

# Near Infrared Spectroscopy Modeling Based on Adaptive Elastic Net Method

ZHENG Nian-nian, LUAN Xiao-li\*, LIU Fei

Key Laboratory for Advanced Process Control of Light Industry of Ministry of Education,  
Institute of Automation, Jiangnan University, Wuxi 214122, China

**Abstract** When near infrared spectral information is much larger than the sample size, it is both important and challenging to make automatic variable selection of spectral information and coefficient estimation to establish a sparse linear model between spectra and sample concentration. In this paper, adaptive Elastic Net, a variable selection method, is used to establish a quantitative calibration model between near infrared spectroscopy and o-cresol content, which is a kind of trace component and is difficult to measure in the production of polyphenylene ether. Then, the model performance is compared with the Elastic Net method. Under the circumstance that the number of variables is much larger than the sample size, although Elastic Net method can achieve variable selection, due to the fact that its coefficient estimation does not have the Oracle property, the interpretability and prediction accuracy of the model are affected. The adaptive Elastic Net method solves the above problem and improves the model performance by applying adaptive weights to L1 penalty. In order to verify model performance indicators of adaptive Elastic Net method, the number of selected independent variables (NSIV) is used to evaluate the model complexity and the complex correlation coefficient  $R^2$  is used to evaluate the interpretability of the model. Meanwhile, the prediction accuracy of the model is evaluated by using the mean relative prediction error (MRPE) and the prediction correlation coefficient ( $R_p$ ). The performance indicators of Elastic Net Method are: NSIV=529,  $R^2=0.96$ , MRPE=3.22%,  $R_p=0.97$ ; adaptive Elastic Net method's performance indicators are: NSIV=139,  $R^2=0.99$ , MRPE=2.00%,  $R_p=0.99$ . The results show that adaptive Elastic Net's model is better than that of Elastic Net. A simpler sparse linear model with better interpretability and higher prediction accuracy can be obtained by the adaptive Elastic Net regression.

**Keywords** Near infrared spectroscopy; Adaptive Elastic Net; Interpretability; Prediction accuracy  
中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2019)01-0319-06

## Introduction

Near infrared spectroscopy (NIRS) is an independent analytical technique gradually developed in the late 1980s. It has the advantages of fast speed, less pollution and being suitable for on-line analysis of process industry. This technique has been widely used in many fields, such as agricultural testing, drug analysis, petrochemical industry, food industry and bas-

ic chemistry<sup>[1-2]</sup>. Establishing a robust linear model between near infrared spectroscopy information and the sample concentration is very important for quantitative analysis of near infrared spectroscopy. Traditionally, the modeling technically is mainly factor analysis. Typical applications are principal component regression (PCR) and partial least squares (PLS), which are applicable only for linear models<sup>[3-4]</sup>. Recent years, support vector machine (SVM) and artificial neural network (ANN) as nonlinear quantitative calibration methods have also

Received: 2017-11-28; accepted: 2018-05-12

Foundation item: Supported by National Natural Science Foundation of China(61473137, 61722306)

Biography: ZHENG Nian-nian, (1997—), Master degree candidate of automation, Jiangnan University e-mail: 1379924290@qq.com

\* Corresponding author e-mail: xlluan@jiangnan.edu.cn

been used by more and more researchers<sup>[5-7]</sup>. Since the amount of near infrared spectroscopy information (the number of variables which are to be chosen from) is much larger than the sample size (number of samples), in order to reduce the complexity of the model, it is very important to select significant variables from the spectral information (excluding invalid information and retaining variables that have significant effects on sample quality indices). However, the aforementioned methods inevitably require all of the independent variables in the process of fitting a linear model, and cannot be used for variable selection to establish the characteristic sparse model of near infrared spectroscopy<sup>[8]</sup>.

In the fields of multivariate linear analysis and regression modeling, the Elastic Net (EN) method has gained more and more attention in recent years because it can carry out variable selection and parameter estimation automatically and simultaneously<sup>[9-10]</sup>. The proposed methods can not only filter out important variables but also eliminate the influence of multicollinearity between variables on the model. However, due to the lacking of Oracle property in coefficient estimation process of the EN method, interpretability and prediction accuracy of model are affected. To solve this problem, the adaptive Elastic Net (AEN) method was proposed as an improved version to EN in 2009, which applies adaptive weights to the L1 penalty of EN and automatically adjusts the degree of compression to different independent variables. Therefore, the coefficient estimation of AEN method has the Oracle property, which can effectively improve the model's interpretability and prediction accuracy<sup>[11]</sup>.

In the production process of polyphenylene ether, it is a difficult task to measure the concentration of o-cresol which is a kind of trace component. To address this problem, in this article, AEN modeling method is used to establish a sparse linear model between the near infrared spectroscopy and o-cresol content in the case where spectral information is much larger than the sample size. Taking advantage of the correlation between spectra and o-cresol, AEN is exploited to improve the modeling performance of EN. The experimental results show that the model established by AEN is simpler but has better interpretability than EN, and its model prediction accuracy is higher than EN.

## 1 Elastic Net and Adaptive Elastic Net modeling

Assuming the following independent variables (explanatory variables) consist of a matrix

$$\mathbf{X} = [X_1, X_2, \dots, X_p] \in \mathbf{R}^{n \times p}$$

where  $\mathbf{X}_j = [x_{1j} \ x_{2j} \ \dots \ x_{nj}]^T$ ,  $j=1, 2, \dots, p$  represents the  $j^{\text{th}}$  independent variable.

Dependent variable (response variable) consists of a vector

$$\mathbf{Y} = [y_1 \ y_2 \ \dots \ y_n]^T \in \mathbf{R}^{n \times 1}$$

The goal is to establish a multivariate linear model between the dependent variable  $\mathbf{Y}$  and  $p$  independent variables  $X_1, X_2, \dots, X_p$  as follows:

$$\mathbf{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \boldsymbol{\varepsilon} \quad (1)$$

What should be noticed is that  $\beta_0$  is a constant term, standing for the intercept of the linear model;  $\beta_1, \beta_2, \dots, \beta_p$  are regression coefficients, which indicate how much each argument affects the dependent variable. What's more,  $\boldsymbol{\varepsilon} = [\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n]^T$  is a random error term and satisfies the following relationship

$$\begin{cases} E(\boldsymbol{\varepsilon}) = 0 \\ \text{Var}(\boldsymbol{\varepsilon}) = \sigma^2 I \end{cases} \quad (2)$$

Normalizing and centralizing the independent variables, is has

$$\begin{cases} \frac{1}{n} \sum_{i=1}^n x_{ij} = 0 \\ \sum_{i=1}^n x_{ij}^2 = 1 \end{cases} \quad (3)$$

In a linear model, the constant term is estimated by an equation  $\hat{\beta}_0 = \bar{Y}$ . Without loss of generality, it is assumed that the observation of dependent variable is also centered, that is

$$\frac{1}{n} \sum_{i=1}^n y_i = 0 \quad (4)$$

At this point, the constant term  $\beta_0 = 0$ , so the linear model (1) can be re-expressed as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (5)$$

Now the coefficients estimation can be expressed as  $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_p]^T$ .

### 1.1 Elastic Net modeling

For the linear model (5), the EN method can be expressed as

$$\begin{cases} \hat{\boldsymbol{\beta}}^{\text{EN}} = \underset{\boldsymbol{\beta}}{\text{argmin}}\{SSE\} \\ \text{subject to: } \sum_{j=1}^p |\beta_j| + \sum_{j=1}^p \beta_j^2 \leq t \end{cases} \quad (6)$$

In the above formula,  $t \geq 0$  is the tuning parameter, indicating constraint intensity to the coefficients. SSE is the logarithm of "sum of squared residuals" and its detailed expression is

$$SSE = \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 \quad (7)$$

As can be seen from formula (6), EN method is essentially an optimization problem with inequality constraints. Therefore, its basic idea can be expressed as: under the condition that the value of adding the sum of the absolute values of the regression coefficients and the sum of the squares of the regression coefficients together is less than the tuning parameter, and the residual sum of squares is minimized so that the

regression coefficients of nonsignificant arguments shrink to zero. In fact, formula (6) can also be equivalently expressed as:

$$\hat{\beta}^{EN}(\lambda_1, \lambda_2) = \arg \min_{\beta} \left\{ SSE + \lambda_1 \sum_{j=1}^p |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\} \quad (8)$$

In the above formula,  $\lambda_1$  and  $\lambda_2$  are two non-negative real numbers and are related to the selection of the tuning parameter. It can be easily seen from formula (8) that EN penalizes both the L1 norm and the L2 norm of coefficients of independent variables during the coefficient estimation.

In this paper, we transform EN into Lasso problem and use the steepest descent algorithm to estimate the coefficients<sup>[12]</sup>. At the same time, we use cross-validation to find the optimal tuning parameters.

### 1.2 Adaptive Elastic Net modeling

The EN method is widely used because of its excellent ability to implement variable selection and coefficient estimation simultaneously, then to establish a multivariate sparse linear model. However, in spite of its popularity, the EN method does have some drawbacks; the L1 penalty term in this method compresses all coefficients (including the major independent variables and unimportant ones) with the same intensity which can be characterized by  $\lambda_1$ . In this situation, removing unimportant arguments will be at the cost of reducing the regression coefficients of important independent variables, which leads to EN method lacking the Oracle property. When the sample size increases, the variable selection and coefficient estimation of Elastic Net method become discontinuous, which affects the performance of model. Specifically, the interpretability of the model and the prediction accuracy are affected. As an improvement on EN, AEN's basic idea is to compress the coefficients of independent variables having different degrees of importance with different intensity, and the idea is successfully accomplished by implementing adaptive weights for each coefficient in L1 penalty. The AEN estimator can be expressed as

$$\hat{\beta}^{AEN}(\lambda_1, \lambda_2) = \arg \min_{\beta} \left\{ SSE + \lambda_1 \sum_{j=1}^p w_j |\beta_j| + \lambda_2 \sum_{j=1}^p \beta_j^2 \right\} \quad (9)$$

where  $W = \{w_j \mid w_j \geq 0\}$ , ( $j = 1, 2, \dots, p$ ) is an adaptive weight vector and is calculated by following formula:

$$w_j = \frac{1}{(|\hat{\beta}_j^{ini}|)^\gamma}, \quad (j = 1, 2, \dots, p) \quad (10)$$

where  $\gamma$  is a positive constant and  $\hat{\beta}^{ini}$  is an initial estimated coefficient vector. In the calculation process,  $\hat{\beta}^{ini}$  can be replaced by  $\hat{\beta}^{ols}$  or  $\hat{\beta}^{EN}$  (where  $\hat{\beta}^{ols}$  is the ordinary least squares estimation of coefficients and  $\hat{\beta}^{EN}$  stands for EN estimation). What's more, if  $\hat{\beta}_j^{ini}$  equals zero, then  $w_j$  equals to a very large positive number.

When solving AEN problem, we firstly convert it into Lasso problem, but there are still two adjustment parameters:  $\lambda$  and  $\gamma$ . Therefore, the solution process can be completed in two steps:

First, the parameter  $\gamma$  is fixed, and the parameter  $\lambda$  is changed, and the steepest descent algorithm is used to find the path of the solution. The optimal parameter estimation is found by using the cross-validation criterion.

Second, change the parameter  $\gamma$  to find a set of optimal coefficient estimations, and then use the cross-validation criterion again to find the final coefficient estimation.

## 2 Model evaluation index

In fact, for linear model (5), we focus on the performance of the following three aspects:

(1) The complexity of the model

In multivariate linear regression, the fewer the number of independent variables are included in the final model, the simpler the model is. Therefore, the complexity of the model is evaluated by using the number of independent variables (NS-IV).

(2) Interpretability of the model

In this paper, the complex correlation coefficient  $R^2$  is used to evaluate the interpretability of the model and is calculated as follow<sup>[13]</sup>

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij}\beta_j)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (11)$$

Where SST is the logogram of "sum of squares for total".

(3) Prediction accuracy of the model

The mean relative prediction error (MRPE) and the prediction correlation coefficient ( $R_p$ ) are used to judge the prediction accuracy of the established model and can be calculated as follows (variables with wavy symbol come from the forecast set)

$$R_p = \frac{\sum_{i=1}^n (\tilde{y}_i - \sum_{j=1}^p \tilde{x}_{ij}\beta_j)^2}{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2} \quad (12)$$

$$MRPE = \frac{\sum_{i=1}^n (\tilde{y}_i - \sum_{j=1}^p \tilde{x}_{ij}\beta_j)}{\sum_{i=1}^n \tilde{y}_i} \quad (13)$$

## 3 Experimental verification and comparative analysis

### 3.1 Experimental material and experimental equipment

The experimental material used in this study is a kind of

chemical raw material—polyphenylene ether provided by a laboratory. The raw material is in a liquid state and the *o*-cresol concentration in the raw material is the dependent variable. Data acquisition is carried out using the MATRIX-F Fourier Transform Infrared Spectrometer (including OPUS quantitative analysis software package) manufactured by Bruker AG in Germany with a spectral measurement range of  $12\ 800 \sim 4\ 000\ \text{cm}^{-1}$  and a minimum spectral scanning resolution of  $2\ \text{cm}^{-1}$ . The experiment sets the spectral measurement wavelength range of  $10\ 000 \sim 4\ 500\ \text{cm}^{-1}$ , a resolution of  $8\ \text{cm}^{-1}$ .

### 3.2 Spectral information collection and data division

In this experiment, there are a total of 200 samples of chemical raw materials and each sample's *o*-cresol concentration value is measured accurately. While using the near infrared spectroscopy, each sample is scanned three times in a row to obtain 600 sets of spectral sample data. Then the measured 600 sets of samples are divided into two parts; the first 400 sets of samples are used to model as training set and the remaining 200 as forecast set are used to test the model's prediction accuracy.

### 3.3 Experimental results and comparative analysis

#### 3.3.1 Modeling with EN method

When using the EN method for variable selection and modeling, the solution path is shown in Figure 1.

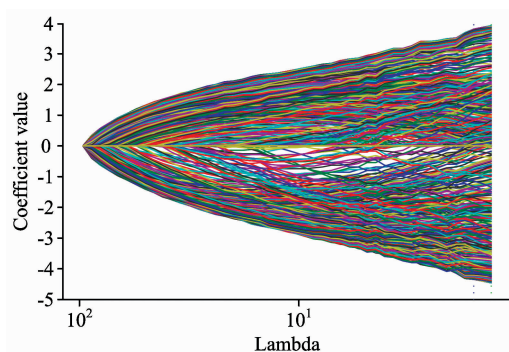


Fig. 1 Coefficients vary with parameter  $\lambda$ (EN)

The abscissa indicates the value of the tuning parameter  $\lambda$ , and the ordinate represents the trend of the coefficient value of each independent variable. From left to right, the abscissa decreases continuously, which implies that the compressive strength or the punishment efforts to independent variable coefficients decreases continuously. Therefore, with the continuous decreasing of  $\lambda$ , independent variables get into the model continually (coefficient value changes from the initial zero to non-zero) and the coefficient value increases continuously. Conversely, from right to left of the abscissa, as  $\lambda$  increases, coefficients of independent variables decrease until they reach zero.

Meanwhile, in the process of solving EN, the optimal adjustment parameter needs to be found through cross-validation

method, that is, the optimal adjustment parameter will be determined when mean squared error (MSE) is minimum. The relationship between the cross-validation criteria and adjustment parameter  $\lambda$  is shown in Figure 2, the black arrow refers to the optimal tuning parameter.

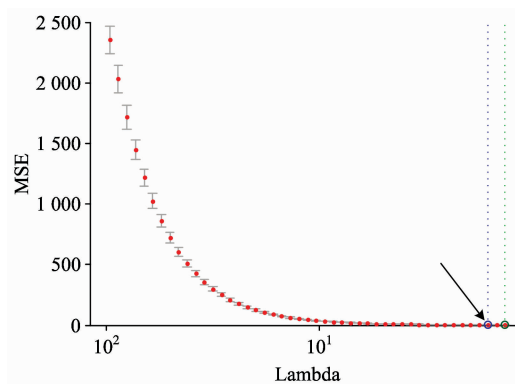


Fig. 2 Optimal adjustment parameter  $\lambda$ (EN)

#### 3.3.2 Modeling with AEN method

When using the AEN method for variable selection and model building, firstly an adaptive weight vector should be constructed using the Elastic Net estimation that has been found. The path of the solution is shown in Figure 3. The abscissa stands for the value of tuning parameter  $\lambda$ , and the ordinate represents the trend of each argument coefficient. In this case, the parameter  $\gamma$  is fixed equal to 0.5. It can be seen from the figure that as the tuning parameter  $\lambda$  decreases, more independent variables enter the model continually.

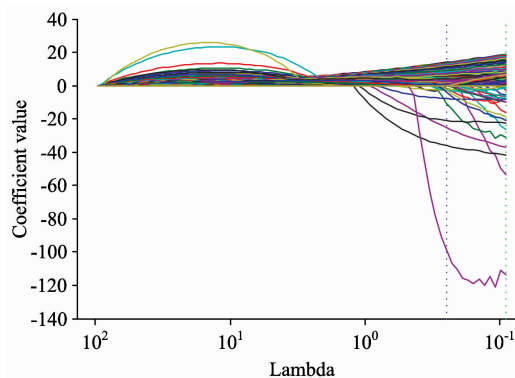


Fig. 3 Coefficients vary with parameter  $\lambda$ (AEN)

Correspondingly, Figure 4 represents the process of selecting the optimal coefficient estimation by cross-validation criteria when  $\gamma=0.5$ . It is noteworthy that the process of solving AEN is only half completed so far. Then, a series of coefficient estimations are obtained by adjusting the value of parameter  $\gamma$ , and the optimum independent variable coefficients are determined by using cross-validation criteria finally.

In addition, it can be easily found by comparing Figure 1 with Figure 3 that the number of color lines in Figure 3 is sig-

nificantly less than Figure 1, indicating that there are fewer independent variables entering the model during the process of solving AEN, so the model established by this method is more streamlined.

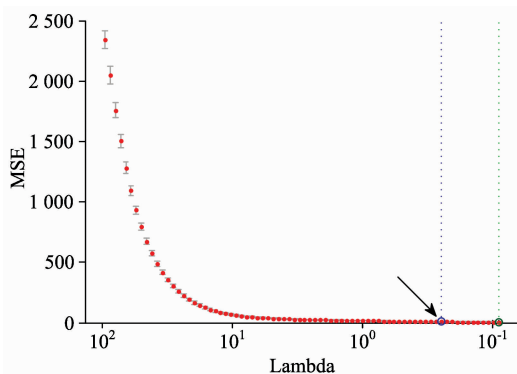


Fig. 4 Optimal adjustment parameter  $\lambda$ (AEN)

3.3.3 Comparing the performance of two modeling methods

Firstly, in order to have an intuitive comparison of the regression effects of the two models, a residual sequence function will be defined and is expressed as follows:

$$e(i) = y(i) - \hat{y}(i), i = 1, 2, \dots, n \quad (14)$$

For the training set and the forecasting set, the residual sequence curves are drawn respectively as shown in Figure 5

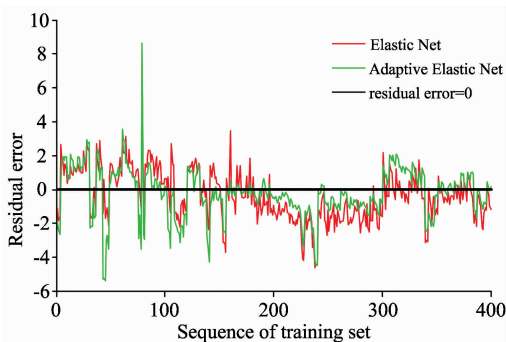


Fig. 5 Residual curves of training set

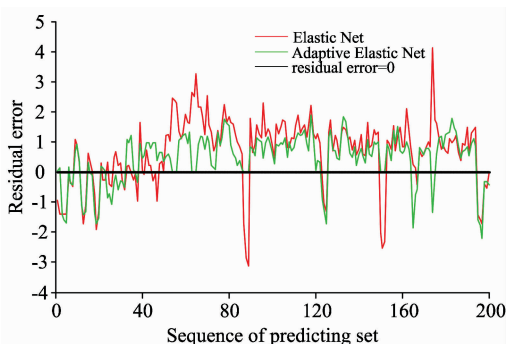


Fig. 6 Residual curves of predicting set

and Figure 6. It can be seen that whether for the training set or for the forecasting set, AEN method's residual has less fluctuation than the EN method and is more concentrated near the line with a residual of zero, which shows that the EN method fits well.

In order to compare the two kinds of variable selection methods ulteriorly and precisely, a list of their respective model evaluation indicators is shown in Table 1. As can be seen from Table 1:

(1) The AEN method makes fewer arguments included in the final model, which indicates that the linear model established by this method is simpler.

(2) The complex correlation coefficient  $R^2$  of AEN method is greater than that of EN method, indicating that the model established by AEN method could reflect the linear relationship between spectrum and o-cresol concentration more accurately, that is, the model has stronger explanatory power.

(3) AEN's mean relative prediction error MRPE is less than EN, and its prediction correlation coefficient  $R_p$  is larger, so it has better prediction performance.

Table 1 The evaluation index of the model established by different methods

Method	Index			
	Complexity	Interpretability	Prediction accuracy	
	NSIV	$R^2$	$R_p$	MRPE/%
Elastic Net	529	0.96	0.97	3.22
Adaptive Elastic Net	139	0.99	0.99	2.00

4 Conclusion

In the process of near infrared spectroscopy modeling, AEN is a well-behaved modeling method if the number of independent variables is much larger than the sample size. This method can not only establish a simple linear model only involving a small number of independent variables, but also select important variables that have a significant influence on the response variables, so a linear model with better interpretation performance can be established. In addition, the model established by this method has high prediction accuracy. When this method is applied to the measurement of o-cresol—a kind of trace component in polyphenylene ether, experiments show that the model built by AEN is not only simpler, but its interpretability and prediction accuracy are superior to EN.

## References

- [ 1 ] SUN Ji-cheng, MA Jin, SHEN Chao, et al. Progress in Modern Biomedicine, 2016, 16(8): 1594.
- [ 2 ] ZHANG Li-pei, LU Xiong. Light Industry Science and Technology, 2016, (2): 103.
- [ 3 ] SHI Ting, LUAN Xiao-li, LIU Fei. Spectroscopy and Spectral Analysis, 2017, 37(4): 1058.
- [ 4 ] SHI Ting, LUAN Xiao-li, LIU Fei. Vibrational Spectroscopy, 2017, 92: 302.
- [ 5 ] TIAN Kuang-da, QIU Kai-xian, LI Zu-hong, et al. Spectroscopy and Spectral Analysis, 2014, 32(12): 3262.
- [ 6 ] YE Shu-bin, XU Liang, LI Ya-kai, et al. Spectroscopy and Spectral Analysis, 2017, 37(3): 749.
- [ 7 ] HUANG Xiao-han, ZHANG Ping, YANG Xiao-li, et al. Gansu Science and Technology, 2017, 33(18): 123.
- [ 8 ] TANG Shou-peng, YAO Xin-feng, YAO Xia, et al. Chinese Journal of Analytical Chemistry, 2009, 37(10): 1445.
- [ 9 ] XU Qing-juan, YANG Bin-bin. Journal of Guangxi Teachers Education University: Natural Science Edition, 2017, 33(4): 36.
- [10] Zou H, Hastie T. Journal of the Royal Statistical Society. Series B (Methodological), 2005, 67(1): 301.
- [11] Zou H, Hao Helen Zhang. The Annals of Statistics, 2009, 37(4): 1733.
- [12] CHEN Shan-xiong, LIU Xiao-juan, CHEN Chun-rong, et al. Journal of Computer Applications, 2017, 37(6): 1674.
- [13] HE Xiao-qun, LIU Wen-qing. Applied Regression Analysis. Beijing:China Renmin University Press,2015.

## 基于 Adaptive Elastic Net 方法的近红外光谱建模技术

郑年年, 栾小丽\*, 刘 飞

江南大学自动化研究所, 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122

**摘 要** 当近红外光谱信息远远大于样本量时, 对光谱信息进行自动变量选择进而建立光谱与微量成分含量之间的稀疏线性模型重要且具有挑战性。针对聚苯醚生产过程中微量成分邻甲酚难以测量的问题, 将变量选择方法 Adaptive Elastic Net 用于建立近红外光谱与邻甲酚含量之间的定量校正模型, 并将其模型性能与 ElasticNet 方法进行对比。在变量数目远远大于样本量的情形下, ElasticNet 方法虽可以实现变量选择, 但由于其系数估计不具备 Oracle 性质, 使得模型的可解释性和预测精度受到影响, 而 Adaptive Elastic Net 方法通过对 L1 惩罚项施加自适应权重从而很好的解决了上述问题并提高了模型性能。为了验证 Adaptive Elastic Net 方法的模型性能指标, 用最终被选中的自变量数目来评价模型复杂度; 利用复相关系数  $R^2$  来评价模型的可解释性, 利用平均相对预测误差 MRPE(mean relative prediction error)和预测相关系数  $R_p$  来评价模型的预测精度。Elastic Net 方法建立的模型性能指标为: NSIV=529,  $R^2=0.96$ , MRPE=3.22%,  $R_p=0.97$ ; Adaptive Elastic Net 方法的性能指标为: NSIV=139,  $R^2=0.99$ , MRPE=2.00%,  $R_p=0.99$ 。结果表明: Adaptive Elastic Net 所建立模型的性能指标优于 Elastic Net 方法, 可以得到更加简单且具有较强可解释性和较高预测精度的稀疏线性模型。

**关键词** 近红外光谱; Adaptive Elastic Net; 可解释性; 预测精度

(收稿日期: 2017-11-28, 修订日期: 2018-05-12)

\* 通讯联系人