Chemometrics and Intelligent Laboratory Systems 153 (2016) 51–57

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A new multivariate calibration model transfer method of near-infrared spectral analysis

Chen Liang [a], Hong-fu Yuan [b,*], Zhong Zhao [a], Chun-feng Song [b], Jia-jun Wang [c]

[a] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China
[b] College of Materials Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China
[c] China Tobacco Yunnan Industrial Co., Ltd., Kunming 650321, China

ARTICLE INFO

Article history:
Received 25 October 2015
Received in revised form 25 January 2016
Accepted 31 January 2016
Available online 23 February 2016

Keywords:
NIR spectroscopy
Calibration model transfer
Piecewise direct standardization
Rank-KS

ABSTRACT

Near-infrared (NIR) has been successfully applied for rapid and nondestructive analyses in diverse fields, but the issue of calibration model transfer has not been clarified thus far. Typically, the calibration model developed for one instrument cannot be applied directly to other instruments in many cases. To achieve calibration model transfer between the master and slave instruments, an improved piecewise direct standardization designated Rank–Kennard–Stone-PDS (Rank-KS-PDS) is proposed in this paper. An optimal sample selection method (Rank-KS) is proposed to select the transfer samples. The process of sample selection is based simultaneously on the distribution of spectral space and of property space. Thus, the samples selected by Rank-KS have greater representativeness and wider coverage. The Rank-KS-PDS method has been used to predict the alkaloid and glycoside content in tobacco. Comparative studies of calibration model transfer with the proposed method, random selection (RS), and KS have also been performed. For the prediction model of alkaloids, the comparative experiment results have verified the reduction in the number of transfer samples required from 12 to 8 and in the root-mean-square error of prediction (RMSEP) from 0.1440 to 0.1250 with the proposed method. For the prediction model of glycosides, the accuracy of model transfer was found to improve significantly, accompanied by a reduction in RMSEP from 1.6945 to 1.5850 without any increase in the number of transfer samples. With the proposed Rank-KS-PDS method, the selected transfer samples based on one property can be directly applied to the calibration model transfer of other properties with satisfactory results.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Near-infrared (NIR) has been successfully used to achieve rapid and nondestructive analysis of complex samples. Since the early 1990s, quality control and qualitative recognition techniques based on NIR analysis have developed rapidly in various fields, such as the petrochemical industry, the pharmaceutical industry [1], the military industry, food and drug administration [2], and agriculture [3], among others [4,5]. To date, however, the issue of NIR calibration model transfer has not been resolved. Typically, the multivariate calibration NIR model developed on the master instrument cannot be applied directly to the slave instrument in many cases. The relative spectral changes (10^{-3} to 10^{-4} A), crucial to NIR analysis, are so weak that a small difference between the spectra of the master and slave instruments can introduce an error. To obtain accurate measurement changes with time, the calibration model should be updated for NIR analysis even with the same instrument [6]. However, due to the high cost and limited resources, developing a multivariate model on each instrument is not feasible in practice [7,8]. As multivariate calibration models cannot be applied to different instruments and cannot adapt to aging equipment, the large-scale application of NIR analysis, such as future NIR networks in various industries, is considerably limited.

The main differences in spectra measured on different instruments include baseline drift, wavelength drift, and absorbance fluctuations. Thus, the spectra measured on different instruments can be made more consistent in several ways. First, the differences of the hardware of different instruments can be reduced, although achieving consistency of all instruments is not feasible in most cases. Second, the calibration model developed on the master instrument can be corrected to fit the slave instrument. Third, the predictions for spectra measured on the slave instrument can be corrected, such as slope/bias (S/B) [9]. In the S/B method, the predictions for spectra measured on the master and slave instruments are correlated linearly. Fourth, the spectra can be corrected, which can be achieved in two ways: one involves correcting the spectra obtained on the slave instrument to match the spectra obtained on the master instrument while the calibration model remains unchanged; the other involves correcting the spectra of the master instrument to match the spectra of the slave instrument, while a new model is additionally developed by the corrected spectra. This process is termed as calibration model transfer or instrument standardization. To date, several standardization methods,

* Corresponding author at: No. 15 of North Three-ring East Road, Chao Yang District, Beijing, 100029, China. Tel.: +86 13501215388.
E-mail address: yhf204@126.com (H. Yuan).

http://dx.doi.org/10.1016/j.chemolab.2016.01.017
0169-7439/© 2016 Elsevier B.V. All rights reserved.

such as S/B, direct standardization (DS) [10], piecewise direct standardization (PDS) [11] and Shenk's [12], and canonical correlation analysis (CCA) [13], among others [14,15], have successfully overcome analytical limitations. In most methods currently in use, the spectra of the slave instrument are generally modified to match the spectra of the master instrument. However, some of these methods cannot be applied in practice as they lack accuracy. Among these methods, PDS generally produces superior application results to others.

Several studies have been dedicated to improving the accuracy of model transfer. Some studies have focused on optimizing data preprocessing methods [16], and some have attempted to remove the outliers of the transfer set [17]. However, few studies have optimized the method of transfer sample selection.

In this work, an improved PDS designated Rank–Kennard–Stone-PDS (Rank-KS-PDS) is proposed to optimize the transfer sample selection. PDS primarily creates an effective transformation matrix *F*, which is directly related to the samples of transfer set. Thus, choosing a suitable transfer sample set is crucial to model transfer. The focus must be placed on the selection of transfer samples, as the process of model transfer can be invalidated if the transfer samples behave anomalously. At present, algorithms of KS [18] and random selection (RS) [19] are usually applied to select the transfer samples in PDS. However, the results of the RS algorithm are random. The selection criteria of KS are based on the differences in the spectra. However, the KS method only works at high concentrations, as the differences in the spectra are indistinguishable in the low concentration range. The selection criteria of the Rank–Select algorithm are based on the concentration distribution of the samples without considering the spectral differences. Thus, the samples selected by RS, KS, and Rank–Select lack representativeness.

In this work, based on the Rank-KS-PDS method, an optimal selection method (Rank-KS) [20] is used to select the transfer samples. The selection process of Rank-KS is based simultaneously on the distribution of spectral space and property space. The Rank-KS method can optimize transfer sample selection at a low concentration range where the KS method cannot. Compared with PDS, the accuracy of model transfer improves significantly and the number of transfer samples decreases. The comparative experiments showed that the Rank-KS-PDS method can be applied in practice.

2. Methods

The samples selected by RS, KS, and the Rank–Select method lack representativeness. Thus, in this study, the features of PDS and Rank-KS have been combined in an improved PDS algorithm termed as Rank-KS-PDS. Compared with other methods, the distribution of the sample in both spectral space and property space has been simultaneously considered in the transfer sample selection process of the proposed Rank-KS-PDS method. The samples selected by Rank-KS can cover a wide concentration range evenly and can be more representative.

2.1. Rank-KS algorithm

In the Rank-KS method, the selection of transfer samples is based on the distribution of spectral space and property space simultaneously. Thus, the samples selected by Rank-KS have greater representativeness and wider coverage.

The Rank-KS method comprises the following steps:

Step 1: The property values of the calibration samples are arranged in ascending or descending order. Then, all of the samples are divided into *m* parts:

$$D = \frac{y_{\max} - y_{\min}}{m} \tag{1}$$

$$Re_i \in [y_{min} + (i-1) * D, y_{min} + i * D]; \quad i \in [1, m] \tag{2}$$

where $y_{\min}$ and $y_{\max}$ are the minimum and maximum values of the properties, respectively; *D* is the length of every equal subinterval; and Re_i is the range of the *i*-th subinterval.

Step 2: The representative samples in every subinterval are selected by the KS algorithm [18].

Step 3: Assuming *n* number of transfer samples and *m* number of subintervals, *n*/*m* samples should be selected in every subinterval. If the number of samples in one subinterval is less than *n*/*m*, then all samples in this subinterval should be selected as the transfer set. If the number of all selected samples is less than *n*, then the remaining samples should be selected until the number of samples of the transfer set is *n*.

In the Rank-KS method, because the KS method was used to select the samples in each subinterval, the samples can cover the full concentration range evenly.

2.2. PDS algorithm

In DS [10], the spectra measured on the slave instrument are corrected to match the spectra measured on the master instrument while the calibration model remains unchanged. This correlation is described by the transformation matrix *F*:

$$S_m = S_s \times F \tag{3}$$

where S_m and S_s are the spectra of the samples measured on the master and slave instruments, respectively, and *F* is a square matrix dimensioned wavelengths by wavelengths. The transformation matrix *F* is calculated as follows:

$$F = \left(S_s^T S_s\right)^{-1} S_s^T S_m \tag{4}$$

The underlying principle of PDS and DS is the same. In DS, each spectral point $S_{m,i}$ of the master spectra is simultaneously related to the whole spectrum S_s measured on the slave instrument. In the real spectra, however, the spectral variations are typically limited to small regions. In PDS, each spectral point $S_{m,i}$ at the wavelength *i* on the master instrument is related to spectral measurements $S_{s,i}$ in a small window around the wavelength *i* measured on the slave instrument. The spectrum is transformed by moving the window sequentially through the entire wavelength range.

The steps of PDS are as follows:

Step 1: The subset spectra $s_{m,i}$ are chosen at wavelength index *i* on the master instrument.

Step 2: The subset spectra $S_{s,i}$, which are measured on the slave instrument at nearby wavelengths from index $i - k$ to $i + k$, are chosen:

$$S_{s,i} = \left[s_{s,i-k}, s_{s,i-k+1}, \cdots, s_{s,i+k-1}, s_{s,i+k}\right] \tag{5}$$

then a multiple regression model is developed as follows:

$$s_{m,i} = S_{s,i} \times b_i \tag{6}$$

where $S_{s,i}$ is a matrix dimensioned *num* by $2k + 1$, and the dimension of $s_{m,i}$ is *num* by *1*. *num* denotes the quantity of the transfer samples.

Step 3: The regression vector b_i is calculated via the PLS method. The number of latent variables to be calculated in PLS is denoted by *lv*, which can also be considered the number of transfer factors. Then the transformation matrix *F* is calculated by setting the off-diagonal elements to zero as follows:

$$F = diag\left(b_1^T, b_2^T, \cdots, b_i^T, \cdots b_n^T\right) \tag{7}$$

where *n* is the number of spectral channels included.

Download English Version:

https://daneshyari.com/en/article/7562700

Download Persian Version:

https://daneshyari.com/article/7562700

Daneshyari.com