Chemometrics and Intelligent Laboratory Systems 149 (2015) 22–27

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Designing classification filters for integrated sensing and processing using optimal discriminant vectors

Yan Liu, Wensheng Cai, Xueguang Shao *

Research Center for Analytical Sciences, College of Chemistry, Nankai University, Tianjin Key Laboratory of Biosensing and Molecular Recognition, State Key Laboratory of Medicinal Chemical Biology, Collaborative Innovation Center of Chemical Science and Engineering (Tianjin), Tianjin 300071, PR China

ARTICLE INFO

Article history:
Received 17 July 2015
Received in revised form 25 September 2015
Accepted 27 September 2015
Available online 3 October 2015

Keywords:
Integrated sensing and processing
Near-infrared spectroscopy
Discriminant vectors
Optical filters
Classification

ABSTRACT

Integrated sensing and processing (ISP) is a new strategy for instrument design to simplify quantitative or qualitative analysis. One of the ISP approaches is processing the optical spectrum with filters to obtain analytical results directly. ISP filters based on optimal discriminant vectors are designed in this study for the problem of classification. The method starts with performing principal component analysis (PCA) on the spectra of multi-class samples, and then constructs the optimal orthogonal discriminant vectors using the PCA scores by maximizing Fisher's discriminant function. Therefore, the filters for discriminating the samples can be obtained by transforming the loadings with the discriminant vectors. Applying the filters onto the spectra of new samples, the difference between samples of different class can be obtained and the difference can be used for discrimination of these samples. NIR datasets of vitamins, cephalosporins and Chinese patent medicines are used to test the performance of the filters. The results show that, for each of the datasets, the samples of different class can be correctly identified.

1. Introduction

Integrated sensing and processing (ISP) are receiving more and more attention for designing new instruments to simplify the procedures of spectral analysis. ISP aims to design and optimize sensing systems that integrate the traditionally independent units of sensing, signal processing, communication and modeling. By employing ISP, the quantitative or qualitative analysis within traditional sensing system can be achieved. One of the ISP approaches involves processing the optical spectrum with filters to calculate the analytical results about the samples in sensing stage directly [1]. Hieftje et al. reported that an organic solvent can be used as a filter for quantitative determination of the trace substances in solutions [2–4]. Myrick et al. designed a pair of optical filters to simulate the regression vector produced by PCA. Light passing through the paired filters produces an analog detector signal that is directly proportional to the chemical or physical property [5]. In later works, the design was improved and used in UV-visible and NIR spectroscopy [6–9]. The key of the design is based on the solid-state optical filters fabricated with thin films, termed multivariate optical elements (MOEs). On the other hand, Lodder et al. demonstrated that absorption filters can be constructed with different molecules and used as mathematical factors to generate a factor-analytic optical calibration in a high-throughput spectrometer, termed molecular factor computing (MFC) [10–12].

For either MFC or MOEs, a digital filter that describes the transmittance curve of the filter should be obtained before making the real filter. Although studies on MFC and MOEs were reported, few methods on designing the digital filter were proposed. In order to design the digital filter, Small et al. constructed multivariate calibration models through using Gaussian basis functions to extract relevant information from single-beam spectral data [13]. The widths and peak positions of the basis functions were optimized by genetic algorithm (GA). The method was used for quantitative analysis of glucose in mimic biological samples, and the results were found to be better than partial least-squares regression (PLSR).

Fisher linear discriminant analysis has been the widely used method to design filters for discriminant analysis. The method can effectively classify the samples from different classes [14]. It consists of a series of non-orthogonal discriminant vectors. Okada et al., however, have shown that orthogonal vectors are more powerful than the classical non-orthogonal ones, in terms of both discriminant ratio and mean error probability [15]. As a result, the optimal discriminant vectors are developed as an improvement of Fisher linear discriminant analysis [16–18]. The method may be a good choice to design the filters in MFC or MOEs.

Near-infrared (NIR) spectroscopy is a rapid and nondestructive analytical technique and has been widely used in food, agriculture, industrial production etc. The weak signals, broad and overlapped peaks of near-infrared spectra, however, contain not only chemical information

* Corresponding author at: College of Chemistry, Nankai University, Tianjin 300071, PR China. Tel.: +86 22 23503430, fax: +86 22 23502458.
E-mail address: xshao@nankai.edu.cn (X. Shao).

http://dx.doi.org/10.1016/j.chemolab.2015.09.015
0169-7439/

Table 1
Information of samples in the datasets.

Data	Medicine	Manu.	Label	Num.	Cal.	Pred.
1	Vitamin B1	1	Va1	18	12	6
	Vitamin B1	2	Va2	18	12	6
	Vitamin B2	3	Vb3	18	12	6
	Vitamin B2	4	Vb4	18	12	6
	Vitamin B4	5	Vc5	18	12	6
	Vitamin B12	6	Vd6	18	12	6
	Vitamin C	7	Ve7	18	12	6
	Vitamin C	8	Ve8	18	12	6
	Total			144	96	48
2	Cefaclor	1	Ca1	18	12	6
	Cefixime	2	Cb2	36	24	12
	Cefixime	3	Cb3	19	12	7
	Cefalexin	4	Cc4	24	16	8
	Cefalexin	5	Cc5	18	12	6
	Cefalexin	6	Cc6	18	12	6
	Total			133	88	45
3	CPM1	1	Pa1	45	30	15
	CPM1	2	Pa2	36	24	12
	CPM1	3	Pa3	35	23	12
	CPM2	4	Pb4	43	28	15
	CPM2	5	Pb5	48	32	16
	CPM3	6	Pc6	31	20	11
	CPM3	7	Pc7	53	35	18
	Total			291	192	99

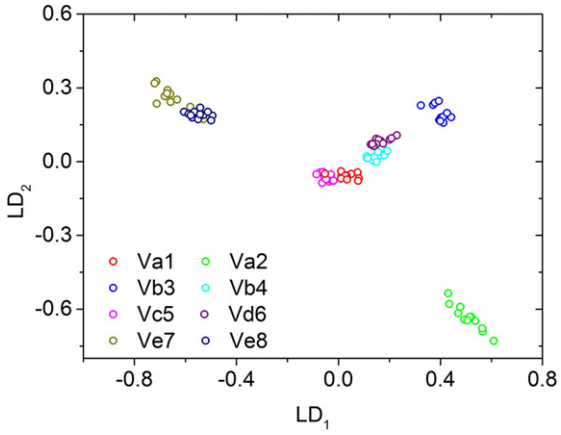

Fig. 2. Scores of the calibration samples in LD_1–LD_2 subspace.

but also noise and variant background. Therefore, to obtain an exact result of quantitative or qualitative analysis from the spectra, a reliable multivariate calibration model is indispensable [19,20]. The most commonly used methods to construct a multivariate calibration model, for quantitative and qualitative analyses, respectively, are partial least

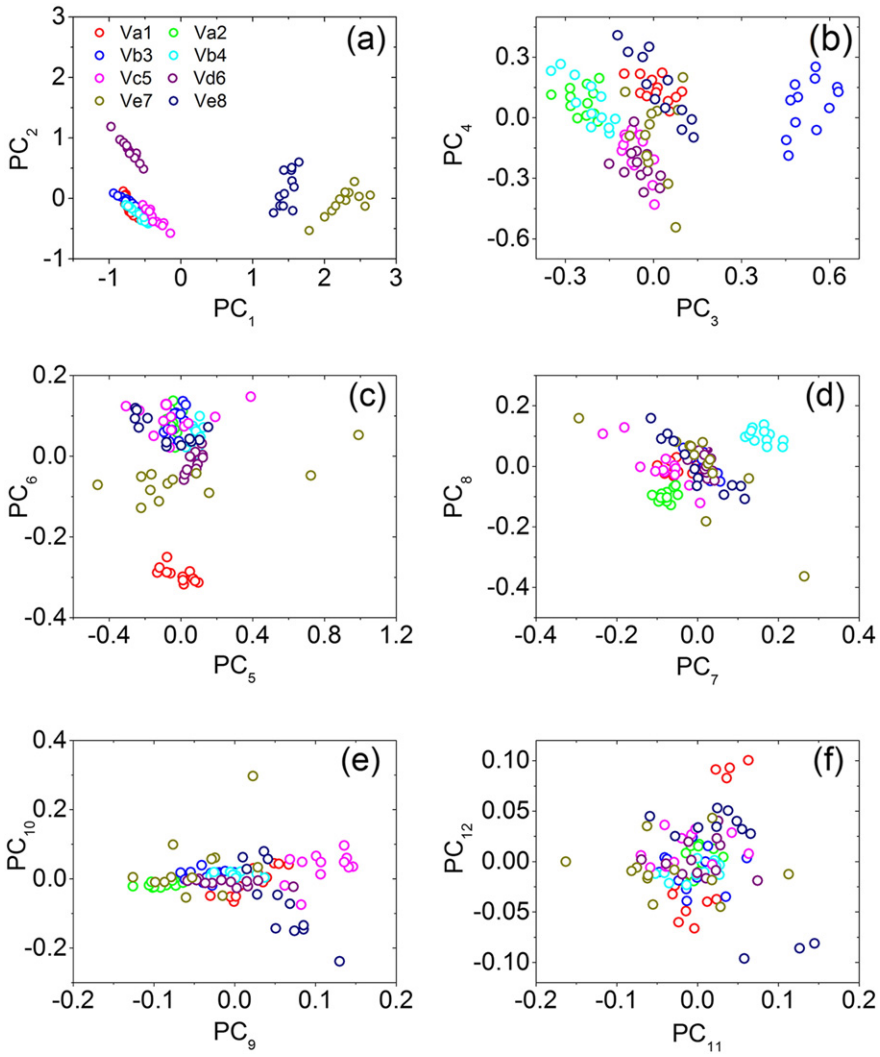

Fig. 1. Scores of the calibration samples in PC_1–PC_2 (a), PC_3–PC_4 (b), PC_5–PC_6 (c), PC_7–PC_8 (d), PC_9–PC_{10} (e) and PC_{11}–PC_{12} (f) subspaces.