Journal of Pharmaceutical and Biomedical Analysis 50 (2009) 158–163

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Quantitative analysis of diclofenac sodium powder via near-infrared spectroscopy combined with artificial neural network

Bin Wang[a], Guoliang Liu[b], Ying Dou[c], Liwen Liang[a], Haitao Zhang[a], Yulin Ren[a,*]

[a] College of Chemistry, Jilin University, 2519 Jiefang Road, Changchun, 130021, China
[b] School of Public Health, Jilin University, Changchun, 130021, China
[c] College of Science, Tianjin University of Science & Technology, Tianjin, 300222, China

ARTICLE INFO

Article history:
Received 18 January 2009
Received in revised form 9 April 2009
Accepted 10 April 2009
Available online 19 April 2009

Keywords:
Orthogonal projection to latent structures
Artificial neural network
Near-infrared spectroscopy
Diclofenac sodium
Degree of approximation
Partial least squares regression

ABSTRACT

A method for quantitative analysis of diclofenac sodium powder on the basis of near-infrared (NIR) spectroscopy is investigated by using of orthogonal projection to latent structures (O-PLS) combined with artificial neural network (ANN). 148 batches of different concentrations diclofenac sodium samples were divided into three groups: 80 training samples, 46 validation samples and 22 test samples. The average concentration of diclofenac sodium was 27.80%, and the concentration range of all the samples was 15.01–40.55%. O-PLS method was applied to remove systematic orthogonal variation from original NIR spectra of diclofenac sodium samples, and the filtered signal was used to establish ANN model. In this model, the concentration of diclofenac sodium was determined. The degree of approximation was employed as selective criterion of the optimum network parameters. In order to compare with O-PLS–ANN model, principal component artificial neural network (PC-ANN) model and calibration models that use different preprocessing methods (first derivative, standard normal variate (SNV) and multiplicative scatter correction (MSC)) of the original spectra were also designed. In addition, partial least squares regression (PLS) models were also established to compare with ANN models. Experimental results show that O-PLS–ANN model is the best.

1. Introduction

Near-infrared (NIR) spectroscopy has been proved to be a powerful analytical tool for analyzing a wide variety of samples that are used in agricultural, food, chemical and pharmaceutical industries [1–10], mainly due to its advantages over other analytical techniques, such as being expeditious, without destruction, low cost, being adaptable for almost all kinds of samples in all states and with little or no sample preparation. Frequently, the objective with this characterization is to determine the concentrations of different components in the samples. However, NIR spectra often contain serious systematic variation that is unrelated to the response data set, and the analyte of interest absorbs only in small parts of the spectral region. For solid samples this systematic variation is mainly caused by light scattering and differences in spectroscopic path length. Furthermore, the baseline and slope variations may often constitute the major part of the variation of the NIR spectra. The variation in X (a given data set) that is unrelated to y (the response set) may disturb the multivariate modeling and cause imprecise predictions for new samples. So the first step of a multivariate calibration based on NIR spectra is often to preprocess the original data.

For the preprocessing of NIR spectral data, conventional methods that are commonly used including smoothing, derivation, multiplicative scatter correction (MSC) and standard normal variate (SNV). These signal corrections are different cases of filtering, practical effect of the first derivative is that it removes an additive baseline. The second derivative removes also a multiplicative baseline. But the drawbacks of using derivatives are the inevitable change of the shape of the spectra and the noise is seriously enlarged. SNV and MSC remove both additive and multiplicative baseline variation without altering the shape of the spectra. Common for all these methods is that they do not require a response variable in the preprocessing step, which is a prerequisite when orthogonal projection to latent structures (O-PLS) method [11,12] is applied. Being a generally applicable preprocessing and filtering method, O-PLS provides a way to remove systematic orthogonal variation from a given data set X without disturbing the correlation between X and the response set y. Compared with the original data, because the orthogonal variation is removed by applying O-PLS method, the filtered data which is used as input data for the calibration model is simplified, thus the complexity of the calibration model is reduced and the predictive ability is preserved,

* Corresponding author. Tel.: +86 0431 85659459; fax: +86 0419 85156744.
E-mail address: ryl@jlu.edu.cn (Y. Ren).

0731-7085/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2009.04.014

effectively improved the interpretational ability of the model for both correlated and non-correlated variation in NIR spectra.

Artificial neural networks (ANNs) technique is considered one of the best approaches to non-linear calibration and fitting problem in every field of chemistry. The distinct characteristic of ANNs is their ability to learn from experience and examples and to get adapted with changing situations accordingly. In quantitative analysis, ANNs have been more and more widely applied during the past several years [13–20], mainly due to their anti-jamming, anti-noise and robust non-linear transfer ability. Generally, proper ANNs models result in lower mapping errors and prediction errors. They are an alternative for modeling non-linear data sets when the more classical multivariate calibration methods fail. ANNs also suffer from some drawbacks: the predictive properties of ANNs strongly depend on the learning parameters and the topology of the network, and ANNs models are complex and difficult to interpret.

2. Theory

2.1. Preprocessing methods

In order to simplify the interpretation of NIR spectral data, O-PLS uses the input data set X and the response set y to filter and remove variability in X that is orthogonal to y. The O-PLS preprocessing method with a single response set y is described as following:

1. Optional transformation, centering and scaling of the raw data to give the matrices X and y.

2. Calculation of the parameters w, t, p, u and c with the normal NIPALS method [21] for single y, where w represents the weight vector of X; t is the score vector of X; p is the loading vector of X; u is the score vector of y and c is the loading vector of y.

3. Calculation of weight vector of the orthogonal variation.

$$w_{ortho} = p - \left[\frac{w^T p}{w^T w}\right] w \tag{1}$$

Then make normalization of w_{ortho}, where w_{ortho} represents the weight vector of orthogonal variation.

4. Calculation of score vector and loading vector of the orthogonal variation and saving of found parameters.

$$t_{ortho} = \frac{X w_{ortho}}{w_{ortho}^T w_{ortho}} \tag{2}$$

$$p_o = \frac{X^T t_{ortho}}{t_{ortho}^T t_{ortho}} \tag{3}$$

t_{ortho} is the score vector of orthogonal variation and p_{ortho} is the loading vector of orthogonal variation.

5. Removal of orthogonal variation from X.

$$E_{\text{O-PLS}} = X - t_{ortho} p_{ortho}^T \tag{4}$$

$t_{ortho} p_{ortho}^T$ represents the matrix of orthogonal components, $E_{\text{O-PLS}}$ represents the residual matrix, for additional orthogonal components, return to step 2 and set, run the circle till the orthogonal variation $X = E_{\text{O-PLS}}$ does not exist in X.

After preprocessing with O-PLS method, the filtered data $E_{\text{O-PLS}}$ does not contain any variation that is orthogonal to y, so the stability of the calibration model is greatly improved.

SNV is a mathematical transformation method used to remove slope variation and to correct for scatter effects, and MSC corrects for difference in light scatter between samples before calibration. The SNV theory and MSC theory are described in Refs. [22–26] in detail.

2.2. Artificial neural networks

The current interest in artificial neural networks is largely due to their ability to mimic natural intelligence in its learning from experience [27]. They learn from examples by constructing an input–output mapping without explicit derivation of the model equation. Artificial neural networks are parallel computational devices consisting of groups of highly interconnected processing elements called neurons. Traditional neural networks have neurons arranged in a series of layers: input, hidden(s), and output layers. The layers work parallel in time, taking input from the previous layer and passing their output to the next layer in a synchronous manner at every time step. The number of neurons in the input layer and the output layer are determined by the number of input and output parameters, respectively. In order to find the optimal architecture, number of neurons in the hidden layer has to be determined (this number will be determined based on the ANN during the training process by taking into consideration the convergence rate, mapping accuracy, etc.). In each neuron, the sum of the weighted signals is calculated and when it overcomes a certain value, or threshold, it is processed by a so-called transfer function and sent to all neurons in the next layer, and during training, the weight coefficients and threshold values are adjusted to fit the training data. Of all the ANNs, the most widely used network type is multilayered feed-forward network [28,29] trained with the back-propagation (BP) learning algorithm [30–32]. The BP algorithm is based on the selection of a suitable error function, whose values are determined by the actual and predicted outputs of the network. The model with lowest prediction error is being used as the final and optimal model. Generally, the root mean squared error (RMSE) is used as the error function for finalizing the training and testing process [29].

2.3. Evaluation of artificial neural networks

The present criterion of optimization of the network is to minimize the performance error measured on the training set. However, it is very easy to choose an overfitting model, namely, the error of testing set is not at the minimum. This kind of network is unsteady when it is used to predict an unknown sample. To avoid this kind of situations, a new evaluation criterion of the network, the degree of approximation, is employed [33–35]. The definition of this criterion is given by Eqs. (5) and (6):

$$e_a = \left(\frac{n_1}{n}\right) e_1 + \left(\frac{n_c}{n}\right) e_c + |e_1 - e_c| \tag{5}$$

where e_a is the error of the approximation; e_1 and e_c are the relative standard errors of training set and validation set, n_1 and n_c are the numbers of samples in the training set and validation set, n is the number of all known samples, and n_1/n and n_c/n are the weights contributed to the error of approximation (e_a) by training set and validation set:

$$D_a = \frac{c}{e_a} \tag{6}$$

where D_a represents the degree of approximation and c is a constant number by which D_a is adjusted to get a good chart, here the value of c is set with 0.08. It is very obvious that the smaller e_a, the larger D_a can obtain the better ANN models. Therefore, the effects of both training set and validation set are considered in this evaluation criterion.

The predictive ability of calibration model for training set, validation set and test set are evaluated in terms of the relative standard error (RSE) [36,37], defined as:

$$\text{RSE} = \sqrt{\frac{\sum_{i=1}^{n} (C_{NIR_i} - C_{REF_i})^2}{\sum_{i=1}^{n} C_{REF_i}^2}} \tag{7}$$