

近红外光谱定量分析的新方法: 半监督最小二乘支持向量回归机

李 林¹, 徐 硕^{2*}, 安 欣³, 张录达⁴

1. 中国农业大学信息与电气工程学院, 北京 100193
2. 中国科学技术信息研究所信息技术支持中心, 北京 100038
3. 对外经济贸易大学国际经济与贸易学院, 北京 100029
4. 中国农业大学理学院, 北京 100193

摘 要 在近红外光谱定量分析中, 样品化学值测定的准确度是运用数学模型进行定量分析精确度的理论极限。但能够准确获取化学值的样品数量比较少, 许多模型在建模时只考虑这部分样品数据, 而不考虑大量的无化学值的样品数据。针对该问题, 本文在 LS-SVR 的基础上, 提出了可以同时利用有化学值(标签)和无化学值样品数据的半监督 LS-SVR(S^2 LS-SVR)模型。类似于 LS-SVR, 该模型也只需求解一个线性方程组。最后, 以烤烟样品数据集为实验材料, 建立了四种样品成分(总糖、还原糖、总氮和烟碱)的定量分析模型。四种样品成分的预测值与实际值的平均误差分别为 6.62%, 7.56%, 6.11% 和 8.20%, 相关系数分别为 0.974 1, 0.973 3, 0.923 0 和 0.948 6。经分析比较发现 S^2 LS-SVR 模型优于 PLS 和 LS-SVR, 从而验证了 S^2 LS-SVR 模型的可行性和有效性。

关键词 近红外光谱; 化学计量学; 半监督 LS-SVR(S^2 LS-SVR)

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)10-2702-04

引 言

利用有机物在近红外光谱区的振动吸收, 近红外光谱技术^[1]几乎可以用于所有与含氢基团有关的化学性质与物理性质分析。它具有操作简单、分析速度快以及测定一次光谱可同时获得样品多种成分含量的独特优点, 使其在作物品质分析上得到了广泛应用。对于近红外光谱定量分析, 准确测定样品的化学值极其重要, 因为它们是运用数学模型进行定量分析精确度的理论极限^[1]。但与获取样品的近红外光谱数据相比, 样品化学值的准确测定代价高、效率低, 比如根据 GB/T 5009.9—2003《食品中淀粉的测定》中酸水解法测定玉米总淀粉的含量, 需要大量的化学试剂, 操作复杂。换句话说, 可以获取近红外光谱数据的样品是大量的, 而准确测定化学值的样品却是少量的。

在机器学习中, 通常将化学值称为标签。传统的监督学习方法, 比如偏最小二乘(PLS)回归^[2]、支持向量回归机

(SVR)^[3]及最小二乘 SVR(LS-SVR)^[4]等, 虽然能够解决一定的实际问题, 但它们只能利用有标签样品的信息, 而不能利用无标签样品的信息。为了解决这类问题, Blum 和 Mitchell^[5]于 1998 年提出了互训练(co-training)算法, 自此以后大量的半监督学习方法^[6,7]被提出, 比如半监督支持向量分类机(S^3 VC)^[8]等。但绝大部分方法是针对分类问题提出的, 有关半监督回归模型的研究非常少, 从文献中能够找到的只有半监督岭回归^[9,10]和互训练 kNN^[11], 而且半监督分类模型难以直接推广到对应的回归模型^[9]。

LS-SVR 用等式约束代替了 SVR 中的不等式约束, 这样只需求解一个线性方程组而不是二次规划问题, 大大加快了训练速度, 而且大量的实验研究表明性能与 SVR 的性能相当^[12], 受到了人们越来越多的重视。本文在 LS-SVR 的基础上, 提出了半监督 LS-SVR(S^2 LS-SVR)模型, 类似于 LS-SVR, 该模型也只需求解一个线性方程组。最后建模分析了烤烟样品的四种样品成分含量, 平均误差和相关系数指标均优于 PLS 和 LS-SVR。

收稿日期: 2010-12-18, 修订日期: 2011-03-21

基金项目: 国家“十一五”科技支撑计划(2007BAD36B01-04), 中央高校基本科研业务费专项资金(2009-2-05)和公益性行业(农业)科研专项(200903021)资助

作者简介: 李 林, 女, 1963 年生, 中国农业大学信息与电气工程学院计算机系副教授 e-mail: lilincau@gmail.com

* 通讯联系人 e-mail: xush@istic.ac.cn

© 1994-2012 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

1 最小二乘支持向量回归机(LS-SVR)

给定训练数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, 其中 $(x_i, y_i) \in R^d \times R (i = 1, 2, \dots, m)$, 记 $y = (y_1, y_2, \dots, y_m)^T$, 则 LS-SVR^[4] 的原始问题可表述为

$$\min_{w \in R^{n_h}, b \in R, \xi \in R^m} J(w, \xi) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \xi^T \xi \quad (1)$$

$$\text{s.t. } y_i = w^T \Phi(x_i) + b + \xi_i, \quad i = 1, 2, \dots, m \quad (2)$$

其中 $\Phi: R^d \rightarrow R^{n_h}$ 为输入空间到某一高维(可能为无穷维)特征空间 H 的映射, n_h 表示特征空间 H 的维度, $\xi = (\xi_1, \xi_2, \dots, \xi_m)^T$ 为由松弛变量组成的向量, γ 为取正值的正则化调节参数。

通过 Lagrange 函数, 式(1)和(2)的求解可转换为以下线性方程组的求解

$$\begin{bmatrix} 0 & e^T \\ e & K + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (3)$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$ 为 Lagrange 乘子向量, $e = (1, 1, \dots, 1)^T$, I 为单位矩阵, $K_{i,j} = K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) (i, j = 1, 2, \dots, m)$, $K(\cdot, \cdot)$ 为满足 Mercer 定理^[3]的核函数。

记线性方程组(3)的解为 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)^T$ 和 b^* , 则决策函数为

$$\begin{aligned} f(x) &= w^{*T} \Phi(x) + b^* = \sum_{i=1}^m \alpha_i^* \Phi(x_i)^T \Phi(x) + b^* \\ &= \sum_{i=1}^m \alpha_i^* K(x_i, x) + b^* \end{aligned} \quad (4)$$

2 半监督最小二乘支持向量回归机(S²LS-SVR)

给定有标签的训练数据集 $L = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 以及无标签的训练数据集 $U = \{x_{m+1}, x_{m+2}, \dots, x_{m+u}\}$, 其中集合 L 对应于准确测定化学值的样品数据, 而集合 U 对应于无化学值的样品数据, 这些样品数据在建模时通常都是已知的, 而且通常 $u > m$ 。为方便, 记 $y = (y_1, y_2, \dots, y_m)^T$ 。

2.1 原始及对偶问题

假设可以通过某种方式估计得到训练数据集 U 中每个样品的标签(化学值), 记这些估计得到的标签为 $\{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_u\}$ 。为方便, 记 $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_u)^T$ 。类似于半监督岭回归^[9,10], S²LS-SVR 的原始问题可表述为

$$\min_{w \in R^{n_h}, b \in R, \xi \in R^m, \zeta \in R^u} J(w, \xi, \zeta) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \xi^T \xi + \lambda \frac{1}{2} \zeta^T \zeta \quad (5)$$

$$y_i = w^T \Phi(x_i) + b + \xi_i, \quad i = 1, 2, \dots, m \quad (6)$$

$$\hat{y}_i = w^T \Phi(x_{m+i}) + b + \zeta_i, \quad i = 1, 2, \dots, u \quad (7)$$

其中 $\xi = (\xi_1, \xi_2, \dots, \xi_m)^T$ 为对应于训练数据集 L 的松弛变量, $\zeta = (\zeta_1, \zeta_2, \dots, \zeta_u)^T$ 为对应于训练数据集 U 的松弛变量, γ 和 λ 为取正值的正则化调节参数。

式(5)~(7)的 Lagrange 函数为

$$L(w, \xi, \zeta, \alpha, \beta) = J(w, \xi, \zeta) - \sum_{i=1}^m \alpha_i [w^T \Phi(x_i) + b + \xi_i - y_i] - \sum_{i=1}^u \beta_i [w^T \Phi(x_{m+i}) + b + \zeta_i - \hat{y}_i] \quad (8)$$

其中 $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$ 为对应于训练数据集 L 的 Lagrange 乘子, $\beta = (\beta_1, \beta_2, \dots, \beta_u)^T$ 为对应于训练数据集 U 的 Lagrange 乘子, 则对应的 KKT 条件为

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^m \alpha_i \Phi(x_i) + \sum_{i=1}^u \beta_i \Phi(x_{m+i}) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^m \alpha_i + \sum_{i=1}^u \beta_i = 0 \\ \frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \alpha_i = \gamma \xi_i, & i = 1, 2, \dots, m \\ \frac{\partial L}{\partial \zeta_i} = 0 \Rightarrow \beta_i = \lambda \zeta_i, & i = 1, 2, \dots, u \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T \Phi(x_i) + b + \xi_i - y_i = 0, & i = 1, 2, \dots, m \\ \frac{\partial L}{\partial \beta_i} = 0 \Rightarrow w^T \Phi(x_{m+i}) + b + \zeta_i - \hat{y}_i = 0, & i = 1, 2, \dots, u \end{cases} \quad (9)$$

类似于 LS-SVR, 式(9)可表示为线性方程组

$$\begin{bmatrix} 0 & e_m^T & e_u^T \\ e_m & K_{1,1} + \gamma^{-1} I_m & K_{1,2} \\ e_u & K_{2,1} & K_{2,2} + \lambda^{-1} I_u \end{bmatrix} \begin{bmatrix} b \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \\ \hat{y} \end{bmatrix} \quad (10)$$

其中 e_m 和 e_u 分别为包含 m 和 u 个元素的全 1 列向量, I_m 和 I_u 分别为 m 阶和 u 阶单位矩阵。记 $Z_1 = (\Phi(x_1), \Phi(x_2), \dots, \Phi(x_m))$, $Z_2 = (\Phi(x_{m+1}), \Phi(x_{m+2}), \dots, \Phi(x_{m+u}))$, 则 $K_{1,1} = Z_1^T Z_1$, $K_{1,2} = K_{2,1}^T = Z_1^T Z_2$, $K_{2,2} = Z_2^T Z_2$ 。

记线性方程组(10)的解为 $\alpha^* = (\alpha_1^*; \alpha_2^*; \dots; \alpha_m^*)$, $\beta^* = (\beta_1^*; \beta_2^*; \dots; \beta_u^*)$ 和 b^* , 则决策函数为

$$\begin{aligned} f(x) &= w^{*T} \Phi(x) + b^* = \sum_{i=1}^m \alpha_i^* K(x_i, x) + \\ &\quad \sum_{i=1}^u \beta_i^* K(x_{m+i}, x) + b^* \end{aligned} \quad (11)$$

2.2 估计训练数据集 U 的标签

由 2.1 节知, 线性方程组(10)的求解需要利用训练数据集 U 的标签, 但这些标签事先是未知的。无论是半监督分类还是半监督回归, 都需要通过某种方式对集合 U 的标签进行估计^[7-11]。本文利用了特征空间 H 中近邻的思想, 具体来说, 对于 $\forall x \in U$, 记集合 L 中离 $\Phi(x)$ 最近的 k 个元素为 $N_k = \{\Phi(x_{i_1}), \Phi(x_{i_2}), \dots, \Phi(x_{i_k})\}$, 则 $x \in U$ 的标签可由集合 N_k 中元素的标签按距离远近加权得到, 即 $\hat{y} = \left[\sum_{i=1}^k d_i^{-1} y_{i_i} \right] / \left[\sum_{i=1}^k d_i^{-1} \right]$, 其中 d_i 为在特征空间 H 中 $\Phi(x)$ 与 $\Phi(x_{i_i})$ 间的距离, 利用核技巧^[13]在原始空间该距离可计算为

$$\begin{aligned} \|\Phi(x) - \Phi(x_{i_i})\| &= \sqrt{(\Phi(x) - \Phi(x_{i_i}))^T (\Phi(x) - \Phi(x_{i_i}))} \\ &= \sqrt{K(x, x) - 2K(x_{i_i}, x) + K(x_{i_i}, x_{i_i})} \end{aligned} \quad (12)$$

3 实验方法及实验数据

样品来自 2005 年云南省昆明地区的 115 份优质烤烟。近红外光谱数据由德国 Bruker 公司生产的 Vector/22N 型傅

里叶变换近红外光谱仪测定, 其波谱范围为 4 000~ 12 000 cm^{-1} , 近红外光谱的波长点有 2 099 个。光谱数据由 Origin 软件经过一阶导数、13 点平滑的预处理。化学组分值有四个, 分别为总糖、还原糖、总氮和烟碱的各自含量百分比, 由云南省烟草科学研究所应用荷兰 Skalar 公司的 SAN plus 流动注射自动分析仪测定。

4 实验结果及分析

本文从 115 个烤烟样品中随机选取 69 个作为训练集合 L , 23 个作为训练集合 U , 其余 23 个作为测试集。为了说明 $\text{S}^2\text{LS-SVR}$ 模型的优势, 以 PLS 和 LS-SVR 两种模型做参照, 平均相对误差 δ 和相关系数 R 作为评价指标。核函数采用径向基 (RBF) 核: $K(\mathbf{x}, \mathbf{z}) = \exp(-p \|\mathbf{x} - \mathbf{z}\|^2)$, $p > 0$ 。之所以选择该核函数, 原因有三方面, (1) 线性核是 RBF 核的一个特例^[14]; (2) Sigmoid 核不是正定核, 而且对于一定的参数取值, Sigmoid 核和 RBF 核的性能相当^[15]; (3) 相对来说, 多项式核的参数比较多, 优化起来比较困难, 而且它在运算时可能会溢出。

4.1 参数选择

参数(γ , λ , p , k)对 LS-SVR 和 $\text{S}^2\text{LS-SVR}$ 模型的性能

影响比较大。为了选择一组合适的参数, 本文对 LS-SVR 采用格网搜索选参法^[16-17]: 令参数 γ 和 p 按对数增长的方式分别从集合 $\{2^{-5}, 2^{-3}, \dots, 2^{15}\}$ 及 $\{2^{-15}, 2^{-13}, \dots, 2^3\}$ 中取值, 根据平均相对误差 δ 最小的原则采用留一法 (LOO) 确定合适的参数 γ^* 和 p^* 。而对于 $\text{S}^2\text{LS-SVR}$, 本文采用两层格网搜索选参法, 首先仅根据训练集合 L , 利用 LS-SVR 模型采用格网搜索选择参数 γ^* 和 p^* , 然后固定 γ^* 和 p^* , 引入训练集合 U , 令 $\lambda \in \{2^{-5}, 2^{-3}, \dots, 2^{15}\}$, $k \in \{1, 2, \dots, 5\}$, 根据在集合 L 上平均相对误差 δ 最小的原则采用留一法确定合适的参数 λ^* 和 k^* 。对于 PLS, 本文令主成分的个数 $nFactor \in \{1, 2, \dots, \minRank\}$, 其中 \minRank 表示输入矩阵的行秩与列秩的最小值, 根据平均相对误差 δ 最小的原则采用留一法确定合适的参数 $nFactor^*$ 。

4.2 结果分析

根据 4.1 节所介绍的方法确定的最优参数, 分别利用 PLS, LS-SVR 和 $\text{S}^2\text{LS-SVR}$ 建立定量分析模型, 确定烤烟样品各成分含量的预测值并计算相应的评价指标, 详见表 1。其中 Part 是指 PLS 和 LS-SVR 在建模时只考虑了训练集合 L 中的信息, 而 Both 是指 PLS 和 LS-SVR 在建模时不仅考虑了训练集合 L 中的信息, 而且考虑了训练集合 U 以及对应的标签 (化学值) 信息。

Table 1 Comparisons of the predicted results with PLS, LS-SVR and $\text{S}^2\text{LS-SVR}$

	总糖		还原糖		总氮		烟碱	
	$\delta/\%$	R	$\delta/\%$	R	$\delta/\%$	R	$\delta/\%$	R
PLS (Part)	9.39	0.770 1	9.47	0.569 6	8.14	0.517 3	9.20	0.636 5
PLS (Both)	7.17	0.933 1	7.73	0.896 4	5.70	0.883 0	7.38	0.589 9
LS-SVR (Part)	9.00	0.894 3	9.56	0.877 4	7.85	0.819 1	9.33	0.886 7
LS-SVR (Both)	6.47	0.969 8	7.00	0.964 9	5.77	0.933 4	7.61	0.955 8
$\text{S}^2\text{LS-SVR}$	6.62	0.974 1	7.56	0.973 3	6.11	0.9230	8.20	0.948 6

从表 1 容易看出, 在训练集合 U 的标签信息未知的条件下, $\text{S}^2\text{LS-SVR}$ 的效果最好, LS-SVR 和 PLS 的性能相当。原因有两方面: (1) $\text{S}^2\text{LS-SVR}$ 在建模时同时考虑了有标签训练集 L 和无标签训练集 U , 集合 U 极大地提高了 $\text{S}^2\text{LS-SVR}$ 模型的性能; (2) 尽管在建模时集合 U 通常是已知的, 但由于其标签信息的缺失使得 LS-SVR 无法利用集合 U 中的近红外光谱信息。当引入训练集合 U 的标签信息时, $\text{S}^2\text{LS-SVR}$ 可逼近 PLS 和 LS-SVR 的性能, 从理论上来说, 建立在这些信息基础上的模型是对应的半监督模型的上限^[7]。

5 结论及讨论

受半监督岭回归模型的启发, 本文在 LS-SVR 的基础上提出了一种新的近红外光谱定量分析方法—— $\text{S}^2\text{LS-SVR}$ 。

首先利用特征空间中的近邻特征估计无化学值 (标签) 样品的成分含量, 然后将其引入到模型的建模过程中, 一方面可以扩大训练集合的大小, 提供更多的训练数据, 另一方面可以大大减少因样品成分含量测定不准引入的噪声, 提高模型的鲁棒性。最后, 以烤烟样品数据集为实验材料, 通过与 PLS 和 LS-SVR 的对比分析, 验证了 $\text{S}^2\text{LS-SVR}$ 的可行性和有效性, 同时为发展化学计量学提供了一种新的定量分析建模方法。

另外, 近红外光谱包含了生物样品中所有样品成分的光谱信息, 本文提出的模型为每种样品成分含量分开进行建模分析, 尽管利用了无化学值样品数据的信息, 但失去了各种样品成分含量间可能存在的联系, 如何将样品成分含量间的联系引入到半监督回归模型中是我们下一步的研究内容之一。

References

[1] YAN Yan-lu, ZHAO Long-lian, HAN Dong-hai, et al (严衍禄, 赵龙莲, 韩东海, 等). Foundation of Near Infrared Spectral Analysis and its Applications (近红外光谱分析基础与应用). Beijing: China Light Industry Press (北京: 中国轻工业出版社), 2005.

- [2] Abdi H. Partial Least Squares (PLS) Regression. Encyclopedia for Research Methods for the Social Sciences, Lewis-Beck M, Bryman A, Futing T, eds. Sage, Thousand Oaks, CA, 2003. 792.
- [3] Vapnik V N. The Nature of Statistical Learning Theory, 2nd Edition. New York: Springer Verlag, 1999.
- [4] Suykens J A K, Van Gestel T, Brabanter J D, et al. Least Squares Support Vector Machines. World Scientific Pub. Co., Singapore, 2002.
- [5] Blum A, Mitchell T. Combining Labeled and Unlabeled Data with Co-Training. Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT), Madison, Wisconsin, United States, 1998. 92
- [6] Zhu X. Semi-Supervised Learning Literature Survey. Technical Report 1530, Department of Computer Sciences, University of Wisconsin, Madison, 2008.
- [7] Chapelle O, Schölkopf B, Zien A. Semi-Supervised Learning. Cambridge: MIT Press, 2006.
- [8] Chapelle O, Sindhwani V, Keerthi S S. Journal of Machine Learning Research, 2008, 9(2): 203.
- [9] Cortes C, Mohri M. On Transductive Regression. Advances in Neural Information Processing Systems 19, Schölkopf B, Platt J, Hoffman T, eds. MIT Press, Cambridge, MA, 2007. 305.
- [10] Brefeld U, Gärtner T, Scheffer T, et al. Efficient Co-Regularised Least Squares Regression. Proceedings of the 23rd International Conference on Machine Learning (ICML), 2006. 137.
- [11] Zhou Z-H, Li M. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(11): 1479.
- [12] Van Gestel T, Suykens J A K, Baesens B, et al. Machine Learning, 2004, 54(1): 5.
- [13] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge: Cambridge University Press, 2004.
- [14] Keerthi S S, Lin C J. Neural Computation, 2003, 15(7): 1667.
- [15] Lin H T, Lin C J. A Study on Sigmoid Kernels for SVM and the Training of Non-PSD Kernels by SMO-Type Methods. Technical Report, Department of Computer Science, National Taiwan University, 2003.
- [16] Hsu G-W, Chang G-C, Lin G-J. A Practical Guide to Support Vector Classification. Available [online]: <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
- [17] Xu S, Ma F J, Tao L. Learn from the Information Contained in the False Splice Sites as well as in the True Splice Sites using SVM. Proceedings of the International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Chengdu, China, 2007. 1360.

A Novel Approach to NIR Spectral Quantitative Analysis: Semi-Supervised Least-Squares Support Vector Regression Machine

LI Lin¹, XU Shuo^{2*}, AN Xin³, ZHANG Lu-da⁴

1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100193, China
2. Information Technology Supporting Center, Institute of Scientific and Technical Information of China, Beijing 100038, China
3. School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China
4. College of Science, China Agricultural University, Beijing 100193, China

Abstract In near infrared spectral quantitative analysis, the precision of measured samples' chemical values is the theoretical limit of those of quantitative analysis with mathematical models. However, the number of samples that can obtain accurately their chemical values is few. Many models exclude the amount of samples without chemical values, and consider only these samples with chemical values when modeling sample compositions' contents. To address this problem, a semi-supervised LS-SVR (S^2 LS-SVR) model is proposed on the basis of LS-SVR, which can utilize samples without chemical values as well as those with chemical values. Similar to the LS-SVR, to train this model is equivalent to solving a linear system. Finally, the samples of flue-cured tobacco were taken as experimental material, and corresponding quantitative analysis models were constructed for four sample compositions' content (total sugar, reducing sugar, total nitrogen and nicotine) with PLS regression, LS-SVR and S^2 LS-SVR. For the S^2 LS-SVR model, the average relative errors between actual values and predicted ones for the four sample compositions' contents are 6.62%, 7.56%, 6.11% and 8.20%, respectively, and the correlation coefficients are 0.974 1, 0.973 3, 0.923 0 and 0.948 6, respectively. Experimental results show the S^2 LS-SVR model outperforms the other two, which verifies the feasibility and efficiency of the S^2 LS-SVR model.

Keywords Near infrared spectrum; Chemometrics; Semi-supervised LS-SVR(S^2 LS-SVR)

* Corresponding author

(Received Dec. 18, 2010; accepted Mar. 21, 2011)