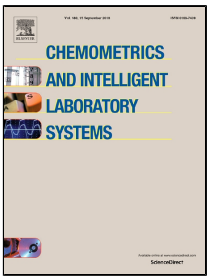
Accepted Manuscript



Collaborative representation based classifier with partial least squares regression for the classification of spectral data

Weiran Song, Hui Wang, Paul Maguire, Omar Nibouche

PII:	S0169-7439(18)30164-3
DOI:	10.1016/j.chemolab.2018.08.011
Reference:	CHEMOM 3671
To appear in:	*Chemometrics and Intelligent Laboratory Systems*
Received Date:	11 March 2018
Accepted Date:	19 August 2018

Please cite this article as: Weiran Song, Hui Wang, Paul Maguire, Omar Nibouche, Collaborative representation based classifier with partial least squares regression for the classification of spectral data, *Chemometrics and Intelligent Laboratory Systems* (2018), doi: 10.1016/j.chemolab.2018.08.011

Collaborative representation based classifier with partial least squares regression for the classification of spectral data

Weiran Song[a], Hui Wang[a], Paul Maguire[b], Omar Nibouche[a]

[a]School of Computing, [b]School of Engineering, Ulster University, BT37 0QB, Newtownabbey, Co. Antrim, UK

ABSTRACT

The need to classify high-dimensional spectral data is an increasingly common occurrence in rapid and non-destructive detection of object features and chemical species using spectroscopy. Partial least squares discriminant analysis (PLS-DA) is an effective method for spectral data classification, which is based on a multivariate regression model. Although powerful, PLS-DA suffers from performance degradation under complex conditions such as nonlinearity, imbalance and multiclass, which are common in real applications. Collaborative representation-based classifier (CRC) is a new machine learning algorithm which represents a query by a linear combination of training samples and classifies the query based on the representation. It offers the possibility of classifying even under nonlinearity, imbalance and multiclass conditions. In this paper, we present a novel method for spectral data classification, namely CRC-WPLS, which reaps the benefits of both PLS regression and CRC. This method searches for a weighted, linear combination of all training samples to represent the query by using PLS regression, and then assigns the query to the class which yields the least approximation error. CRC-WPLS is compared to PLS-DA, kernel PLS-DA and representation-based classifiers on fourteen general machine learning datasets and three spectral datasets. Experimental results show the proposed method can outperform 5 baseline methods in most cases, and achieve a high classification accuracy (> 92%) for low grade spectra obtained from portable instrumentation.

Keywords: Classification, Partial least squares, Collaborative representation, Spectral data.

1. Introduction

The combination of spectroscopy and chemometrics provides an effective tool for identifying the chemical compositions of a material in many fields such as food, pharmaceutical and biomedical science. It aims to reveal the qualitative or quantitative relationship between the high-dimensional spectra and corresponding identities by means of a classification model. Recently, the utilization of low-cost and portable spectrometers is gaining increasing attention for many field-based applications. However, variable environmental conditions and inevitable instrument limitations pose serious challenges for implementation