Improved Partial Least Squares Regression Recommendation Algorithm

Liao Chunhua, Du Jianqiang, Jin Guohua, and Chen Chunlei

School of Computer Science, Jiangxi University of Traditional Chinese Medicine, Nanchang Jiangxi 330004 China

Abstract - This paper aims to improve the performance of partial least squares regression, and then, improve efficiency of its implementation. In this paper we provide a novel derivation based on optimization for the partial least squares (PLS) algorithm. The derivation shows that only one of either the X- or the Y- matrix needs to be deflated during the sequential process of computing latent factors. And then, based on this derivation, an improved recursive exponentially weighted PLS regression algorithm was proposed. And the improved algorithm is obviously superior to traditional PLS regression algorithm on performance.

Index Terms - partial least squares (PLS), kernel algorithm, algorithms improvement, recursive exponentially weighted algorithms

1. Introduction

Research in science and engineering often involves using some variables (commonly known as factors or prediction variables) which are easy-to-measure and/or controllable to explain, regulate, or predict other variables (commonly known as dependent variables or response variables). Given having many variables and ill-understood relationships, researchers tend to select appropriate variables to construct a good predictive model. However, a large part of the information contained in the data gets lost in the process of data collection. Interactions and correlations between factors are omitted, although they are of great important for the solution of problems. When the factors are few in number, are not significantly redundant(collinear), and have a well-understood relationship to the responses, methods based on Least-squares regression, such as multiple linear regression(MLR), principle component analysis(PCA),canonical correlation analysis (CCA) and so on, can be a good way to turn data into information. However, if any of these three conditions was broken down, these methods mentioned above can be inefficient or inappropriate. The damaging effects of multiple linear correlations in the regression model are summarized by Wang Huiwen [1]. So, algorithms and programs must be refined and optimized to meet the demands of research. We can learn from the thought of PCA and CCA, combined with theory of correlation between explanatory variables and the response variable in CCA and the method of component extraction in PCA. The method is partial least squares regression (PLSR).

PLS method was first presented by econometrician Herman Wold in 1966 and used for social science research [2]. A new multivariate statistical analysis method based on PLS was presented by S.Wold and C.Albano [3], i.e. partial least squares regression (PLSR). Subsequently PLSR measurement is widely used and further developed in many fields, such as chemistry, biology, medicine. Researches show that in the process of using multiple predictor variables to create response variable regression model, when there are high correlations between variables sets, compared to general multivariate regression, the analysis using PLSR is more effective, and the conclusion is more reliable[4,5,6,7].

To illustrate problems conveniently, the following notation is used. Uppercase bold variables are matrices and lowercase bold variables are column vectors. $\mathbf{X}$: predictor variable matrix (n×p), $\mathbf{Y}$: response variable matrix (n×q), $\mathbf{B_{PLS}}$: PLS regression coefficient matrix (p×q), $\mathbf{W}$:PLS weights matrix for $\mathbf{X}$ (p×m), $\mathbf{P}$: PLS loadings matrix for $\mathbf{X}$(p×m), $\mathbf{Q}$:Y matrix of PLS loadings matrix for $\mathbf{Y}$(q×m), $\mathbf{R}$: PLS weights matrix to calculation scored $\mathbf{T}$ directly from original matrix $\mathbf{X}$ (p×m), $\mathbf{T}$:PLS was score matrix of $\mathbf{X}$(n×m), $\mathbf{w_a}$: a column vector of $\mathbf{W}$, $\mathbf{p_a}$: a column vector of $\mathbf{P}$, $\mathbf{q_a}$: a column vector of $\mathbf{Q}$, $\mathbf{t_a}$: a column vector of $\mathbf{T}$, $\mathbf{r_a}$: a column vector of $\mathbf{R}$, p: number of $\mathbf{X}$-variables, q: number of Y- variables, n: number of samples, m: number of components in PLS model, h: integer counter for latent variable dimension.

2. Principle and Algorithm Research

In order to explain better the variability of the response variable, PLS regression must find some linear combination of the predictive variables space. However, it is not directly establish the linear regression model in PLS regression, but establish the linear regression model of response variables with latent variables of predictor variable. The model reflects the relationship between predictive variables and response variables indirectly. That is there are two groups of latent variables were extracted from predictor variables and response variables in one time, and they are linear combination of predictor variables and response variables respectively, and are called factors or components. There are two requirements must be meet in the extraction: ① 2 groups of latent factors were maximally expressed their variable information respectively; ② To maximize the covariance between corresponds to the latent factors of predictor variables and the latent factors of response variables, The mathematical model can be expressed as:

$$\max <X_1 w_1, Y_1 c_1> = w_1' X_1' Y_1 c_1 \quad s.t. \begin{cases} w_1' w_1 = 1 \\ c_1' c_1 = 1 \end{cases} \quad (1)$$

Using the classical NIPALS algorithm or kernel algorithm [8], we can get a PLS regression model. Hoskuldsson [9] discussed the calculation method of the latent factors vector in PLSR. The method is a continuous process as following. Firstly, calculate the latent factors of $\mathbf{X}$- and

Y-matrices separately, secondly, deflate the **X**- and **Y**-matrices by subtracting the computed latent factors from them in the next calculation. And then calculate the new latent factors using the deflated **X**- and **Y**- matrices. Note that we assume that X- and Y- matrices have been scaled to have zero column means so that no bias terms necessary. The detail description of typical PLSR algorithm is shown below:

① Calculated **w**, **t**, **q**, **u** and **p** separately with NIPALS algorithm, or kernel algorithms;

② Subtracting the latent factors from **X**- and **Y**-matrices respectively, then we can get the deflated matrices of **X**- and **Y**- matrices.

$$X_{h+1} = X_h - t_h p_h^T, \quad Y_{h+1} = Y_h - t_h q_h^T \qquad (2)$$

③ Calculated Predicted Residual Sum of Squares (PRESS) as below.

$$PRESS_{(h)} = \sum_{j=1}^{q} \sum_{i=1}^{n} (Y_{ij} - \hat{y}_{j(h)(-l)})^2 \qquad (3)$$

④ If these residuals are sufficiently small, i.e. **PRESS$_{(h)}$- PRESS$_{(h-1)}$**<ε ,whereε is predictive accuracy, or if the last dimension was insignificant according to cross validation (CV), go to step 5. Else back to step 1.

⑤ Build linear regression equation of **Y** on **r**

$$Y = t_1 r_1' + t_2 r_2' + \cdots + t_h r_h' + F \qquad (4)$$

⑥ Establish PLSR equation. Because **t$_h$** is the linear combination of **X**, so that we can get the PLSR equation by standard inverse transform.

$$y_j^* = a_{jp} x_1^* + \cdots + a_{jp} x_p^* + F_{Mj}, \quad j = 1, 2, \cdots, q \qquad (5)$$

F$_{Mj}$ is the j$_{th}$ column of the residual matrix F$_M$. Typically, in the PLS regression process, **X**- and **Y**- matrices is deflated in step 2 after calculating the latent factors.

3. Development of the Kernel Algorithm of the PLS Regression

Equations (2) are one of deflated process continually about the X- and Y- matrices. Therefore when the number of predictor variables and response variables is great, we should try to avoid the deflated operation of **X**- and **Y**- matrices in computation process. Especially when the number of latent variables is large, the operation will be calculated after several iterations. A proof is given that only one of either the **X**- or the **Y**- matrix in PLS algorithms needs to be deflated during the sequential process of computing latent vector[10].Also that

$$X_{h+1}^T Y_{h+1} = X_h^T Y_{h+1} = X_{h+1}^T Y_h, \quad X_{h+1}^T X_{h+1} = X_{h+1}^T X_h \qquad (6)$$

Kernel algorithm was proposed by Lindgren [8]. It has been proved that $X_{h+1}^T Y_{h+1}$ *and* $X_{h+1}^T X_{h+1}$ needed not to be deflate operation in PLS regression. Deflation of the **X**- and **Y**- matrices is given as equation (2). Thus, $X_{h+1}^T Y_{h+1}$ and $X_{h+1}^T X_{h+1}$ can be represented by (7) and (8) formula:

$$X_{h+1}^T Y_{h+1} = (I - w_h p_h^T)^T X_h^T Y_h \qquad (7)$$

$$X_{h+1}^T X_{h+1} = (I - w_h p_h^T)^T X_h^T X_h (I - w_h p_h^T) \qquad (8)$$

By equations (9) and (10), in the calculation of eigenvector extraction, only the need for iterative multiplication, instead of deflate operation of $X^T X$ and $X^T Y$.

Equations (9) and (10) included matrix multiplication in the deflating calculation; the computational cost is relatively high. De Jong [11] improved above algorithm as follows:

$$(X^T X)_{h+1} = (X^T X)_h - p_h p_h^T (t_h^T t_h) \qquad (9)$$

$$(X^T Y)_{h+1} = (X^T Y)_h - p_h q_h^T (t_h^T t_h) \qquad (10)$$

Therefore, matrix multiplication in deflation phrase can be accomplished by vector product iteratively, so as to decrease the cost of calculation. It is obvious that this improved kernel algorithms faster than the original kernel algorithm.
Kernel algorithm given as below:

① Computing covariance matrices $X^T X$ and $X^T Y$. Kernel matrix $X^T Y Y^T X$ can be calculated with $X^T Y$ multiplied by $(X^T Y)^T$.

② PLS weight vector w_h are calculated by the maximum eigenvalue corresponding eigenvector of $(X^T Y Y^T X)_h$:

$$w \propto (X\,YY\,X)\,w \qquad (11)$$

③ PLS load vector p_h and q_h is calculated as bellow:

$$p_h^T = \frac{w_h^T (X^T X)_h}{w_h^T (X^T X)_h w_h}, \quad q_h^T = \frac{w_h^T (X^T Y)_h}{w_h^T (X^T X)_h w_h} \qquad (12)$$

④ After extracting each latent factors, the residual between covariance matrices $X^T X$ and $X^T Y$ can be represented as equations (9) and (10).

In addition, A. Hoskuldsson [5] doesn't think must be degraded **Y**. finds that $X_h^T Y_h^T Y_h X_h$ and $X_h^T Y_1^T Y_1 X_h$ is equivalence in each calculation of eigenvectors w for covariance matrix $X_h^T Y_h^T Y_h X_h$. Hereby, Tenenhaus M [7], Lindgren [8], De Jong [12] and the Zhun Yun-hua [13] make SIMPLS algorithm, its mathematical model as below.

$$\max_{\|w\|=1} \sum_{j=1}^{q} \text{cov}(Y, X_h w_h) \qquad (13)$$

Advantages of this method are: Component $t_i = X w_i$ directly linked with the initial **X** and non deflated $X_i = X_{i-1} - t_{i-1} p_{i-1}^T$ [3], And the obtained results is easier to interpret. The solving process of **W** does not involve variables **X** and **Y** regression on components of $t_i = X w_i$, and save the computing on c_h, u_h and F_h, thus, the calculation process is simplified greatly, the calculation is simpler.

4. Kernel Algorithm Improvements

All PLS algorithm is optimized for different types of problems. Because there are too many regression coefficients, the importance analysis of predictive variables or factors is very difficult. In order to meet the real-time modeling online system demand, timely, efficient processing of large amounts

of data relates to the system modeling, to reduce the computation time, and reduce the need for computer memory, there is a need for a new method based on kernel algorithm and weighted recursive of SIMPLS algorithm, PLS algorithm is put forward to solve this kind of problem. The method effectively uses the calculated results to make recursive computation; amend the new model parameters in real time, thus obtain perfect parameter estimation value, and then meet the real-time modeling system requirements. B.hupinder and S.Dayl [10] have demonstrated a relationship between score matrix $\mathbf{T}$ and predictor matrix $\mathbf{X}$ can be represented by the following equation.

$$t_h = XR = X \prod_{j=1}^{h-1} (I - w_j p_j^T) w_h \tag{14}$$

By literature [14], there exist loop compute relationship between computing score vector t_h and u_h, load vector p_h, q_h, w_h, c_h and the regression coefficient r_i, just know one of them, we can infer other vectors.

$$r_1 = w_1 \tag{15}$$

$$r_i = w_i - \sum_{j=1}^{i-1} p_j^T w_i r_j, \quad i > 1 \tag{16}$$

Considering the strength, a key step which calculation speed is determined in kernel algorithm is construction of matrix $X^{\mathrm{T}}X$ and $X^{\mathrm{T}}Y$. By equation (9), construction of $X^{\mathrm{T}}X$ is not required. If the sample number is much larger than the number of variables ($n >> k$) in $\mathbf{X}$- matrix, $X^{\mathrm{T}}X$ can be easily calculated. Because the storage of $X^{\mathrm{T}}X$ takes up less than storage of $\mathbf{X}$.

For regular data collection tasks, the ideal approach is recursively update PLS model using each new multivariate as is becomes available. This process may be slowly changing over time. One would like to weight recent data more heavily and discount past data in an exponentially weighted manner. Using algorithm with the covariance updating equations, one can obtain a fast kernel algorithm for recursive exponentially weighted updating of a PLS regression model.

$$(X^T X)_{h+1} = \lambda_t (X^T X)_h + x_t^T x_t \tag{17}$$

$$(X^T Y)_{h+1} = \lambda_t (X^T Y)_h + x_t^T y_t \tag{18}$$

The past data in the covariance matrices are being exponentially discounted with a forgetting factor $\lambda_t (0 < \lambda_t \leq 1)$ and the recent data are being added in each new sampling period. For $\lambda_t = 1$ no discounting of the old data is done. As the covariance matrices in equation (17) and (18) are updated with very little computational effort, this algorithm will be extremely fast in these applications. Dayal and MacGregor[15,16] apply this algorithm to the adaptive multivariable control of a simulation continuous stirred tank reactor and to the updating of an online multi-output prediction model for an industrial mineral flotation circuit.

The description of PLS algorithm for recursive weighted of is as below:

① Calculate covariance matrices $X^{\mathrm{T}}X$, this step is optional.

② If the variable number of $\mathbf{Y}$ is less, computed q_h as eigenvectors corresponding to the largest eigenvalue of $(Y^T X^T X Y)_h$, w_h can be calculated by the following formula.

$$w = (XY)q, \quad w = \frac{w}{|w|} \tag{19}$$

When the variable number of $\mathbf{X}$ are less, computed w_h as eigenvectors corresponding to the largest eigenvalue of $(X^T Y Y^T X)_h$, as shown in equation (10).

③ The calculation formula for r_h with formula (16)

④ The calculation of load vectors p_h and q_h.

Only once computation, one can construct covariance matrix $X^T X$ with original matrix $\mathbf{X}$, and then applied in all dimensions.

$$p_h^T = \frac{r_h^T (X^T X)}{r_h^T (X^T X) r_h}, \quad q_h^T = \frac{r_h^T (X^T Y)_h}{r_h^T (X^T X) r_h} \tag{20}$$

⑤ to update the covariance matrix $X^T Y$.
By equations (9), (10) and (17) (18) have

$$(1-\lambda)(XY) = x y + p q (t t) \tag{21}$$

$$(1-\lambda)(XX) = x x + p p (t t) \tag{22}$$

⑥ store w, p, q and r in W, P, Q and R:

$$W = [w_1 \, w_2 \cdots w_m], P = [p_1 \, p_2 \cdots p_m] \tag{23}$$

$$Q = [q_1 \, q_2 \cdots q_m], R = [r_1 \, r_2 \cdots r_m] \tag{24}$$

⑦ If $PRESS_{(h)}$- $PRESS_{(h-1)} < \varepsilon$, ε is predictive accuracy, then go to step 8; Else turn to step 2 to calculate next latent factors;

⑧ As calculating the latent factors, PLS model regression coefficients (factor) are calculated by equation (25).

$$B_{PLS} = RQ^T \tag{25}$$

Simulation using MATLAB® environment, generated 1000 samples by a random function RAND as experimental data in systems analysis. Compared with traditional kernel algorithm, improved recursive weighted algorithm has improved significantly for calculation speed. It shown that recursive weighted PLS algorithm is suitable for real-time modeling and analysis process of mass-data.

5. Discussion

Compared to typical PLS algorithm, each new PLS algorithm has been almost as revolutionary [17]. To prove new algorithm is more useful than the typical PLS algorithm, modeling process repeatedly must be considered, which involved such as cross validation [18, 19], variable select [20] and missing data in sample data, and so on. Cross-validation techniques can help to select appropriate number of latent

factors in case of without knowing the number of latent variables, and then ensure the accuracy of model and efficiency of calculation. Rannar [21] research the problem of missing data in kernel algorithm proposed by Lindgren. They used methods is the EM algorithm [22]. All these key measures can accelerate the speed of repeated modeling [23]. These techniques have been described in above literature; it is just not talk about here.

Operation of matrices deflation in duplicated modeling must be considered in improvement of PLS algorithm. Hoped that regardless of increase or decrease in number of samples or variable; it does not need to compute the variance/covariance and adjoin matrix repeatedly.

In specific applications, forgetting factor $\lambda_t (0 < \lambda_t \leqslant 1)$ was introduced for weaken the influence of old data, the measure improves robustness of recursive algorithm, speeds algorithm convergence, and makes the estimate parameters more reasonable. Usually determine the overall data range based on other ways of handling advance [24].

6. Conclusions

All improvements to the PLS algorithm, is designed to fit a specific task. To select a good algorithm, applied first to understand what are the application fields? And what do you wanted? After answer these questions, then select the appropriate algorithm. For example, when the number of samples is far greater than the variable, and the data is updated in real time of the modeling process, you can select a weighted recursive PLS algorithm. At this point, as long as the covariance matrix $X^T X$ is calculated once, and then can be used in all subsequent calculations of p and q. When predictor matrix is smaller, may directly matrix to calculate the p_h and q_h by **X**. In short, in the PLS algorithm, we deflated only in one of **X**- or **Y**- matrices. The new algorithm Also provides a recursively update PLS model and weighted on the old data processing in order to weaken the influence of old data. Compared in terms of executed time for the calculations of latent factors, cross validation and handing errors between the new algorithm and the traditional kernel algorithm, the new algorithm have displayed a certain degree of superiority.

References

[1]   Wang Huiwen. Zhu Yunhua. PLS Regression in the Function of Eliminating multi-correlation. Application of Statistics and Management. Beijing. China. 1996, 15(6): 48-52.

[2]   Wold S. Partial least squate in Ess. New York: Wiley, 1985: 81-591.

[3]   Wold S, Algano C, Dunn M , et al. Pattern regression finding and using regularities in multivariate data. Landon: Analysis Applied Science Publication, 1983.

[4]   Wold S. Modeling data labels by principal component and PLS: class patterns and quantitative predictive relations. Analysis, 1984, 12: 477-485.

[5]   Hoskuldsson A. PLS regression methods. Journal of Chemometrics, 1988, 2(3): 211-228.

[6]   Geladi P, Qkoulaski B. Partial least squares regression: a tutorial. Analytical Chemical data, 1986, 35: 1-17.

[7]   Tenenhaus M, Gauchi J P, Menardo C. Regression PLS et application. Revue de statistique appliquce 1995, 53(1): 7-63.

[8]   Lindgren F, Geladi P, Wold S. The kernel algorithm for PLS. J Chemometrics, 1993, 7: 45.

[9]   Hoskuldsson A. A combined theory for PCA and PLS. J Chemometrics, 1995, 9: 91.

[10] Hupinder B, Dayal S, John, Macgregor F. Improved PLS algorithms. Journal of Chemometrics, 1997, 11: 73-85.

[11] Jong S D, Braak C J F T.JChemometrics, 1994, 8: 169.

[12] Jong S D. SIMPLS: an alternative approach to partial least squares regression. Chemometrics Intelligent Laboratory Systems, 1993, 18: 251-263.

[13] Zhu Yunhua, Wang Huiwen, Yang Xianglong. A simplified algorithm of PLS regression. Journal of Systems Science and Systems Engineering, 2000, 19(4): 414-419.

[14] Wang Huiwen. Partial Least-Squares Regression-Method and Applications. Beijing: National Defence Industry Press, 1999.

[15] Dayal B S. Department of chemical engineering. McMaster: McMaster University, 1996.

[16] Dayal B S, MacGregor J F. J Process Control, 1996.

[17] Bush B L, Nachbar J R B. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. J Compute-Aided Mol, Design, 1993, 7: 587-619.

[18] StoneM. Cross-validatory choice and assessment for statistical predictions. J Royal Stat, Soc B, 1974, 36: 111-133.

[19] Geisser S. A predictive approach to the random effect model. Biometrika, 1974, 61: 101-107.

[20] Baroni M, Costantino G, Riganelli D, et al. Generationg optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handing 3D QSAR problems. Quant Struct-Act Relat, 1993, 12: 9-20.

[21] Rannar S, Geladi P, Lindgren F, et al. A PLS kernel algorithm for data sets with mahy variables and less objects: part 2.Cross- validation, missing data and examples. J Chemometrics, 1995, 9: 459-470.

[22] Little R J A, Rubin D B. Statistical analysis with missing data. New York: Wiley, 1987.

[23] Fredrik L, Stefan R. Alternative partial least-squares (PLS) algorithms. Perspectives in Drug Discovery and Design, 1998(12/14): 105-113.

[24] Liu Qiang Yin Li. Study on an improved recursive partial least- squares modeling approach and application. Journal of Beijing University of Aeronautics and Astronautics. 2003, 29(7): 640-643.