,这对于成分特别复杂的被测对象的物质成分含量的检测的适用性不佳。

发明内容

[0008] 有鉴于此,本发明的主要目的在于提供一种简单、快速、高效的物质成分含量的近红外光谱无损检测方法及装置,本发明提供的方法简洁,便于近红外光谱无损检测技术中的硬件实现,根据被测对象情况调节模型及其参数,能提高物质成分含量检测结果的精度和适用性。

[0009] 本发明的另一个目的在于提供一种校正集样本优选方法,通过基于自模型混合物分析的技术,消除含重复信息或无用信息的样本,从而解决样本间共线性的技术问题,从而选出少数的具有代表性样本,用于建立多元校正模型,可达到简化训练过程、提高建模速度、降低模型复杂性以便于近红外光谱无损检测过程中的硬件实现和提高物质成分含量近红外光谱无损检测结果的精度的技术效果。

[0010] 为达到上述目的,本发明的技术方案是这样实现的:

[0011] 一种物质成分含量的近红外光谱无损检测方法,该检测方法包括:

[0012] A、利用近红外光谱仪采集大量被测样品的近红外光谱数据,构成校正样本集;

[0013] B、采用标准分析方法测定校正样本集中样本待测组分的化学含量的真值;

[0014] C、对所述近红外光谱数据进行预处理,以去除噪声、基线或其他干扰待测物质成分信息的无用信号;

[0015] D、对上述经预处理后的校正集样本的光谱数据进行基于自模型混合物分析的校正集样本优选;

[0016] E、用优选出的校正样本集建立基于核函数变换的非线性偏最小二乘校正模型;

[0017] F、用所构造的基于核函数变换的非线性偏最小二乘校正模型,检测未知样本的物质成分含量。

[0018] 其中,步骤C所述对所述近红外光谱数据进行预处理,进一步包括小波变换过程,包括:

[0019] C1、选择小波基和小波分解层数,将光谱矩阵X的每条信号进行小波分解;其中光谱矩阵X的每一行中各元素代表一个样本在各个波长下的吸光度值;

[0020] C2、对光谱信号在小波域内的低频段小波系数置零来实现基线校正,利用较高频段小波系数阈值处理来实现噪声去除;

[0021] C3、用分解后去噪和基线校正后的第 N 层低频和高频系数进行信号重构, 重构的各条光谱信号组成新的光谱矩阵 X_{new} 。

[0022] 类似地, 步骤 C 所述对所述近红外光谱数据进行预处理, 进一步包括微分处理过程, 具体为: 选择微分窗口宽度 g 及微分阶次; 并对所述光谱数据进行一阶微分去除与波长无关的漂移, 或进一步进行二阶微分处理去除与波长相关的漂移。

[0023] 其中, 步骤 D 所述的基于自模型混合物分析的校正集样本优选方法, 包括:

[0024] D1、将光谱矩阵 X_{new} 进行转置得到 X'_{new} , 则 X'_{new} 矩阵每一列各元素代表一个样本在各个波长下的吸光度值;

[0025] D2、计算光谱矩阵 X'_{new} 中各列向量即样本 i 的纯度值 $p_{i,1}$, 其计算公式为:

$$[0026] \quad p_{i,1} = \sigma_i / (\mu_i + \alpha)$$

[0027] 其中, μ_i 为均值、 σ_i 为标准差、 α 为补偿因子;

[0028] D3、根据步骤 D2 求得的各列向量 i 的 $p_{i,1}$ 值, 判断 $p_{i,1}$ 值的大小, 将具有最大 $p_{i,1}$ 值的第 i 个列向量作为选出的第一个样本;

[0029] D4、选择第 k ($k \geq 2$) 个样本, 具体为: 根据下列公式计算矩阵 X'_{new} 中列向量 i 的

$$\text{长度 } l_i, l_i = \sqrt{\sum_{j=1}^n (d_{i,j})^2 / n};$$

[0030] 其中, $d_{i,j}$ 为光谱矩阵 X'_{new} 中第 i 行第 j 列元素, 由 $d(l)_{i,j} = d_{i,j} / \sqrt{\mu_i^2 + (\sigma_i + \alpha)^2}$, 得到关系矩阵 $C = D(1)D(1)^T/n$, 根据下列公式计算关系权函数 $\rho_{i,k}$;

$$[0031] \quad \rho_{i,k} = \begin{vmatrix} c_{i,i} & c_{i,p_1} & \cdots & c_{i,p_{k-1}} \\ c_{p_1,i} & c_{p_1,p_1} & \cdots & c_{p_1,p_{k-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ c_{p_{k-1},i} & \cdots & \cdots & c_{p_{k-1},p_{k-1}} \end{vmatrix}$$

[0032] 其中, k 表示待确定的第 k 个样本, p_{k-1} 表示目前已经选定了的第 $(k-1)$ 个样本在 C 矩阵中所在列向量的标号, p_1 表示通过步骤 D3 已选择的第一个样本在 C 矩阵中所在列向量的标号; 则纯度值 $p_{i,k}$ 为:

[0033] $p_{i,k} = \rho_{i,k} (\sigma_i / (\mu_i + \alpha))$; 将具有最大 $p_{i,k}$ 值的第 i 个列向量作为选出的第 k 个样本;

[0034] D5、重复上述步骤 D4, 并通过依次迭代选择的样本建立多元校正模型, 交互验证评价模型的性能, 将具有最小预测均方根误差 (RMSEP, Root Mean Square Error of Prediction) 时选取的样本个数作为最优的样本个数, 所述 RMSEP 计算公式为:

$$[0035] \quad \text{RMSEP} = \sqrt{\frac{\sum_i (\hat{y} - y)^2}{n}}; \text{其中, } \hat{y} \text{ 为预测值, } y \text{ 为参考值, } n \text{ 为样本个数};$$

[0036] D6、用优选出的一定个数的样本组成新的最优的校正样本集。

[0037] 步骤 E 所述建立基于核函数变换的非线性偏最小二乘校正模型, 包括:

[0038] E1、确定初始参数, 包括核函数及其参数、主成分个数;

[0039] E2、对校正集的自变量进行核函数变换, 然后进行中心化处理; 其中, 核函数变换

公式为： $\mathbf{K}_{train} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix}$ ；n 表示校正集样本的个数，

[0040] $k(\cdot, \cdot)$ 表示两个向量的内积； $x_i (i = 1, \dots, n)$ 表示校正集样本的光谱；核函数变换过程即是非线性处理过程，核函数变换之后的矩阵包含了非线性信息；中心化的公式为 $\tilde{\mathbf{K}}_{train} = (\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T) \mathbf{K}_{train} (\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T)$ ； $\mathbf{I}$ 表示单位矩阵， $\mathbf{1}_n$ 表示各元素都为 1 的 n 维向量。

[0041] E3、对经核函数变换后得到的矩阵 $\tilde{\mathbf{K}}_{train}$ ，建立偏最小二乘校正模型，并根据交互验证方法确定最优的核函数、核参数以及主成分个数；所述最终建立的基于核函数变换的非线性偏最小二乘校正模型为 $\hat{\mathbf{y}} = \tilde{\mathbf{K}}_{train} \mathbf{b} + \mathbf{f}$ ；其中， $\mathbf{b}$ 为 $\tilde{\mathbf{K}}_{train}$ 矩阵原始变量的偏最小二乘回归系数向量， $\mathbf{f}$ 为残差向量。

[0042] 其中，所述核函数主要有协方差核、多项式核和高斯核；所述协方差核 $k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ 、多项式核 $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + p)^q$ 、高斯核 $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2 / (2 \times \sigma^2))$ ；

[0043] 其中，i、j 分别表示第 i、j 个样本的光谱向量；p、q、 σ 分别为需根据实际情况确定的核参数。

[0044] 步骤 F 所述检测未知样本的物质成分含量的过程，包括：

[0045] F1、用光谱仪采集未知样本的光谱；

[0046] F2、对于未知样本的光谱，采用与校正集样本相同的预处理方法进行光谱处理，去除噪声、基线以及其他干扰物质成分信息的无用信息；

[0047] F3、对于经预处理后的未知样本的光谱进行核函数变换，然后进行中心化处理；未知样本的核函数变换公式为：

[0048] $\mathbf{K}_{test} = \begin{bmatrix} k(x_{t1}, x_1) & k(x_{t1}, x_2) & \dots & k(x_{t1}, x_n) \\ k(x_{t2}, x_1) & k(x_{t2}, x_2) & \dots & k(x_{t2}, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_{tI}, x_1) & k(x_{tI}, x_2) & \dots & k(x_{tI}, x_n) \end{bmatrix}$ ；

[0049] 其中， $x_{ti} (i = 1, \dots, n)$ 表示未知的样本的光谱； $x_i (i = 1, \dots, n)$ 表示校正集样本的光谱；中心化的公式为 $\tilde{\mathbf{K}}_{test} = (\mathbf{K}_{test} - \frac{1}{n} \mathbf{I}_I \mathbf{I}_n^T \mathbf{K}_{train}) (\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T)$ ；

[0050] F4、通过已建好的模型获得未知样本的物质成分含量的预测值。计算公式为 $\hat{\mathbf{y}}_{test} = \tilde{\mathbf{K}}_{test} \mathbf{b}$ ； $\hat{\mathbf{y}}_{test}$ 为最终通过上述近红外光谱无损检测方法获得的检测结果的值。

[0051] 一种实现权利要求 1 所述的物质成分含量的近红外光谱无损检测方法的物质成分含量的近红外光谱无损检测装置，该装置主要包括校正集样本优选单元 100、模型建立单元 200 和预测值获取单元 300；其中，

[0052] 校正集样本优选单元 100，用于获取用于建立基于核函数变换的非线性偏最小二乘校正模型的最优的校正集样本；

[0053] 模型建立单元 200，用于建立基于核函数变换的非线性偏最小二乘校正模型；以及

[0054] 预测值获取单元 300，用于得到所述的物质成分含量近红外光谱无损检测的结果。

[0055] 其中,所述校正集样本优选单元 100 进一步包括:

[0056] 校正集样本数据获取子单元 110,用于获取大量已知样本的近红外连续光谱数据,以及与之相应的采用标准分析方法测得的校正样本集中样本待测组分的化学含量的真值,构成校正集样本;

[0057] 校正集样本数据预处理子单元 120,用于对所述校正集样本数据获取子单元 110 获取的所述校正集样本光谱数据进行预处理;以及

[0058] 最优校正集样本获取子单元 130,用于根据所述校正集样本数据预处理子单元 120 获取的预处理后的校正集样本光谱数据和所述校正集样本数据获取子单元 110 中获取的相应的真值,确定最优的一定数量的用于建立校正模型的样本并提供给所述模型建立单元 200 和预测值获取单元 300。

[0059] 其中,所述模型建立单元 200 和预测值获取单元 300 进一步包括:

[0060] 校正集样本核函数变换子单元 210,用于对所述最优校正集样本获取子单元 130 得到的一定数量的最优校正集样本进行核函数变换,构成校正集样本的核函数变换后的矩阵;

[0061] 基于核函数变换的非线性偏最小二乘校正模型获取子单元 220,用于对所述校正集样本核函数变换子单元 210 中得到的核函数变换后的矩阵建立非线性偏最小二乘校正模型,并将获得的回归系数提供给预测值获取单元 300;以及

[0062] 所述预测值获取单元 300 进一步包括:

[0063] 未知样本光谱数据获取子单元 310,用于获取待测的未知成分含量的样本的近红外波段的连续光谱;

[0064] 未知样本光谱数据预处理子单元 320,用于对所述未知样本光谱数据获取子单元 310 获取的未知样本数据进行预处理,所采用的预处理方法及其参数与所述校正集样本数据预处理子单元 120 所采用的相同;

[0065] 未知样本核函数变换子单元 330,利用所述最优校正集样本获取子单元 130 获得的最优的校正集样本,用于对所述未知样本光谱数据预处理子单元 320 获取的预处理后的未知样本光谱数据进行核函数变换,所采用的核函数及其参数与所述校正集样本核函数变换子单元 210 所采用的相同,构成未知样本的核函数变换后的矩阵;

[0066] 预测值计算子单元 340,用于根据所述基于核函数变换的非线性偏最小二乘校正模型获取子单元 220 得到的回归系数和所述未知样本核函数变换子单元 330 得到的未知样本的核变换后的矩阵,通过计算得到所述未知样本的被测物质成分含量的检测值。

[0067] 本发明所提供的物质成分含量的近红外光谱无损检测方法及装置,具有以下优点:

[0068] 本发明通过利用基于自模型混合物分析技术的校正集样本优选方法,仅选出少量的样本用于建立多元校正模型,可快速建立校正模型,达到提高近红外光谱无损检测技术中多元校正模型的建模效率的效果,同时由于样本优选使得模型的复杂性降低,从而还可达到便于近红外光谱无损检测过程中的硬件实现的技术效果;另外,若对于采集到的大量的校正集样本,由于本发明提出的样本优选方法可消除样本间的共线性问题,从而能选出少而精的、具有代表性的样本用于建立校正模型,能够用于提高物质成分含量近红外光谱无损检测结果的精度和可靠性。本发明方法采用基于核函数变换的方式建立的非线性校

正模型,仅通过向量的内积完成非线性映射过程,其变换过程简单易实现,从而进一步简化了多元校正模型,降低了物质成分含量近红外光谱无损检测过程硬件实现的困难;仅通过核函数及其参数的选择,可应用于不同的待测量对象,因而本发明提供的方法具有广泛的适应性;另外,本发明所述的非线性校正模型具有适应于小样本、具有较好的泛化能力的特点,尤其适用于经校正集样本优选后的少量样本建模情况,能达到显著提高物质成分含量近红外光谱无损检测结果的精度的效果。

附图说明

- [0069] 图1为本发明物质成分含量的近红外光谱无损检测方法流程图;
- [0070] 图2为采用本发明方法获取的原始近红外光谱示意图;
- [0071] 图3为经微分预处理后得到的近红外光谱示意图;
- [0072] 图4为基于自模型混合物分析的校正集样本优选方法通过最大纯度值选择第一个样本的示意图;
- [0073] 图5为基于自模型混合物分析的校正集样本优选方法通过交互验证得到的预测均方根误差(RMSEP)确定最佳的校正集样本个数图;
- [0074] 图6为本发明校正集样本经核函数变换后的曲线图;
- [0075] 图7为本发明采用优选出的校正集样本建立基于核函数变换的非线性偏最小二乘校正模型的回归系数曲线图;
- [0076] 图8为本发明所述的未知样本经核函数变换后的曲线图;
- [0077] 图9为采用本发明物质成分含量的近红外光谱无损检测方法用于预测未知样本得到的预测结果与参考值的相关性图;
- [0078] 图10为本发明物质成分含量的近红外光谱无损检测装置组成结构示意图。

具体实施方式

- [0079] 下面结合附图及本发明的实施例对本发明的方法及装置作进一步详细的说明。
- [0080] 本发明的核心思想是:利用近红外光谱仪采集大量样品的光谱样本,构成校正样本集,通过预处理方法对校正样本集光谱进行预处理,以去除噪声和基线等无用信息,以改善光谱质量,提高光谱与被测物质成分含量之间的相关性;通过自模型混合物分析技术,从校正样本集中选取有代表性的少量样本的子集,构成最优的校正样本集,用于后续的非线性多元校正模型的建立,以提高建模效率、降低模型复杂性以及便于物质成分含量近红外光谱无损检测装置的硬件实现;采用优选出的校正样本集,建立基于核函数变换的非线性偏最小二乘校正模型,利用采集到的大量的校正集样本,通过选择最优的校正集样本,并根据实际被测对象情况选择预处理和校正模型及其参数,从而达到提高物质成分含量近红外无损检测结果的精度和适用性的目的。
- [0081] 图1为本发明物质成分含量的近红外光谱无损检测方法流程图,如图1所示,该方法包括:
- [0082] 步骤101、利用近红外光谱仪采集大量的被测样品的近红外光谱数据,构成校正样本集。
- [0083] 这里,本发明以谷物蛋白质含量的近红外光谱无损检测过程为例,对被测谷物进

行近红外光谱采集,所述光谱采集范围为 1100 ~ 2498nm,波长变量个数为 700 个,光谱矩阵中每一行各元素代表一个样本在各个波长下的吸光度值。校正集样本总个数为 60 个,其原始光谱的光谱矩阵 X 的图形如图 2 所示。

[0084] 步骤 102、采用标准分析方法测定校正样本集中样本待测组分的化学含量的真值。

[0085] 这里,采用标准分析方法测定上述 60 个校正集样本的蛋白质含量真值范围为 7.6540 ~ 9.7110%,呈随机分布,其浓度标准差为 0.4986%。

[0086] 步骤 103、对所述近红外光谱数据进行预处理,去除噪声、基线或其他干扰待测物质成分信息等的无用信息。

[0087] 这里,所述预处理,主要是用来去除上述样本光谱测量数据中由于仪器噪声、基线漂移和物质中的能导致干扰的物质成分等的无用信息,从而提高光谱质量。所述预处理方式包括小波变换、微分等所有能够去除噪声、基线和其他干扰待测物质成分信息的无用信号的方式,但并不限于小波变换、微分等处理过程;其中,

[0088] 所述小波变换过程,包括如下步骤:

[0089] 步骤 1031:选择小波基和小波分解层数,将所述光谱矩阵 X 的每条信号进行小波分解;其中光谱矩阵 X 的每一行中各元素代表一个样本在各个波长下的吸光度值;

[0090] 步骤 1032:对光谱信号在小波域内的低频段小波系数置零来实现基线校正,利用较高频段小波系数阈值处理来实现噪声去除;

[0091] 步骤 1033:用分解后去噪和基线校正后的第 N 层低频和高频系数进行信号重构,重构的各条光谱信号组成新的光谱矩阵 X_{new} 。

[0092] 所述的光谱微分处理过程,包括如下步骤:

[0093] 步骤 1031':选择微分窗口宽度 g 及微分阶次;

[0094] 步骤 1032':光谱的一阶微分算法如公式 (1) 所示,一阶微分可去除与波长无关的漂移;

$$[0095] \quad X_{\text{new}}(i, j) = [x(i, j+g) - x(i, j)] / g \quad (1)$$

[0096] 光谱的二阶微分算法如公式 (2) 所示,二阶微分可去除与波长相关的漂移。

$$[0097] \quad X_{\text{new}}(i, j) = [x(i, j+g) - 2x(i, j) + x(i, j-g)] / g^2 \quad (2)$$

[0098] 本发明实施例采用微分预处理方式对原始近红外光谱进行处理时,选择微分窗口宽度 $g = 17$,微分阶次为 2 阶,经微分预处理后得到的新的光谱 X_{new} 如图 3 所示,预处理后得到的光谱 X_{new} 的波长变量个数为 666 个,即 X_{new} 的矩阵维数为 60×666 。

[0099] 需要指出的是,本发明中,对近红外光谱进行预处理的方式并不限于上述方法,其他任何去噪和基线校正等无用信息消除的预处理手段,也均适用。

[0100] 步骤 104、对经预处理后的校正集样本的光谱数据进行基于自模型混合物分析技术的校正集样本优选。

[0101] 这里,所述基于自模型混合物分析技术的校正集样本优选方法,是指对经过预处理后的校正集样本的光谱 X_{new} 进行样本优选。

[0102] 进一步地,所述基于自模型混合物分析技术的样本优选方法,包括以下步骤:

[0103] 步骤 1041、将光谱矩阵 X_{new} 进行转置得到 X'_{new} ,则 X'_{new} 矩阵每一列各元素代表一个样本在各个波长下的吸光度值。这样,能够为样本优选自动实现的计算机程序设计带来方便。

[0104] 步骤 1042、计算光谱矩阵 X'_{new} 中各列向量 i 的纯度值 $p_{i,1}$ 。

[0105] 如图 4 所示,是在选第 1 个样本时,求得的各样本的纯度值。这里,光谱矩阵 X'_{new} 中各列向量 i 即样本 i 的纯度值 $p_{i,1}$ 的计算如公式 (3) 所示,纯度值用以表征各样本含被测物质成分含量信息量的大小,也就是说对校正模型的贡献。

$$[0106] \quad p_{i,1} = \sigma_i / (\mu_i + \alpha) \quad (3)$$

[0107] 其中, μ_i 为均值、 σ_i 为标准差、 α 为补偿因子 (一般取均值的 1 ~ 5%)。

[0108] 步骤 1043、选择第一个样本。

[0109] 根据所述步骤 1042 求得的各列向量 i 即样本 i 的 $p_{i,1}$ 值,判断 $p_{i,1}$ 值的大小,具有最大 $p_{i,1}$ 值的第 i 个样本即为选出的第一个样本。从图 4 可知,具有最大纯度值 ($p_{5,1} = 797.12$) 的校正集中的第 5 个样本作为优选出的第 1 个样本,用于后续的校正模型的建立。

[0110] 步骤 1044、选择第 k ($k \geq 2$) 个样本。

[0111] 计算矩阵 X'_{new} 中列向量 i 的长度 l_i 如公式 (4) 所示

[0112]

$$l_i = \sqrt{\sum_{j=1}^n (d_{i,j})^2 / n} \quad (4)$$

[0113] 其中, $d_{i,j}$ 为光谱矩阵 X'_{new} 中第 i 行第 j 列元素,有 $d(l)_{i,j} = d_{i,j} / \sqrt{\mu_i^2 + (\sigma_i + \alpha)^2}$, 得到关系矩阵 $C = D(1)D(1)^T/n$, 计算关系权函数 $\rho_{i,k}$ 如公式 (5) 所示。

$$[0114] \quad \rho_{i,k} = \begin{vmatrix} c_{i,i} & c_{i,p_1} & \cdots & c_{i,p_{k-1}} \\ c_{p_1,i} & c_{p_1,p_1} & \cdots & c_{p_1,p_{k-1}} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ c_{p_{k-1},i} & \cdots & \cdots & c_{p_{k-1},p_{k-1}} \end{vmatrix} \quad (5)$$

[0115] 其中, k 表示待确定的第 k 个样本, p_{k-1} 表示目前已经选定了的第 $(k-1)$ 个样本在 C 矩阵中所在列向量的标号, p_1 表示已选择的第一个样本在 C 矩阵中所在列向量的标号, 则纯度值 $p_{i,k}$ 为:

$$[0116] \quad p_{i,k} = \rho_{i,k} (\sigma_i / (\mu_i + \alpha)) \quad (6)$$

[0117] 具有最大 $p_{i,k}$ 值的第 i 个样本为选出的第 k 个样本。

[0118] 根据公式 (4)、公式 (5)、公式 (6) 可求得与已优选出的前 $k-1$ 个样本具有最小共线性的样本, 作为选出的第 i 个样本, 本实施例中, 预先选出 30 个样本, 为后面确定样本个数, 以确定最优的校正样本集做准备。

[0119] 步骤 1045、最优样本个数判定。重复上述步骤 1044, 并通过依次迭代选择的样本建立多元校正模型, 交互验证评价模型的预测性能, 具有最小预测均方根误差 (RMSEP) 时选取的样本个数即为最优的样本个数。所述 RMSEP 计算公式为:

$$[0120] \quad \text{RMSEP} = \sqrt{\frac{\sum_i (\hat{y} - y)^2}{n}} \quad (7)$$

[0121] 其中, $\hat{y}$ 为预测值, y 为标准方法测得的参考值, n 为样本个数。

[0122] 本实施例中,通过依次迭代选择的 30 个样本建立偏最小二乘多元校正模型,采用交互验证评价模型的性能,实验显示,当选取前 26 个样本建立多元校正模型时,其交互验证得到的 RMSEP (RMSEP = 0.12%) 为最小,如图 5 所示,判定最优样本个数为 26。

[0123] 步骤 1046、组成最优的校正样本集。用优选出的一定个数的样本组成新的最优的校正样本集,其光谱矩阵为 X_{opt} ,为了符合一般的习惯,仍然将光谱矩阵 X_{opt} 每一行元素表示为一个样本在各个波长下的吸光度值。

[0124] 本实施例中,将选取的前 26 个样本组合为最优的校正样本集用于建立最终的多元校正模型,此时得到的光谱矩阵 X_{opt} 的维数变为了 26×666 。

[0125] 步骤 105、用优选出的校正样本集样本的光谱数据和与之对应的标准值,建立基于核函数变换的非线性偏最小二乘校正模型。

[0126] 这里,所述的基于核函数变换的非线性偏最小二乘校正模型的建立,是指将经过光谱预处理和样本优选后得到的优化的校正样本集的光谱矩阵 X_{opt} ,与步骤 102 中采用的标准分析方法测定的校正样本集中样本待测组分的化学含量的真值建立关联关系,以用于预测未知样本的待测成分含量。

[0127] 所述的基于核函数变换的非线性偏最小二乘校正模型的建立,包括以下步骤:

[0128] 步骤 1051、确定初始参数,包括核函数及其参数、主成分个数。常用的核函数有协方差核、多项式核和高斯核等,其表达式分别如公式 (8)、公式 (9)、公式 (10) 所示:

$$[0129] \quad k(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j \quad (8)$$

$$[0130] \quad k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + p)^q \quad (9)$$

$$[0131] \quad k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||^2 / (2 \times \sigma^2)) \quad (10)$$

[0132] 以上三式中的 i 、 j 分别表示第 i 、 j 个光谱向量; p 、 q 、 σ 分别为需根据实际情况确定的核参数。

[0133] 本实施例中,确定初始参数,选用高斯核,核函数 $\sigma = 1$ 、最大主成分个数为 20。

[0134] 步骤 1052、对校正集的自变量进行核函数变换,其变换矩阵如公式 (11) 所示,然后进行中心化处理,其表达式如公式 (12) 所示。

$$[0135] \quad \mathbf{K}_{train} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix} \quad (11)$$

$$[0136] \quad \tilde{\mathbf{K}}_{train} = (\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T) \mathbf{K}_{train} (\mathbf{I} - \frac{1}{n} \mathbf{I}_n \mathbf{I}_n^T) \quad (12)$$

[0137] 以上二式中的 n 表示校正集样本的个数。 $k(\cdot, \cdot)$ 表示两个向量的内积; $\mathbf{x}_i$ ($i = 1, \dots, n$) 表示校正集样本的光谱; $\mathbf{I}$ 表示单位矩阵; $\mathbf{1}_n$ 表示各元素都为 1 的 n 维向量。

[0138] 在实施例中,经核函数变换处理后的矩阵如图 6 所示,经核函数变换后得到的最终用来建立校正模型的光谱矩阵的维数降为了 26×26 ,与最初的原始光谱 X 的维数 60×700 相比,明显有简化光谱模型,降低模型复杂性,有助于提高模型训练速度和降低谷物中蛋白质含量近红外光谱无损检测的硬件实现难度。同时由于采用了高斯核变换,该模型也包含了非线性信息。

[0139] 步骤 1053、对经核函数变换后得到的矩阵 $\tilde{\mathbf{K}}_{train}$,建立偏最小二乘校正模型,并根据

交互验证方法确定最优的核函数、核参数以及主成分个数。如公式 (13) 所示为最终建立的基于核函数变换的非线性偏最小二乘校正模型,

$$[0140] \quad \hat{y} = \tilde{K}_{train} b + f \quad (13)$$

[0141] 其中, b 为 $\tilde{K}_{train}$ 矩阵原始变量的偏最小二乘回归系数向量, f 为残差向量。在本实施例中, 对经核函数变换后得到的矩阵 $\tilde{K}_{train}$, 建立偏最小二乘校正模型, 并根据交互验证方法选定最优的核函数为高斯核、核参数 $\sigma = 1$ 以及主成分个数为 19。采用优选出的校正集样本建立的基于核函数变换的非线性偏最小二乘校正模型的回归系数曲线如图 7 所示。

[0142] 步骤 106、用所构造的基于核函数变换的非线性偏最小二乘校正模型, 来检测未知样本的物质成分含量。

[0143] 这里, 所述的检测未知样本的物质成分含量, 是指采用光谱仪采集未知样本的光谱, 将该光谱进行和校正集样本相同的预处理, 通过已建立的校正模型, 得到该未知样本的待测成分含量的预测值。

[0144] 进一步地, 所述未知样本的物质成分含量的预测, 包括如下步骤:

[0145] 步骤 1061、用光谱仪采集未知样本的光谱。

[0146] 本实施例中, 采集 20 个未知的样本光谱。

[0147] 步骤 1062、对于未知样本的光谱, 采用与校正集样本相同的预处理方法进行光谱处理, 去除无用信息。

[0148] 本实施例中, 采用微分预处理方法对未知样本的原始近红外光谱进行处理, 选择微分窗口宽度 $g = 17$, 微分阶次为 2 阶。

[0149] 步骤 1063、对于经预处理后的未知样本的光谱进行核函数变换, 核函数及其参数的选择与步骤 105 中所选的一致, 其变换矩阵如公式 (14) 所示, 然后进行中心化处理, 其表达式如公式 (15) 所示。

$$[0150] \quad K_{test} = \begin{bmatrix} k(xt_1, x_1) & k(xt_1, x_2) & \dots & k(xt_1, x_n) \\ k(xt_2, x_1) & k(xt_2, x_2) & \dots & k(xt_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(xt_l, x_1) & k(xt_l, x_2) & \dots & k(xt_l, x_n) \end{bmatrix} \quad (14)$$

$$[0151] \quad \tilde{K}_{test} = (K_{test} - \frac{1}{n} I_l I_n^T K_{train}) (I - \frac{1}{n} I_n I_n^T) \quad (15)$$

[0152] 其中, xt_i ($i = 1, \dots, n$) 表示未知的样本的光谱; x_i ($i = 1, \dots, n$) 表示校正集样本的光谱。

[0153] 本实施例中, 对于经预处理后的未知样本的光谱进行核函数变换, 核函数及其参数的选择与步骤 1053 中确定的一致, 即选择核函数为高斯核、核参数 $\sigma = 1$, 则其核变换后的矩阵如图 8 所示。

[0154] 步骤 1064、获得未知样本的物质成分含量的检测值, 其计算的表达式如公式 (16) 所示:

$$[0155] \quad \hat{y}_{test} = \tilde{K}_{test} b \quad (16)$$

[0156] 本实施例中, 获得未知样本的物质成分含量的检测值, 按公式 (16) 求得该 20 个未知样本的蛋白质含量, 采用本发明的物质成分含量近红外光谱无损检测方法, 得到的检测值与采用标准方法测得的参考值的相关性如图 9 所示。

[0157] 另外,为与一般的近红外光谱无损检测方法进行比较,表一给出了不同方法建立的模型对未知样本进行预测的参数,结果表明,本发明提出的物质成分含量近红外光谱无损检测方法的检测精度明显优于一般方法,且本发明提出的物质成分含量近红外光谱无损检测方法简单、快速,校正集样本优选和核函数变换过程均能降低模型的复杂性,便于硬件的实现,且通过选择适当的预处理方法及其参数、核函数及其参数,该物质成分含量的近红外光谱无损检测方法还可适用于人体生化参数如血糖、植物生化参数如叶片水分、苹果糖度等的以及其他生物物质成分含量的无损检测,并可显著提高检测结果的精度。

[0158] 表一 :选用不同光谱分析方法的预测参数

实验	校正集 样本选择方法	多元校正模型	样本个数	RMSEP(%)	相关系数
[0159] 谷物 近红外光谱	无	偏最小二乘	80 个	0.147	0.968
	自模型混合物分析	核函数变换非 线性偏最小二乘	26 个	0.076	0.989

[0160] 本发明还提供了所述的物质成分含量的近红外无损检测装置,如图 10 所示。因现有近红外光谱无损检测装置采用的预测模型通常是离线训练的校正模型,模型复杂,且通过硬件中写入的校正模型通常是固定不可调的,这对于成分特别复杂的被测对象的物质成分含量的检测的适用性不佳。而采用本发明提供的装置,可自动的实现校正样本优选和非线性校正模型的训练过程,参数设置不复杂、模型简单,具有较高的无损检测速度和准确度。且本发明提供的装置还可根据实际被测对象情况调整预处理方法及其参数、以及调节校正模型及其参数,能提高物质成分含量检测结果的精度和适用性。

[0161] 该装置主要包括光源、近红外光谱仪、微处理器和检测结果显示屏四个模块,由光源发出的近红外光照射在被测物体上,由近红外光谱仪获取光源经被测物体后反射或者透射后的吸光度数据,通过微处理器处理获取的该光谱吸光度数据、建立并保存校正模型、计算未知样本的被测物质成分含量的检测值,并由检测结果显示屏显示该检测值。其中微处理器模块包括 :校正集样本优选单元 100、模型建立单元 200 和预测值获取单元 300。其中,
[0162] 校正集样本优选单元 100,用于获取一定数量的用于建立基于核函数变换的非线性偏最小二乘校正模型的最优的校正集样本。所述校正集样本优选单元 100 进一步包括 :校正集样本数据获取子单元 110、校正集样本数据预处理子单元 120 和最优校正集样本获取子单元 130。

[0163] 所述校正集样本数据获取子单元 110,用于获取大量样本的近红外连续光谱数据,以及与之相应的采用标准分析方法测得的校正样本集中样本待测组分的化学含量的真值,构成校正集样本并提供给校正集样本数据预处理子单元 120 ;校正集样本数据预处理子单元 120,连接校正集样本数据获取子单元 110,用于对校正集样本数据获取子单元 110 获取的校正集样本数据进行预处理,包括预处理方法及其参数的选择,将预处理后得到的光谱数据提供给最优校正集样本获取子单元 130,同时还将选择的预处理方法及其参数提供给未知样本光谱数据预处理子单元 320 ;最优校正集样本获取子单元 130,连接校正集样本数据预处理子单元 120,用于根据校正集样本数据预处理子单元 120 获取的预处理后的校正集样本数据和校正集样本数据获取子单元 110 中获取的相应的真值,确定一定数量的最优校正集样本并提供给模型建立单元 200 和预测值单元 300。

[0164] 模型建立单元 200, 连接校正集样本优选单元 100, 用于利用校正集样本优选单元 100 获得的最优的校正集样本建立基于核函数变换的非线性偏最小二乘校正模型, 并将该模型提供给预测值获取单元 300。所述模型建立单元 200 进一步包括: 校正集样本核函数变换子单元 210 和基于核函数变换的非线性偏最小二乘校正模型获取子单元 220。其中,

[0165] 所述校正集样本核函数变换子单元 210, 用于对最优校正集样本获取子单元 130 获得的最优校正集样本进行核函数变换, 包括核函数及其参数的选择, 将获得的校正集样本的核函数变换后的矩阵提供给基于核函数变换的非线性偏最小二乘校正模型获取子单元 220, 同时将选择的核函数及其参数提供给未知样本核函数变换子单元 330。

[0166] 所述基于核函数变换的非线性偏最小二乘校正模型获取子单元 220, 连接校正集样本核函数变换子单元 210, 用于根据校正集样本核函数变换子单元 210 获取的核函数变换后得到的矩阵, 建立非线性偏最小二乘校正模型, 并将获得的回归系数提供给预测值获取单元 300。

[0167] 预测值获取单元 300, 连接校正集样本优选单元 100 和模型建立单元 200, 利用校正集样本优选单元 100 得到的最优的校正集样本和模型建立单元 200 得到的模型参数, 获取未知样本的物质成分含量近红外光谱无损检测的结果。所述预测值获取单元 300 进一步包括: 未知样本光谱数据获取子单元 310、未知样本光谱数据预处理子单元 320、未知样本核函数变换子单元 330 和预测值计算子单元 340。其中,

[0168] 所述未知样本光谱数据获取子单元 310, 用于获取待测的未知成分含量的样本的近红外波段范围的连续光谱, 并将该光谱数据提供给未知样本光谱数据预处理子单元 320。

[0169] 所述未知样本光谱数据预处理子单元 320, 连接校正集样本数据预处理子单元 120 和未知样本光谱数据获取子单元 310, 用于根据校正集样本数据预处理子单元 120 提供的所选择的预处理方法及其参数, 对未知样本光谱数据获取子单元 310 获取的未知样本光谱数据进行预处理。

[0170] 所述未知样本核函数变换子单元 330, 连接校正集样本核函数变换子单元 210、最优校正集样本获取子单元 130 和未知样本光谱数据预处理子单元 320, 用于根据校正集样本核函数变换子单元 210 提供的所选的核函数及其参数, 并利用最优校正集样本获取子单元 130 提供的最优校正集样本, 对未知样本光谱数据预处理子单元 320 提供的预处理后的未知样本光谱数据进行核函数变换, 将获得的核函数变换后的矩阵提供给预测值计算子单元 340。

[0171] 所述预测值计算子单元 340, 连接未知样本核函数变换子单元 330 和基于核函数变换的非线性偏最小二乘校正模型获取子单元 220, 用于利用基于核函数变换的非线性偏最小二乘校正模型获取子单元 220 获得的回归系数和未知样本核函数变换子单元 330 获得的待测的未知样本核函数变换后的矩阵, 计算得到未知样本的待测成分含量的检测值。

[0172] 以上所述, 仅为本发明的较佳实施例而已, 并非用于限定本发明的保护范围。

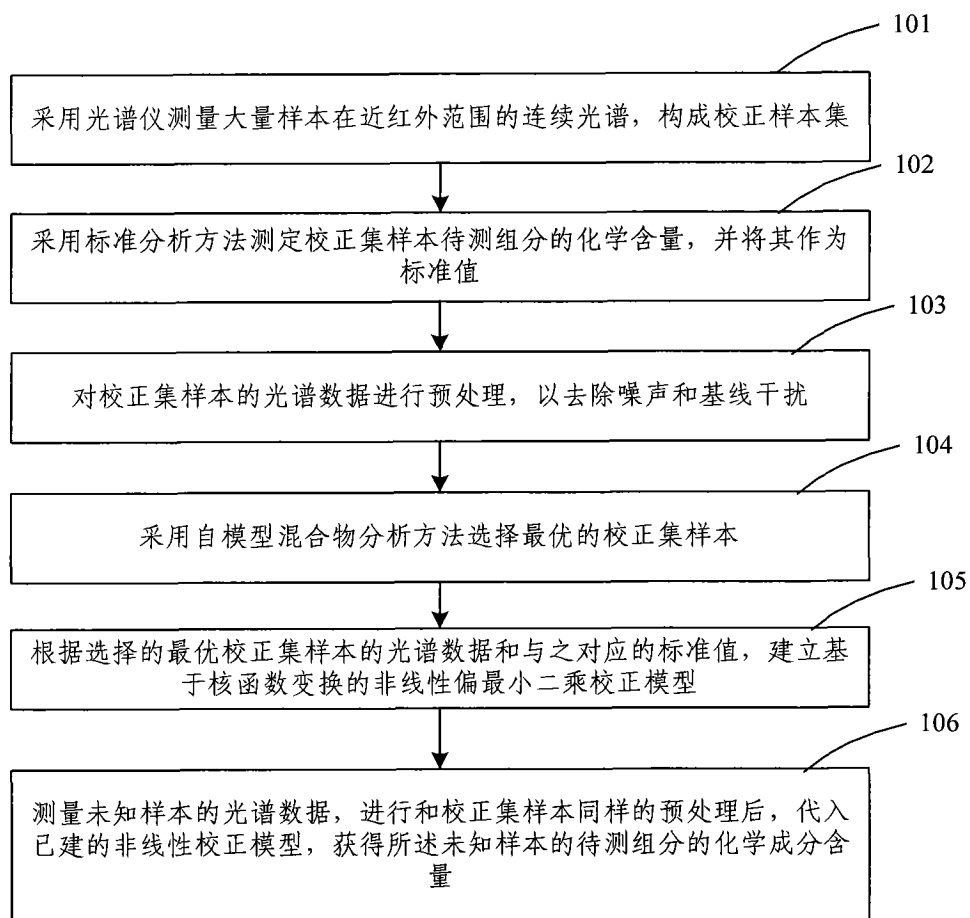


图 1

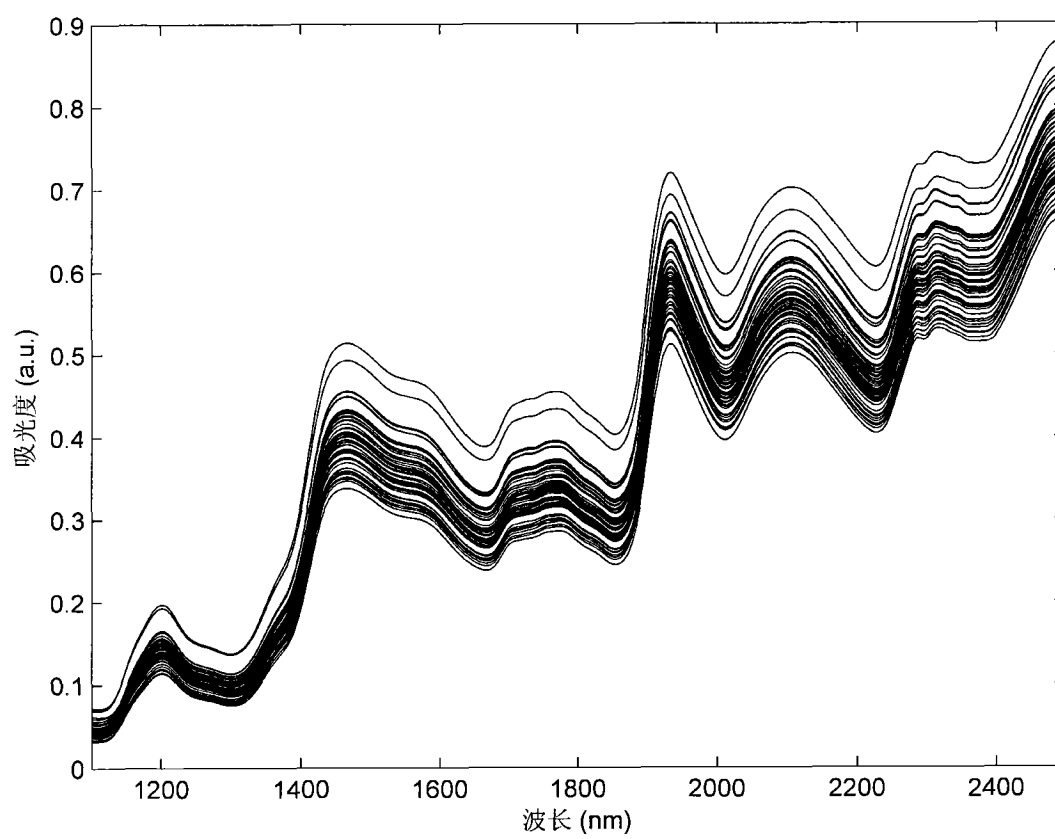


图 2

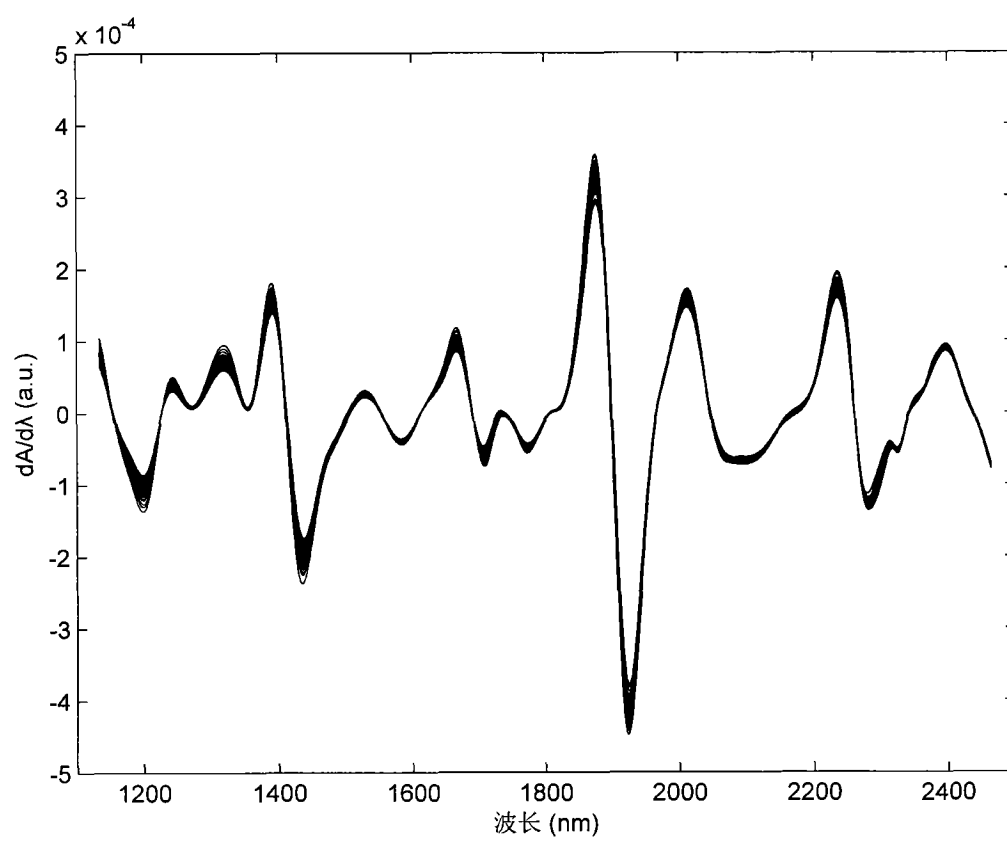


图 3

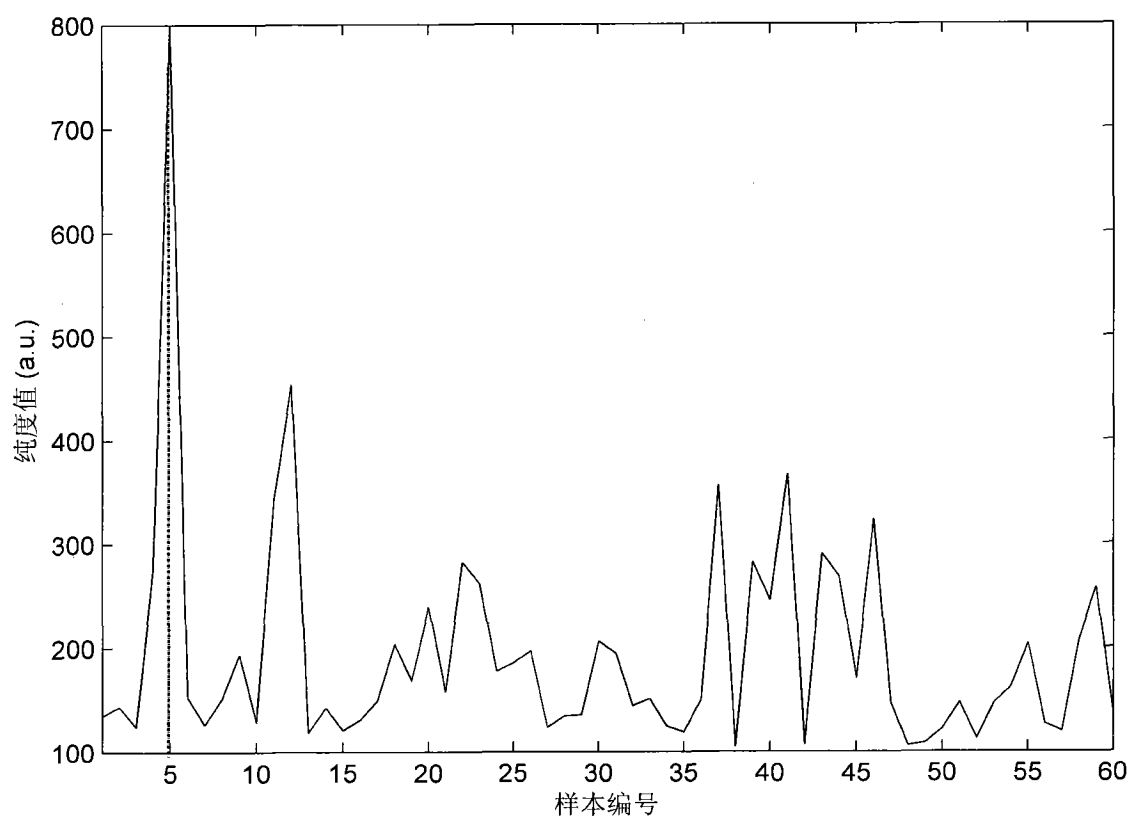


图 4

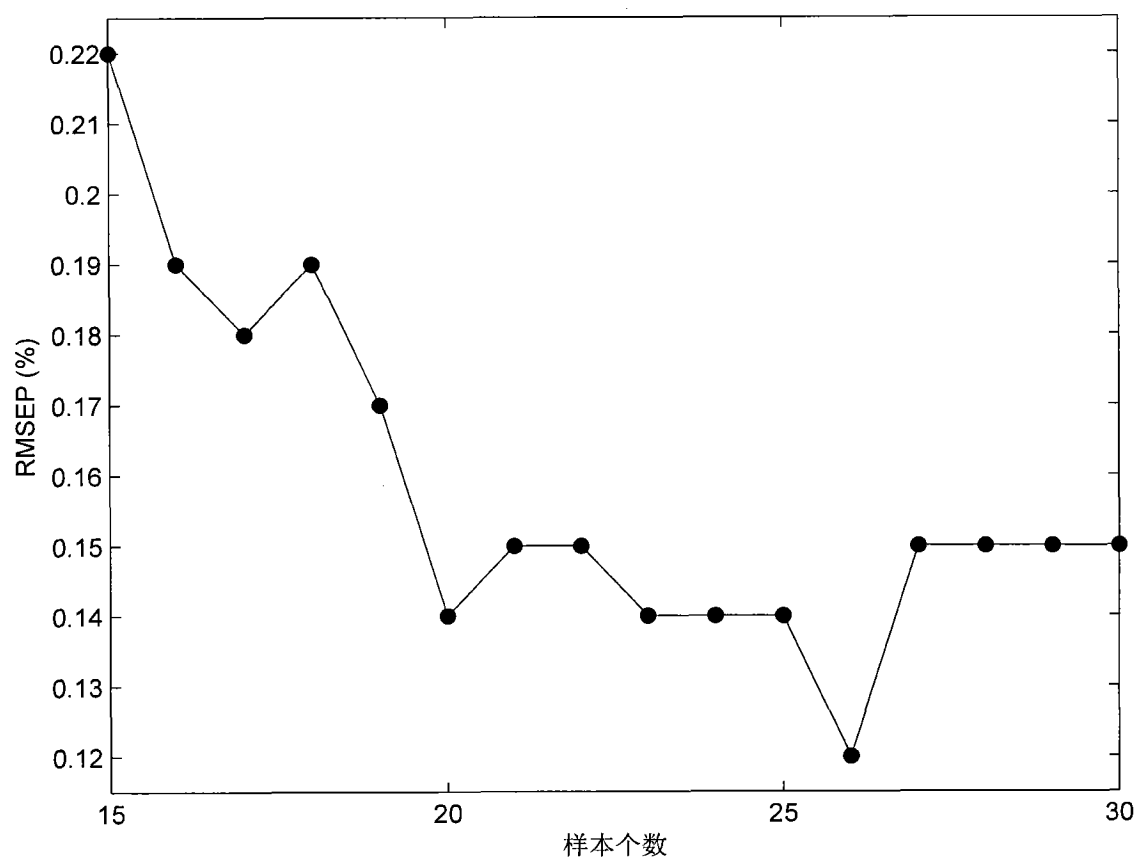


图 5

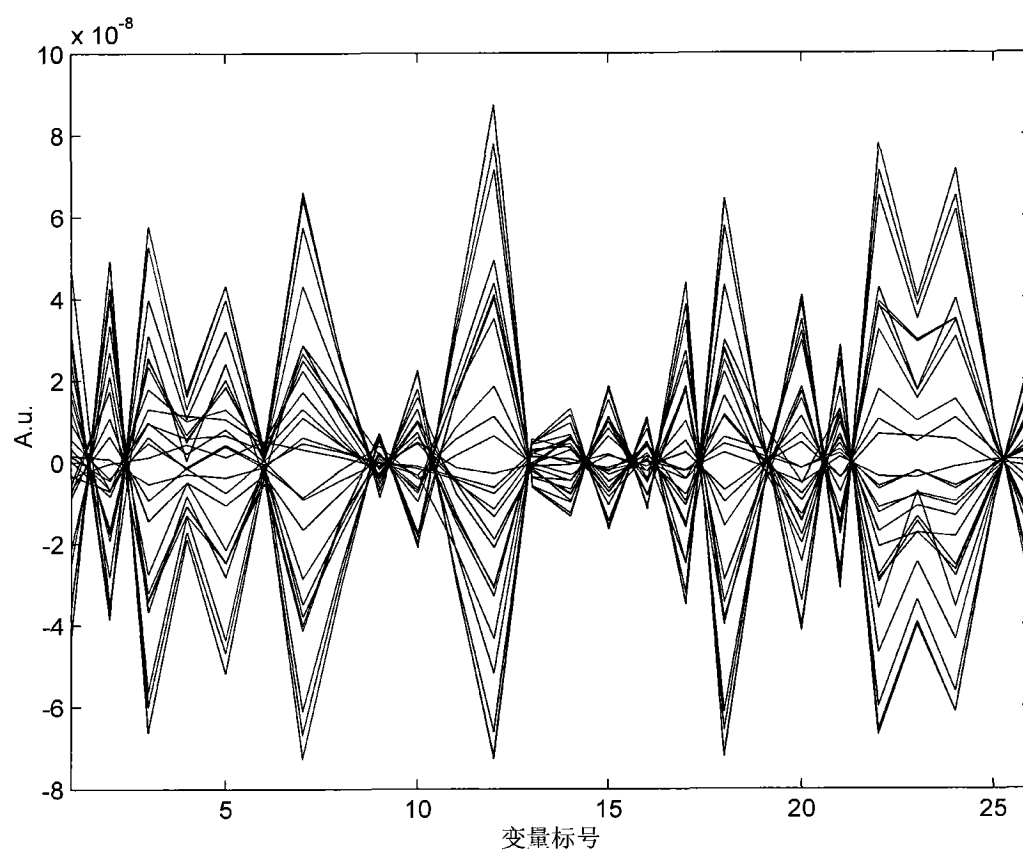


图 6

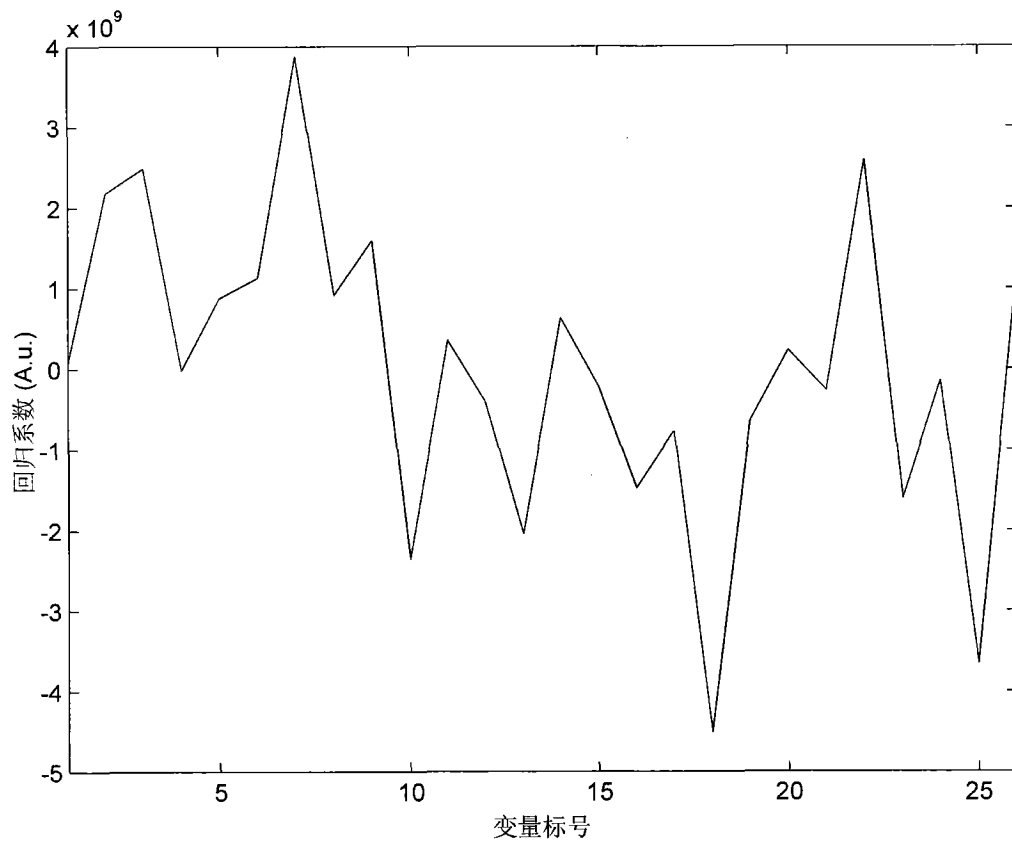


图 7

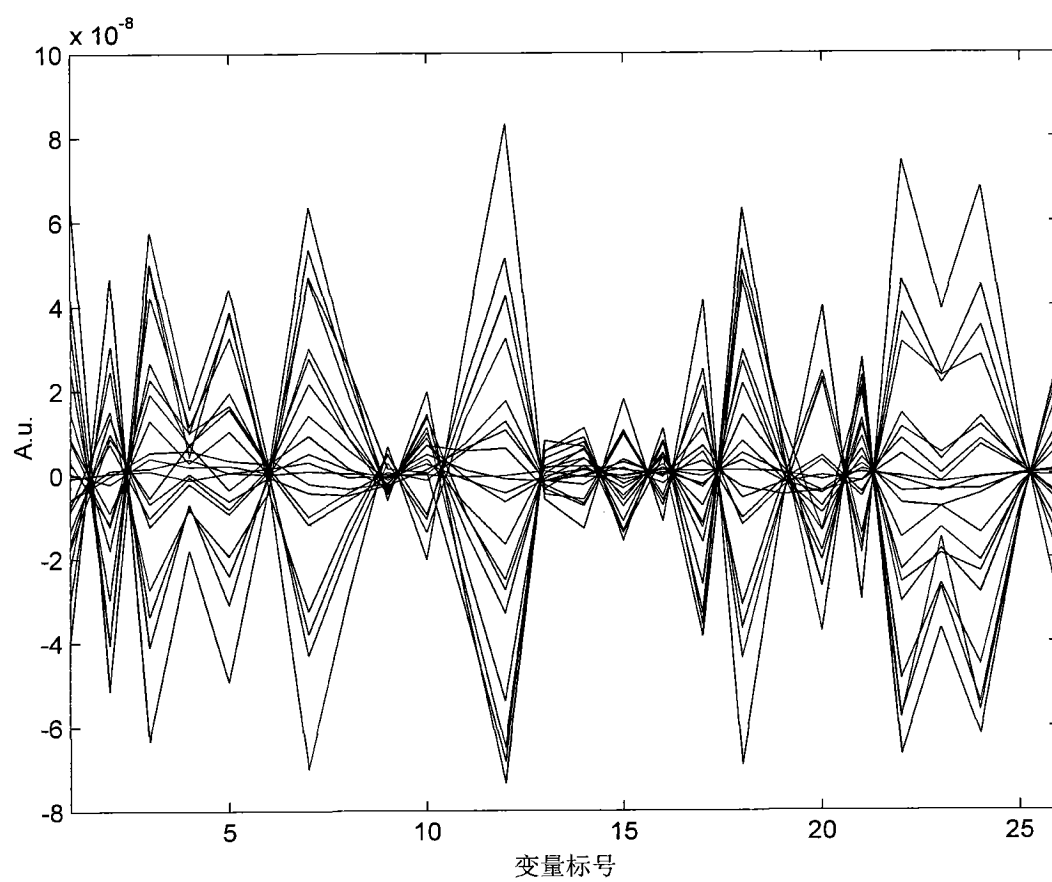


图 8

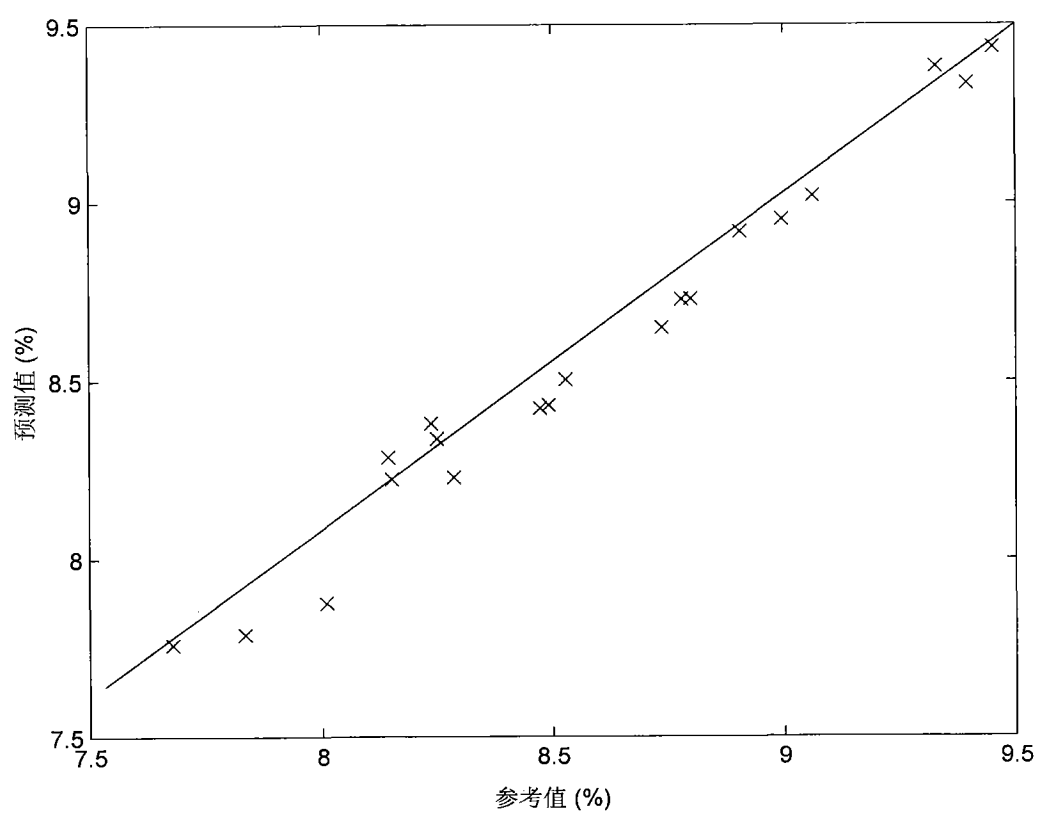


图 9

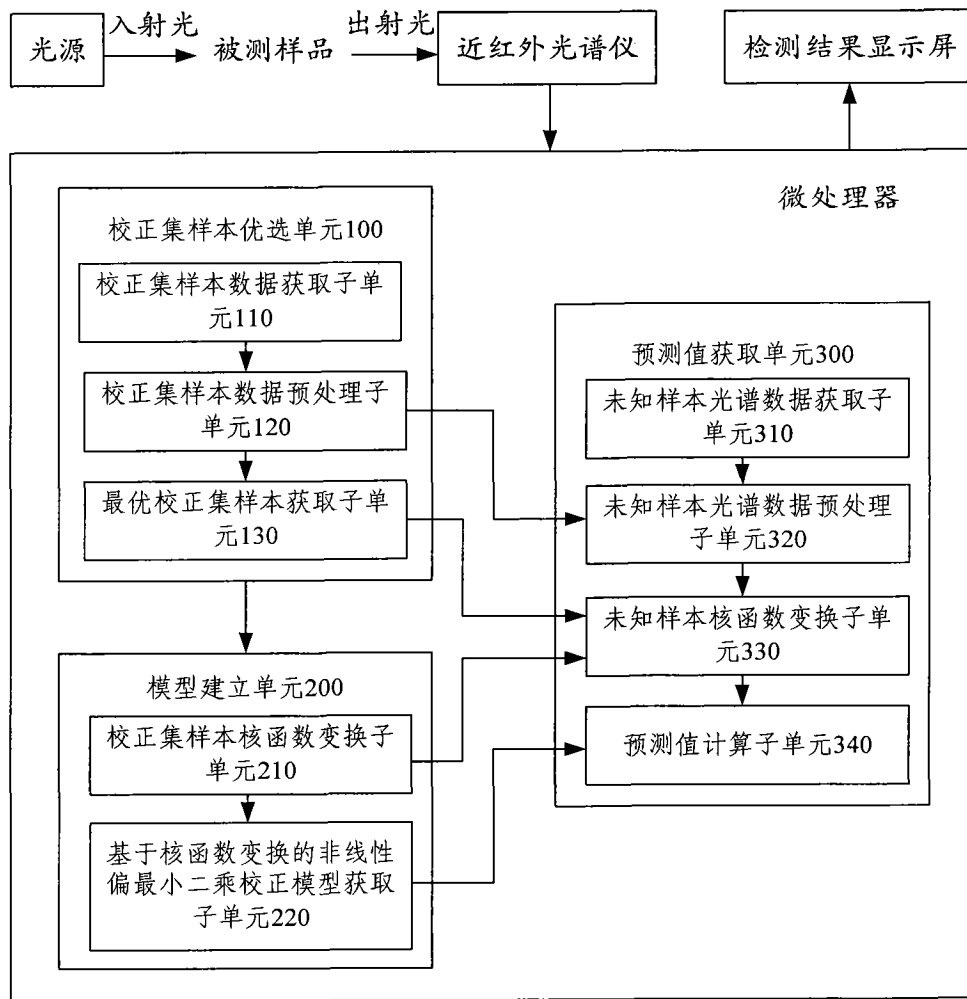


图 10