Chemometrics and Intelligent Laboratory Systems 143 (2015) 136–145

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

A unified framework for contrast research of the latent variable multivariate regression methods

Zhangming He [a,b,*], Haiyin Zhou [a], Jiongqi Wang [a], Shouchao Zhai [c]

[a] College of Science, National University of Defense Technology, Yanwachi 47, 410073 Changsha, China
[b] Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, 47057 Duisburg, Germany
[c] Department of Automation, Tsinghua University, Beijing 100084, China

A R T I C L E I N F O

Article history:
Received 20 July 2014
Received in revised form 2 February 2015
Accepted 6 February 2015
Available online 14 February 2015

Keywords:
Latent variable multivariate regression
Principal component regression
Partial least square regression
Canonical correlation regression
Reduced rank regression
Calibration and prediction

A B S T R A C T

This paper focuses on contrast research of four latent variable multivariate regression (LVMR) methods, *i.e.*, principal component regression (PCR), partial least square regression (PLSR), canonical correlation regression (CCR) and reduced rank regression (RRR). The performances are evaluated by mean square error (MSE). A unified framework, called weight-framework, is proposed, where each LVMR method as well as the ordinary least square regression (OLSR) can be represented by a specific **W**eight matrix. Moreover, three theorems are proved delicately. The first one is coefficient theorem which reveals the relations between the coefficients estimated by the four LVMR methods and OLSR; the second one is MSE theorem which contrasts the calibration performances of the different methods; the third one is fault detection rate (FDR) theorem, which tells the different FDR when LVMR is applied for fault detection. Finally, two simulated data sets and one real data set collected from a benchmark system validate the correctness of our theoretical results.

1. Motivations and introduction

1.1. Motivations

Consider the general multivariate linear model

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{E}, \tag{1}$$

where $\mathbf{X} \in \mathbb{R}^{N \times n_x}$ is the explanatory data with N samples and n_x variables, $\mathbf{Y} \in \mathbb{R}^{N \times n_y}$ is the response data with N samples and n_y variables, $\mathbf{E} \in \mathbb{R}^{N \times n_y}$ is the noise, and $\beta \in \mathbb{R}^{n_x \times n_y}$ is the coefficient matrix which maps $\mathbf{X}$ onto $\mathbf{Y}$. Note that prior information about $\mathbf{E}$ is scarce, thus we make no more assumption about it. $\mathbf{X}$ and $\mathbf{Y}$ are centered, *i.e.*, subtracted from the mean values. Ordinary least square regression (OLSR) is a standard full rank regression method which is an unbiased estimation technique. OLSR works well if the calibration samples are sufficient and are not linearly correlated [5]. However, OLSR tends to estimate β with large uncertainty in the presence of sample deficiency and linear correlation [28].

Unlike OLSR, latent variable multivariate regression (LVMR) applies dimension reduction techniques for chemometric analysis [14,20], process monitoring [29], and prediction [31,11,6] and experimental design [6]. This paper mainly focuses on four LVMR methods, *i.e.*, principal component regression (PCR) [18], canonical correlation regression (CCR) [13], partial least square regression (PLSR) [12] and reduced rank regression (RRR) [24].

There are two questions with using LVMR:

1) Can a unified framework for LVMR be proposed, where all the four LVMR methods can be interpreted?
2) Can some theorems be developed, which reveal clearly the relation of the coefficients estimated, the calibration MSE and the detection performance by different LVMR methods and by OLSR?

As for the second question, there are few such theorems yet. As for the first question, there are some existing results, *e.g.*, an optimization function framework is proposed in [8], where CCR, RRR and PLSR are perfectly interpreted, however PCR is an exception; another example is a statistical framework based on maximal likelihood in [7], where both PCR to RRR are successfully interpreted, however PLS and CCR do not arise as members of the framework. To sum up, the existing frameworks are confined to the optimization functions of LVMR, but some other important aspects, *e.g.*, the latent variable (LV) structure and the unified regession procedure and performance contrast, are sometimes neglected.

1.2. Introduction

PCR is based on principal component analysis (PCA), which was developed by Harold Hotelling in the 1930s [15]. It is a technique for

* Corresponding author at: College of Science, National University of Defense Technology, Yanwachi 47, 410073 Changsha, China. Tel.: +86 18932487430.
E-mail address: hzmnudt@sina.com (Z. He).

http://dx.doi.org/10.1016/j.chemolab.2015.02.006
0169-7439/

finding LVs, also called principal components (PCs), from X with *maximal variance* [1]. PCA does not consider the information between X and Y, *e.g.*, the covariance information, this explains why PCR usually does not guarantee high calibration accuracy in the sense of the reconstruction mean square error (MSE).

CCR is based on canonical correlation analysis (CCA), which was first proposed in 1936 by Harold Hotelling [16]. It is a technique for finding the LV pair from X and Y with *maximal correlation* [23,26,10]. As introduced in [4], CCA is one of the most popular methods for investigating the relations between two sets of variables. Since the correlation information between X and Y is maximally extracted, CCR tends to give smaller calibration MSE, compared with PCR.

PLSR is based on partial least square (PLS), which was first proposed in 1975 by Herman Wold and then developed by Svante Wold [27]. PLS is a technique for finding the LV pair from $\mathbf{X}$ and $\mathbf{Y}$ with *maximal covariance*. PLSR is particularly suitable when $\mathbf{X}$ has more variables than observations, *i.e.*, $n_x > N$. PLS is frequently applied for analyzing data from near-infrared reflection (NIR) spectrometry in [22,11,20]. Note that the critical difference between PLS and CCA is that PLS extracts the LV pair with maximal covariance and CCA extracts the LV pair with maximal correlation [4]. Recently, an error-in-variable (EIV) data structure model is introduced in [21], where a modified PLS formulation are computationally more efficient than existing PLS-based approach.

RRR is based on redundancy analysis or the best rank approximation [25,8]. As introduced in [17], RRR is a technique for finding the LVs from $\mathbf{X}$ which can *optimally reconstruct* $\mathbf{Y}$ in terms of MSE. We will see in Theorem 2 Section 4.3 that RRR guarantees the highest calibration accuracy on condition that all LVMR methods extract the same number of LVs.

The rest of this paper mainly focuses on the three questions mentioned in Section 1.1 and it is organized as follows. In Section 2, we propose a unified framework, called weight-framework, for LVMR and performance evaluation. In Section 3, PCR, CCR, PLSR, RRR as well as OLSR are interpreted in weight-framework. In Section 4, three theorems about the coefficients, the MSE and the FDR are proposed and proved. In Section 5, two simulations and one practical benchmark case are used to validate the theory results.

1.2.1. Notations

This paper uses the standard notations. The transpose, the Moore–Penrose pseudo-inverse and the Frobenius-norm of matrix $\mathbf{A}$ are respectively denoted as $\mathbf{A}^{\mathrm{T}}$, $\mathbf{A}^{+}$ and $||\mathbf{A}||$; $\mathbf{A}_i$ denotes the ith column of A; I_n denotes the $n \times n$ identity matrix; $tr(\mathbf{A})$ denotes trace function of $\mathbf{A}$, defined to be the sum of the elements on the main diagonal; $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B}$ is a positive semi-definite matrix.

1.2.2. Acronyms

Several frequently-used acronyms are listed in Table 1.

Table 1
Acronyms.

Acronym	Meaning
LVs/LVMR	Latent variables/latent variable multivariate regression
PCA/PCR	Principal component analysis/principal component regression
CCA/CCR	Canonical correlation analysis/canonical correlation regression
PLS/PLSR	Partial least square/partial least square regression
RRR	Reduced rank regression
OLSR	Ordinary least square regression
MSE	Mean square error
QRD	QR decomposition
SVD	Singular value decomposition
GEVD	Generalized eigen-value decomposition
FDR	Fault detection rate

2. Weight-framework for LVMR and performance evaluation

A unified framework, called weight-framework, is proposed for LVMR and performance evaluation. We will see that each LVMR method can be represented by a special weight matrix.

LVMR is used to estimate the coefficient matrix in Eq. (1). Unlike ordinary least square regression (OLSR), the coefficient estimated by LVMR is biased and not full rank. Suppose that r_x denotes the rank of $\mathbf{X}$ and 'a' denotes the number of LVs with $a < r_x$.

For each LVMR method, LVs are extracted according to some specific criteria. LVs take the form of

$$\mathbf{T} = \mathbf{X}\mathbf{W}, \tag{2}$$

where $\mathbf{W} \in \mathbb{R}^{k \times a}$ is the weight matrix relying on the LV extraction criteria. Since any nonsingular transformation does not not change the spanned space of $\mathbf{T}$, we assume that $\mathbf{T}$ is orthonormal, *i.e.*,

$$\mathbf{T}^{\mathrm{T}}\mathbf{T} = \mathbf{I}_a. \tag{3}$$

Suppose that $\beta_{\mathbf{T}}$ is a linear mapping which maps $\mathbf{T}$ onto $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{T}\beta_{\mathbf{T}} + \mathbf{F}, \tag{4}$$

then the transpose of $\beta_{\mathbf{T}}$ is usually called 'loads' in PCA, CCA and PLS, and the ordinary least square estimate of $\beta_{\mathbf{T}}$ is

$$\hat{\beta}_{\mathbf{T}} = \left(\mathbf{T}^{\mathrm{T}}\mathbf{T}\right)^{-1}\mathbf{T}\mathbf{Y}. \tag{5}$$

From Eqs. (5), (3) and (2), we have

$$\hat{\beta}_{\mathbf{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \tag{6}$$

When Eq. (2) is put into Eq. (4), we have

$$\mathbf{Y} = \mathbf{X}\mathbf{W}\beta_{\mathbf{T}} + \mathbf{F}. \tag{7}$$

When Eq. (7) is compared with Eq. (1), it is reasonable to estimate β as follows

$$\hat{\beta} = \mathbf{W}\hat{\beta}_{\mathbf{T}}. \tag{8}$$

From Eqs. (8) and (6), we have

$$\hat{\beta} = \mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \tag{9}$$

When β is estimated, the calibration of $\mathbf{Y}$ is

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}. \tag{10}$$

From Eqs. (10) and (9), we have

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{Y}. \tag{11}$$

From Eqs. (11) and (2), we have

$$\hat{\mathbf{Y}} = \mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{Y}. \tag{12}$$

which means that $\hat{\mathbf{Y}}$ is the orthogonal projection of $\mathbf{Y}$ onto $\mathbf{T}$.

Like $\hat{\mathbf{Y}}$, the new response vector $\mathbf{y}_{new}$ can be predicted based on the new explanatory vector $\mathbf{x}_{new}$

$$\hat{\mathbf{y}}_{new} = \mathbf{x}_{new}\hat{\beta}. \tag{13}$$