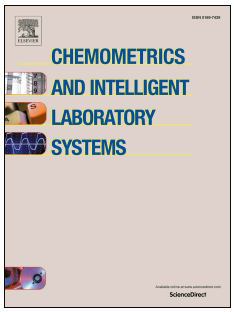
Accepted Manuscript



A selective review and comparison for interval variable selection in spectroscopic modeling

Li-Li Wang, You-Wu Lin, Xu-Fei Wang, Nan Xiao, Yuan-Da Xu, Hong-Dong Li, Qing-Song Xu

PII: S0169-7439(17)30367-2

DOI: 10.1016/j.chemolab.2017.11.008

Reference: CHEMOM 3540

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 29 May 2017

Revised Date: 7 September 2017

Accepted Date: 2 November 2017

Please cite this article as: L.-L. Wang, Y.-W. Lin, X.-F. Wang, N. Xiao, Y.-D. Xu, H.-D. Li, Q.-S. Xu, A selective review and comparison for interval variable selection in spectroscopic modeling, *Chemometrics and Intelligent Laboratory Systems* (2017), doi: 10.1016/j.chemolab.2017.11.008.

A Selective Review and Comparison for Interval Variable Selection in Spectroscopic Modeling

Li-Li Wang[1,†], You-Wu Lin[2,†], Xu-Fei Wang[3], Nan Xiao[1,4], Yuan-Da Xu[5], Hong-Dong Li[6], Qing-Song Xu[1,*]

[1]School of Mathematics and Statistics, Central South University, Changsha 410083, P. R. China
[2]School of Mathematics and Statistics, Guangxi Teachers Education University, Nanning 530023, P.R. China
[3] Department of Statistics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA
[4] Seven Bridges Genomics, 1 Main Street, Cambridge, MA 02142, USA
[5] Program of Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA
[6] School of Information Science and Engineering, Central South University, Changsha 410083, P.R. China

Abstract

Dimension reduction and variable selection are two types of effective methods that deal with high-dimensional data. In particular, variable selection techniques are of wide-spread use and essentially consist of individual selection methods and interval selection methods. Given the fact that the vibrational spectra have continuous features of spectral bands, interval selection instead of individual spectral wavelength point selection allows for more stable models and easier interpretation. Numerous methods have been suggested for interval selection recently. Therefore, this paper is devoted to a selective review on interval selection methods with partial least squares (PLS) as the calibration model. We described the algorithms in the five classes: classic methods, penalty-based, sampling-based, correlation-based, and projection-based methods. Finally, we compared and discussed the performances of a subset of these methods on three real-world spectroscopic datasets.

Keywords

Spectroscopy; PLS; interval variable selection

1. Introduction

In recent years, the extensive use of multivariate calibration methods in multi-component spectral analysis has made them extremely popular techniques, especially for vibrational spectral

† These authors contributed equally to this paper.
* Correspondence to: Q.-S. Xu. Email: qsxu@csu.edu.cn