Optik 157 (2018) 343–352

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Original research article

Low-rank based infrared spectral feature extraction framework for quantitative analysis

Yi Mou*, Long Zhou*, Hailin Huang, Weizhen Chen, Jijun Fan

School of Electrical and Electronic Engineering, Wuhan Polytechnic University, Wuhan 430023, China

ARTICLE INFO

Article history:
Received 22 September 2017
Accepted 13 November 2017

Keywords:
Low-rank
Nuclear norm
Feature extraction
Quantitative analysis

ABSTRACT

Feature extraction is a key problem in spectral analysis. Spectrum collected with spectrometer have latent low-rank component. If spectrum can be represented as a superposition of low-rank component and an approximation term, the spectrum feature is obtained. In this paper, a novel low-rank based infrared spectral feature extraction method is proposed. Employing a slide window to convert a single spectrum into a matrix, which can be decomposed as the superposition of a low-rank component and feature. In machine learning, nuclear norm is employed to approximate to low-rank minimization. Thus, the model can be written as a combination of the nuclear norm and an approximation term. We have proposed an efficient algorithm with singular value decomposition to the model. Solving the model, we obtain the latent low-rank component in spectrum. The feature is obtained via derivative of original and low-rank approximation. Then the quantitative analysis model is directly built with the feature. The advantage of proposed method is that extraction procedure of one spectrum is not affected by other spectrum. Extensive experiments are conducted with four public data sets and experimental results demonstrate that our proposed feature extraction method can lead to accuracy improvements over state-of-the-art methods.

1. Introduction

Feature extraction plays a central role in infrared spectral analysis [1–5]. The desire for efficient infrared feature extraction stems from two principal application areas: (1) classification and recognition [6,7], (2)quantitative analysis [8]. Thus, feature extraction becomes a very hot field in spectral analysis. There are three categories of infrared spectra feature, including local feature, global feature and transformed feature.

Local infrared spectral feature is geometrical feature of spectrum, includes location, height, width and area of peak. Chemical bond can be identified with location of peaks, height, width, especially area of peak are key feature for regression based quantitative analysis.

The simplest global feature is infrared spectrum itself, which can be employed to build a regression based quantitative analysis model with regularized least squares, principal component regression (PCR) and partial least squares (PLS).

Transformed feature includes filtering based feature and subspace learning feature. In infrared spectral analysis, high pass, low pass and wavelet filters are commonly utilized filtering methods. The simplest high pass filter is first or second order derivative, which can amplify detail hidden in original spectrum. But unfortunately, noise is also amplified. The sim-

* Corresponding authors.
E-mail addresses: mouyi@whpu.edu.cn (Y. Mou), profzh@126.com (L. Zhou).

https://doi.org/10.1016/j.ijleo.2017.11.078

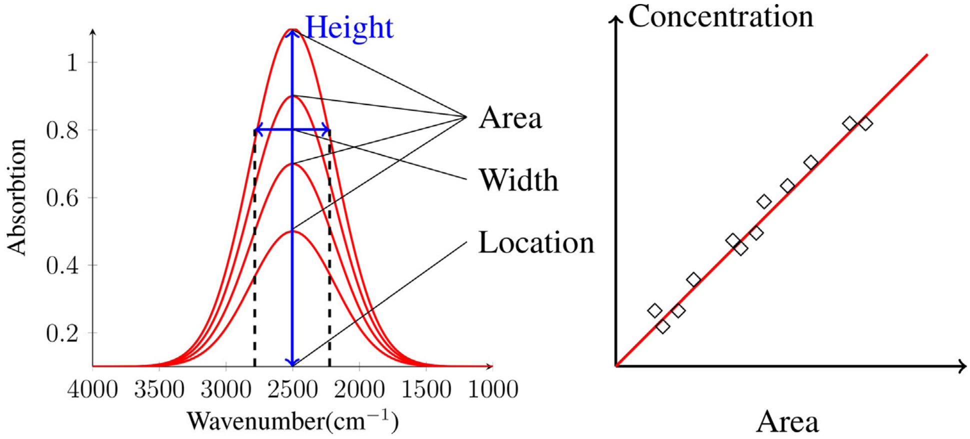

Fig. 1. Local feature.

plest low pass filter is average filter, which can remove noise. Meanwhile, some details are also removed. Wavelet is a more advanced tool than Fourier transform, which can be implemented with two-channel perfect construction filter banks. Yiuming Cheungemploy wavelet to extract high frequency to identify the characteristics of infrared spectra of herba epimedii icariin [9]. Subspace learning based feature extraction includes PCA [10,11], PLS-DA, LDA, CCA. Spectral classification can be realized with these features.

Data collected in many fields are low-ranked. Low-rank based model are widely utilized in machine learning, pattern recognition and image processing [12–14]. Thus, if low-rank component of spectra is extracted, then the feature is obtained. Previous feature extraction methods are not able to extract low-rank and row full rank components. Consequently, in some occasion quantitative analysis accuracies are low. Thus, we employ a low-rank approximation model to extracted the low-rank component in infrared spectra, and solving the model using singular value decomposition. In summary, our contributions are: (1) We have proposed a new low-rank based feature extraction method for infrared quantitative analysis. (2) Low-rank component and row full rank component of matrix formed by original spectra are obtained. (3) We then proposed an efficient solution algorithm to the proposed model using singular value threshold.

The reminder of the paper is organized as follows. We review general feature extraction methods in Section 2. We then formulate proposed low-rank based feature extraction model and provide an efficient algorithm for solving the proposed model in Section 3. Experiments and results analysis are provided in Sections 4 and conclusions are drawn in Section 5.

2. Related work

2.1. Local feature extraction

Local feature employed in infrared spectra analysis is demonstrated in Fig. 1. Locations of peaks are employed for identifying chemical bonds. For an example, peaks around 3300 cm^{-1} is related with —OH. The height, width and area can be employed for regression based quantitative analysis.

2.2. Transformed feature

First order derivative can be regarded as a high pass filter $h=\{1, -1\}$ with linear phase. Frequency response is $H(e^{j\omega})=e^{-j\omega}-e^{-2j\omega}$. Similarly, two point average filter can be regarded as a low pass filter $h=\{1, 1\}$ with linear phase [15,16]. Frequency response is $H(e^{j\omega})=e^{-j\omega}+e^{-2j\omega}$, which is shown in Fig. 2;

Wavelet can be implemented with the two-channel perfect reconstruction filter banks [17,18], as shown in Fig. 3. H_0, H_1, F_0 and F_1 are two low and two high filters, which satisfy perfect construction condition.

$$\begin{aligned} F_0(z)H_0(z)+F_1(z)H_1(z) &= 2z^{-L} \\ F_0(z)H_0(-z)+F_1(z)H_1(-z) &= 0 \end{aligned} \tag{1}$$

Principal component analysis (PCA) is a widely used feature extraction algorithm, which are widely employed in Chemometrics, pattern recognition, machine learning. It finds a direction vector $\boldsymbol{w}$, which can extract the maximum energy of data matrix $\mathbf{X}$. The model is expressed as [19]:

$$\begin{aligned} &\max \quad \boldsymbol{w}^T\mathbf{X}^T\mathbf{X}\boldsymbol{w} \\ &s.t. \quad \boldsymbol{w}^T\boldsymbol{w}=1 \end{aligned} \tag{2}$$