



# (12) 发明专利申请

(10) 申请公布号 CN 104091089 A

(43) 申请公布日 2014. 10. 08

(21) 申请号 201410362602. 1

(22) 申请日 2014. 07. 28

(71) 申请人 温州大学

地址 325035 浙江省温州市高教园区

(72) 发明人 陈孝敬

(74) 专利代理机构 长沙正奇专利事务所有限责任

公司 43113

代理人 马强

(51) Int. Cl.

G06F 19/00 (2011. 01)

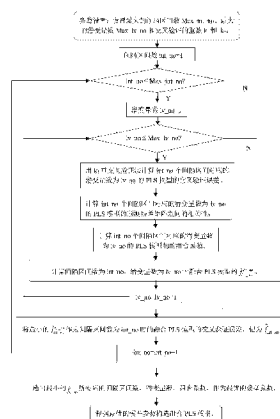
权利要求书2页 说明书6页 附图1页

## (54) 发明名称

一种红外光谱数据 PLS 建模方法

## (57) 摘要

本发明公开了一种红外光谱数据 PLS 建模方法,结合各个间隔区间的 PLS 模型的误差和误差之间的相关性来确定各个间隔区间的 PLS 模型的权系数,从而能使所得的融合 PLS 模型具有最小的误差。本发明的方法可以最好的利用各个间隔区间的光谱信息,简便、可视化、运算量小,可以很快的找到特征波长区间;本方明中的权系数的确定方法由于同时考虑到了各个参与融合的模型的误差以及误差之间的相关性,能保证融合之后的模型具有最小的误差。



1. 一种红外光谱数据 PLS 建模方法,其特征在于,包括以下步骤:

1) 设置最大间隔区间数 max\_int\_no、最大潜变量数 max\_lv\_no、交叉法的重数  $k_1$  和  $k_2$ ; 其中,  $k_1$ 、 $k_2$  均不小于 2;

2) 按照步骤 2.1) 和步骤 2.2) 计算间隔区间数为 int\_no 时,对应的融合 PLS 模型的交叉验证误差,其中  $1 \leq \text{int\_no} \leq \text{max\_int\_no}$ ;

2.1) 将红外光谱样本集数据中的光谱矩阵 X 平均分为 int\_no 个间隔区间  $X_i$ ; 每个间隔区间的列数  $l = \left\lceil \frac{\text{光谱的变量长度}}{\text{int\_no}} \right\rceil$ ,  $\lceil \cdot \rceil$  表示取整; 第 i 个间隔区间  $X_i$  对应光谱矩阵 X 的第  $[(i-1) \times l + 1] \sim (i \times l)$  列的数据;  $1 \leq i \leq \text{int\_no}$ ;

2.2) 按照步骤 2.2.1) ~ 步骤 2.2.5) 计算潜变量数为 lv\_no 时,融合 PLS 模型的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$ , 其中  $1 \leq \text{lv\_no} \leq \text{max\_lv\_no}$ ;

2.2.1) 用  $k_1$  重交叉法计算间隔数为 int\_no, 潜变量数为 lv\_no 时,各个间隔区间对应的 PLS 模型的交叉验证误差  $S(e_i) = \sqrt{\frac{1}{n} \|e_i\|^2}$ , 其中  $e_i = y - \hat{y}_i$ ; y 表示红外光谱样本集数据中的因变量矩阵的实际值,  $\hat{y}_i$  表示第 i 个间隔区间对应的潜变量数为 lv\_no 的 PLS 模型根据  $k_1$  重交叉法得到的因变量矩阵的预测值,  $e_i$  是相应的预测残差矩阵, n 是红外光谱样本集数据的样品数;

2.2.2) 计算间隔数为 int\_no, 潜变量数为 lv\_no 时,各个间隔区间对应的 PLS 模型的预测残差矩阵之间的相关性  $r_{ij} = \frac{\text{cov}(e_i, e_j)}{S(e_i)S(e_j)}$ ,  $i, j = 1, 2, \dots, \text{int\_no}$ ; 其中,

$$\text{cov}(e_i, e_j) = \frac{1}{n} \langle e_i, e_j \rangle, i, j = 1, 2, \dots, \text{int\_no};$$

2.2.3) 通过非线性优化的方法计算下式:

$$f = \min \left( \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_i) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{p>i}^{\text{int\_no}} \omega_i \omega_p r_{ip} S(e_i) S(e_p) \right)$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{\text{int\_no}} \omega_i = 1 \\ 0 \leq \omega_i \leq 1 \end{cases};$$

得到间隔数为 int\_no, 潜变量数为 lv\_no 时,各个间隔区间对应的 PLS 模型的组合系数  $\omega = [\omega_1, \dots, \omega_{\text{int\_no}}]'$ ;

2.2.4) 用  $k_2$  重交叉法计算间隔数为 int\_no, 潜变量数为 lv\_no 时,各个间隔区间对应的 PLS 模型的预测残差矩阵  $e_{2i} = y - \hat{y}_{2i}$ , 其中  $\hat{y}_{2i}$  表示第 i 个间隔区间对应的潜变量数为 lv\_no 的 PLS 模型根据  $k_2$  重交叉法得到的因变量矩阵的预测值, 计算  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$ ;

$$\hat{f}_{\text{int\_no}}^{\text{lv\_no}} = \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_{2i}) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{p>i}^{\text{int\_no}} \omega_i \omega_p r_{ip} S(e_{2i}) S(e_{2p});$$

2.2.5) 选出最小的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$  作为间隔区间数为 int\_no 时的融合 PLS 模型的交叉验证误差, 记为  $\hat{f}_{\text{int\_no}}$ ;

3) 选出所有间隔区间数下最小的  $\hat{f}_{\text{int\_no}}$ , 该最小的  $\hat{f}_{\text{int\_no}}$  对应的间隔区间数 int\_bt、潜变量数 lv\_bt 和组合系数  $\omega\_bt$  作为最优的模型参数;

4) 根据最优的模型参数构造融合 PLS 模型: 将光谱矩阵 X 平均分为 int\_bt 个间隔区间, 融合 PLS 模型如下:

$$y^* = \sum_{g=1}^{\text{int\_bt}} \omega\_bt_g (x_g \times b_g + c_g)$$

其中,  $\omega\_bt_g$  是  $\omega\_bt$  的第 g 个分量,  $y^*$  是融合 PLS 模型对样品的因变量的预测值;  $b_g$ 、 $c_g$  分别是间隔区间  $X_g$  和因变量矩阵 Y 对应潜变量数为 lv\_bt 时的偏最小回归系数和截距;  $x_g$  是第 g 个间隔区间对应的红外光谱数据。

## 一种红外光谱数据 PLS 建模方法

### 技术领域

[0001] 本发明属于红外光谱识别领域,具体是一种能提升红外光谱偏最小二乘建模效果的数据处理方法。

### 背景技术

[0002] 在小样本多变量的红外光谱数据中,PLS 模型可以很好的解决其它建模方法所遇到的变量共线性和维数灾难,因此在红外光谱识别中得到了广泛的用。虽然 PLS 可以直接对全光谱建模,但是理论和大量的实验证明波长选择仍是一种有效的提高 PLS 模型的方法。波长优化选择是指通过一定的方法在建模之前进行特征波长或波段的筛选。经波长选择后所建模型由于剔除了不相关或非线性变量,因此较全波长模型更为简化,预测能力和稳健性也更好。其中 iPLS(interval PLS-iPLS) 是一种常用的波长选择方法。iPLS 方法的优点是简便、可视化、运算量小,可以很快的找到特征波长区间。缺点是只利用一个间隔区间的光谱信息,可能丢失其他间隔区间的有用的光谱信息。因此如何最好的利用各个间隔区间的光谱信息是亟待解决的问题。

### 发明内容

[0003] 本发明所要解决的技术问题是,针对上述现有技术的不足,提供一种红外光谱数据 PLS 建模方法。

[0004] 为解决上述技术问题,本发明所采用的技术方案是:一种红外光谱数据 PLS 建模方法,包括以下步骤:

[0005] 1) 设置最大间隔区间数 max\_int\_no、最大潜变量数 max\_lv\_no、交叉法的重数  $k_1$  和  $k_2$ ;其中,  $k_1$ 、 $k_2$  均不小于 2;

[0006] 2) 计算间隔区间数为 int\_no 时,对应的融合 PLS 模型的交叉验证误差,计算的步骤都是 2.1 至 2.2,其中  $1 \leq \text{int\_no} \leq \text{max\_int\_no}$ ;

[0007] 2.1) 将红外光谱样本集数据中的光谱矩阵  $X$  平均分为 int\_no 个间隔区间  $X_i$ :每个

间隔区间的列数  $l = \left\lfloor \frac{\text{光谱的变量长度}}{\text{int\_no}} \right\rfloor$ ,  $\lfloor \cdot \rfloor$  表示取整;第  $i$  个间隔区间  $X_i$  对应光谱矩阵

$X$  的第  $[(i-1) \times l + 1] \sim (i \times l)$  列的数据;  $1 \leq i \leq \text{int\_no}$ ;

[0008] 2.2) 计算潜变量数为 lv\_no 时,融合 PLS 模型的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$ ,其中  $1 \leq \text{lv\_no} \leq \text{max\_lv\_no}$ ,计算的步骤都是 2.2.1 至 2.2.5;

[0009] 2.2.1) 用  $k_1$  重交叉法计算间隔数为 int\_no,潜变量数为 lv\_no 时,各个间隔区间

对应的 PLS 模型的交叉验证误差  $S(e_i) = \sqrt{\frac{1}{n} \|e_i\|^2}$ ,其中  $e_i = y - \hat{y}_i$ ;  $y$  表示红外光谱样本集数

据中的因变量矩阵的实际值,  $\hat{y}_i$  表示第  $i$  个间隔区间对应的潜变量数为 lv\_no 的 PLS 模型根据  $k_1$  重交叉法得到的因变量矩阵的预测值,  $e_i$  是相应的预测残差矩阵,  $n$  是红外光谱样

本集数据的样品数；

[0010] 2.2.2) 计算间隔数为  $\text{int\_no}$ , 潜变量数为  $\text{lv\_no}$  时, 各个间隔区间对应的 PLS 模型的预测残差矩阵之间的相关性  $r_{ij} = \frac{\text{cov}(e_i, e_j)}{S(e_i)S(e_j)}, i, j = 1, 2, \dots, \text{int\_no}$ ; 其中,

$$\text{cov}(e_i, e_j) = \frac{1}{n} \langle e_i, e_j \rangle, i, j = 1, 2, \dots, \text{int\_no};$$

[0011] 2.2.3) 通过非线性优化的方法计算下式,

$$[0012] \quad f = \min \left( \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_i) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{p>i}^{\text{int\_no}} \omega_i \omega_p r_{ip} S(e_i) S(e_p) \right)$$

$$[0013] \quad s.t. \begin{cases} \sum_{i=1}^{\text{int\_no}} \omega_i = 1 \\ 0 \leq \omega_i \leq 1 \end{cases};$$

[0014] 得到间隔数为  $\text{int\_no}$ , 潜变量数为  $\text{lv\_no}$  时, 各个间隔区间对应的 PLS 模型的组合系数  $\omega = [\omega_1, \dots, \omega_{\text{int\_no}}]'$ ;

[0015] 2.2.4) 用  $k_2$  重交叉法计算间隔数为  $\text{int\_no}$ , 潜变量数为  $\text{lv\_no}$  时, 各个间隔区间对应的 PLS 模型的预测残差矩阵  $e_{2i} = y - \hat{y}_{2i}$ , 其中  $\hat{y}_{2i}$  表示第  $i$  个间隔区间对应的潜变量数为  $\text{lv\_no}$  的 PLS 模型根据  $k_2$  重交叉法得到的因变量矩阵的预测值, 计算  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$ ,

$$[0016] \quad \hat{f}_{\text{int\_no}}^{\text{lv\_no}} = \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_{2i}) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{p>i}^{\text{int\_no}} \omega_i \omega_p r_{ip} S(e_{2i}) S(e_{2p});$$

[0017] 2.2.5) 选出最小的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$  作为间隔区间数为  $\text{int\_no}$  时的融合 PLS 模型的交叉验证误差, 记为  $\hat{f}_{\text{int\_no}}$ ;

[0018] 3) 选出所有间隔区间数下最小的  $\hat{f}_{\text{int\_no}}$ , 该最小的  $\hat{f}_{\text{int\_no}}$  对应的间隔区间数  $\text{int\_bt}$ 、潜变量数  $\text{lv\_bt}$  和组合系数  $\omega_{\text{bt}}$  作为最优的模型参数;

[0019] 4) 根据最优的模型参数构造融合 PLS 模型: 将光谱矩阵  $X$  平均分为  $\text{int\_bt}$  个间隔区间, 融合 PLS 模型如下:

$$[0020] \quad y^* = \sum_{g=1}^{\text{int\_bt}} \omega_{\text{bt}_g} (x_g \times b_g + c_g)$$

[0021] 其中,  $\omega_{\text{bt}_g}$  是  $\omega_{\text{bt}}$  的第  $g$  个分量,  $y^*$  是融合 PLS 模型对样品的因变量的预测值;  $b_g$ 、 $c_g$  分别是间隔区间  $X_g$  和因变量矩阵  $Y$  对应潜变量数为  $\text{lv\_bt}$  时的偏最小回归系数和截距;  $x_g$  是第  $g$  个间隔区间对应的红外光谱数据。

[0022] 本发明的融合 PLS 模型是多个成员模型的加权组合。成员模型就是各个间隔区间对应的 PLS 模型。间隔区间数对应成员模型的数量。第  $i$  个成员模型的具体形式由第  $i$  个间隔区间的光谱数据以及提取的潜变量确定。

[0023] 与现有技术相比, 本发明所具有的有益效果为: 本发明的方法可以最好的利用各

个间隔区间的光谱信息,简便、可视化、运算量小,可以很快的找到特征波长区间;本方明中的权系数的确定方法由于同时考虑到了各个参与融合的模型的误差以及误差之间的相关性,能保证融合之后的模型具有最小的误差。

## 附图说明

[0024] 图 1 为本发明的方法流程图。

## 具体实施方式

[0025] 现结合实例对本发明做进一步的描述。

[0026] 光谱数据采用 matlab2012a 自带的 spectra 光谱数据,样品是汽油,因变量是样品的辛烷值。原始的样品数据集包含 60 个样品,每个样品的光谱变量长度是 700。为了便于描述,本例只选用第 1-6 个样本的第 1-6 个光谱变量数据作为样品集的光谱数据矩阵。本例采用的样品集数据由光谱数据矩阵 X 和因变量矩阵 Y 构成,分别如下所示,

$$[0027] \quad X = \begin{bmatrix} -0.0502 & -0.0459 & -0.0422 & -0.0372 & -0.0333 & -0.0312 \\ -0.0442 & -0.0396 & -0.0357 & -0.0309 & -0.0267 & -0.0239 \\ -0.0469 & -0.0413 & -0.0370 & -0.0315 & -0.0265 & -0.0233 \\ -0.0467 & -0.0422 & -0.0386 & -0.0345 & -0.0302 & -0.0277 \\ -0.0509 & -0.0451 & -0.0410 & -0.0364 & -0.0327 & -0.0315 \\ -0.0481 & -0.0427 & -0.0388 & -0.0340 & -0.0301 & -0.0277 \end{bmatrix}$$

$$[0028] \quad Y = \begin{bmatrix} 85.3000 \\ 85.2500 \\ 88.4500 \\ 83.4000 \\ 87.9000 \\ 85.5000 \end{bmatrix}$$

[0029] 本发明的具体实施步骤如下:

[0030] 步骤 1, 参数设置: 设置最大的间隔区间数  $\text{Max\_int\_no} = 2$ 、最大的潜变量数  $\text{Max\_lv\_no} = 2$ 、交叉法的重数  $k_1 = 4$ 、 $k_2 = 6$ 。这些参数的设置可以根据实际需要进行调整。这里这样设置参数只是为了便于说明建模步骤。

[0031] 步骤 2, 计算间隔区间数为  $\text{int\_no}$  时, 对应的融合 PLS 模型的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$ 。计算的步骤都是 2.1 至 2.2, 其中  $1 \leq \text{int\_no} \leq \text{max\_int\_no}$ , 下面以  $\text{int\_no} = 2$  为例进行说明:

[0032] 步骤 2.1 将红外光谱样本集数据中的光谱矩阵 X 平均分为  $\text{int\_no}$  个间隔区间  $X_i$ 。

每个间隔区间的列数  $l = \frac{6}{\text{int\_no}} = 3$ 。  $X_1$  对应光谱数据矩阵 X 的第一列到第三列,  $X_2$  对应光谱数据矩阵 X 的第四列到第六列。  $X_1$ 、 $X_2$  分别如下所示。

$$[0033] \quad X_1 = \begin{bmatrix} -0.0502 & -0.0459 & -0.0422 \\ -0.0442 & -0.0396 & -0.0357 \\ -0.0469 & -0.0413 & -0.0370 \\ -0.0467 & -0.0422 & -0.0386 \\ -0.0509 & -0.0451 & -0.0410 \\ -0.0481 & -0.0427 & -0.0388 \end{bmatrix}$$

$$[0034] \quad X_2 = \begin{bmatrix} -0.0372 & -0.0333 & -0.0312 \\ -0.0309 & -0.0267 & -0.0239 \\ -0.0315 & -0.0265 & -0.0233 \\ -0.0345 & -0.0302 & -0.0277 \\ -0.0364 & -0.0327 & -0.0315 \\ -0.0340 & -0.0301 & -0.0277 \end{bmatrix}$$

[0035] 步骤 2.2 计算潜变量数为  $lv\_no$  时, 融合 PLS 模型的交叉验证误差, 其中  $1 \leq lv\_no \leq \max\_lv\_no$ , 计算的步骤都是 2.2.1 至 2.2.5; 下面以  $lv\_no = 2$  为例进行说明。

[0036] 步骤 2.2.1 用  $k_1$  重交叉法计算间隔数为  $int\_no$ , 潜变量数为  $lv\_no$  时的各个间隔区间对应的 PLS 模型的交叉验证误差。

[0037] 根据  $k_1$  重交叉法计算所得的第一个间隔区间对应的潜变量数为  $lv\_no$  的 PLS 模型对因变量矩阵的预测值  $\hat{y}_1$ , 预测残差矩阵  $e_1$  以及  $e_1$  的标准差  $S(e_1)$  分别如下所示,

$$[0038] \quad \hat{y}_1 = \begin{bmatrix} 83.5924 \\ 86.7554 \\ 85.2694 \\ 82.1904 \\ 89.5037 \\ 87.4816 \end{bmatrix}, \quad e_1 = \begin{bmatrix} 1.7076 \\ -1.5054 \\ 3.1806 \\ 1.2096 \\ -1.6037 \\ -1.9816 \end{bmatrix}, \quad S(e_1) = 2.1490。$$

[0039] 根据  $k_1$  重交叉法计算所得的第二个间隔区间对应的潜变量数为  $lv\_no$  的 PLS 模型对因变量矩阵的预测值  $\hat{y}_2$ 、预测残差矩阵  $e_2$  以及  $e_2$  的标准差  $S(e_2)$  分别如下所示,

$$[0040] \quad \hat{y}_2 = \begin{bmatrix} 80.1147 \\ 86.4685 \\ 86.9897 \\ 86.9970 \\ 81.4383 \\ 84.1570 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 5.1853 \\ -1.2185 \\ 1.4603 \\ -3.5970 \\ 6.4617 \\ 1.3430 \end{bmatrix}, \quad S(e_2) = 3.7823。$$

[0041] 步骤 2.2.2 计算间隔数为  $int\_no$ , 潜变量数为  $lv\_no$  时, 各个间隔区间对应的 PLS 模型的预测残差矩阵之间的相关性,

[0042]  $r_{11} = 1.0000, r_{12} = -0.0900, r_{21} = -0.0900, r_{22} = 1.0000。$

[0043] 步骤 2.2.3 通过非线性优化的方法计算下式,

$$[0044] \quad f = \min \left( \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_i) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{j>i}^{\text{int\_no}} \omega_i \omega_j r_{ij} S(e_i) S(e_j) \right)$$

$$[0045] \quad s.t. \begin{cases} \sum_{i=1}^{\text{int\_no}} \omega_i = 1 \\ 0 \leq \omega_i \leq 1 \end{cases}$$

[0046] 得到间隔数为 int\_no, 潜变量数为 lv\_no 时, 各个间隔区间对应的 PLS 模型的组合系数

$$[0047] \quad \omega = [0.7376 \ 0.2624]'$$

[0048] 步骤 2.2.4 用  $k_2$  重交叉法计算间隔数为 int\_no, 潜变量数为 lv\_no 时, 各个间隔区间对应的 PLS 模型的预测残差矩阵  $e_{2i} = y - \hat{y}_{2i}$ ,

$$[0049] \quad \hat{y}_{21} = \begin{bmatrix} 83.5924 \\ 86.7554 \\ 85.2694 \\ 82.1904 \\ 89.5037 \\ 87.4816 \end{bmatrix}, \quad \hat{y}_{22} = \begin{bmatrix} 80.1147 \\ 86.4685 \\ 86.9897 \\ 86.9970 \\ 81.4383 \\ 84.1570 \end{bmatrix}, \quad e_{21} = \begin{bmatrix} 1.7076 \\ -1.5054 \\ 3.1806 \\ 1.2096 \\ -1.6037 \\ -1.9816 \end{bmatrix}, \quad e_{22} = \begin{bmatrix} 5.1853 \\ -1.2185 \\ 1.4603 \\ -3.5970 \\ 6.4617 \\ 1.3430 \end{bmatrix}$$

[0050]  $\hat{y}_{21}$ ,  $\hat{y}_{22}$  分别是第一个、第二个间隔区间对应的潜变量数为 lv\_no 的 PLS 模型对因变量矩阵的预测值。

$$[0051] \quad \text{计算 } \hat{f}_{\text{int\_no}}^{\text{lv\_no}} = \sum_{i=1}^{\text{int\_no}} \omega_i^2 S^2(e_{2i}) + 2 \sum_{i=1}^{\text{int\_no}} \sum_{p>i}^{\text{int\_no}} \omega_i \omega_p r_{ip} S(e_{2i}) S(e_{2p}) = 1.7250$$

[0052] 步骤 2.2.5 选出最小的  $\hat{f}_{\text{int\_no}}^{\text{lv\_no}}$  作为间隔区间数为 int\_no 时的融合 PLS 模型的交叉验证误差, 记为  $\hat{f}_{\text{int\_no}}$ 。本例中  $\hat{f}_1^1 = 2.4440$ ,  $\hat{f}_1^2 = 2.7208$ ,  $\hat{f}_2^1 = 2.1265$ ,  $\hat{f}_2^2 = 1.7250$ 。因此间隔区间数 int\_no = 1 时, 融合 PLS 模型的交叉验证误差  $\hat{f}_1 = 2.4440$ , 间隔区间数 int\_no = 2 时, 融合 PLS 模型的交叉验证误差为  $\hat{f}_2 = 1.7250$ 。

[0053] 步骤 3, 选出所有间隔区间数 int\_no ( $1 \leq \text{int\_no} \leq \text{Max\_int\_no}$ ) 情况下, 融合 PLS 模型的交叉验证误差最小值。本例中  $\hat{f}_2^2 = 1.7250$  是最小值, 相应的最优的模型参数如下: 间隔区间数 int\_bt = 2, 潜变量数 lv\_bt = 2, 组合系数  $\omega_{\text{bt}} = [0.7376 \ 0.2624]'$ 。

[0054] 步骤 4, 根据最优的模型参数构造融合 PLS 模型。  $b_1 = [64.4 \ -2120.4 \ 443.4]'$ ,  $c_1 = 1565.1$  分别是  $X_1$  和  $Y$  对应的潜变量数为 2 时的偏最小二乘回归系数和截距。  $b_2 = [105.8 \ 596.31404.7]'$ ,  $c_2 = -1544.9$  分别是  $X_2$  和  $Y$  对应的潜变量数为 2 时的偏最小二乘回归系数和截距。最终的融合 PLS 模型如下所示,

$$[0055] \quad y = 0.7376 \times (x_1 b_1 + 64.4) + 0.2624 \times (x_2 b_2 + 105.8)。$$

[0056] 一个样品的完整的光谱数据  $x$  由  $x_1$  和  $x_2$  构成, 即  $x = [x_1 \ x_2]$ 。  $x_1$  是第一个间隔区间对应的的光谱数据,  $x_2$  是第二个间隔区间对应的的光谱数据。  $y$  是融合 PLS 模型对样品



的因变量的预测值。

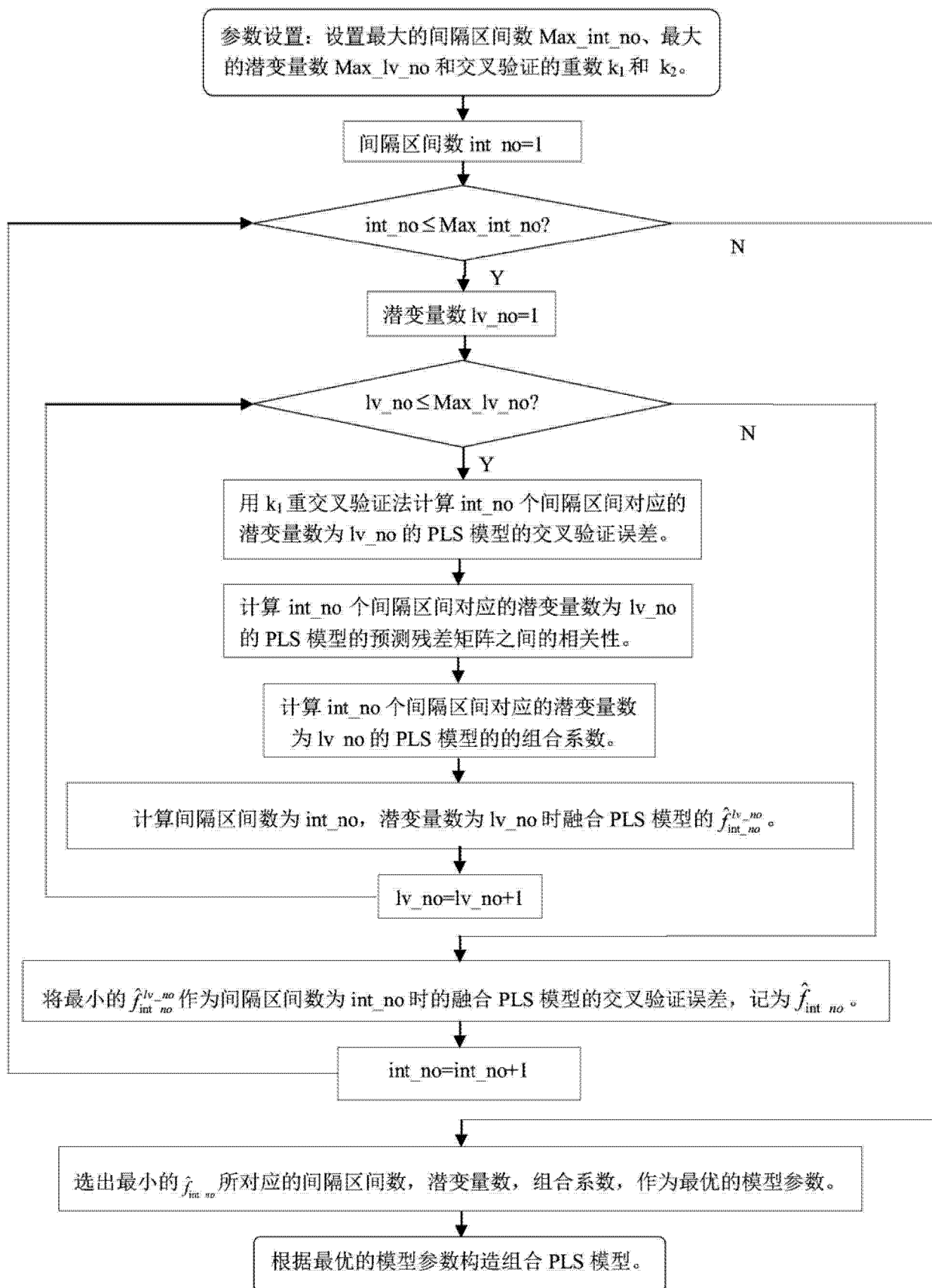


图 1