Multi-Kernel Partial Least Squares Regression based on Adaptive Genetic Algorithm

Shaowei Liu [1, a], Jian Tang [2,b] , Dong Yan[1, b]

[1]Unit 92941, Huludao, Liaoning, China

[2] Northeastern University, Shenyang, China

[a]powernature@126.com, [b] frogdongdong@126.com

Keywords: Multi-kernel learning (MKL), Kernel partial least squares (KPLS), Adaptive genetic algorithm (AGA).

Abstract. Kernel learning has been a focus of machine learning domain recently. Kernel partial least squares (KPLS) algorithm can construct nonlinear model using extract latent variables from the input and output data space simultaneously. However, generalization performance of KPLS model relies mostly on kernel types and kernel parameters, which are difference to modeling of different applicable background. Intelligent optimization algorithm can be used to search these parameters. Thus, a new multi-kernel partial least squares regression approach based on linear multi-kernel construction method and adaptive genetic algorithm (AGA) is proposed in this paper. Normally used global and local kernels are weighed to obtain the mixed multi-kernel of KPLS algorithm. These kernel's parameters and weighting coefficients are selected using AGA optimization algorithm. The experimental results based on Benchmark data set show that the proposed approach has better prediction performance than that of single kernel based modeling method.

Introduction

Kernel learning method has been successfully applied to data-driven based soft sensor problems. However, how to select effective kernel types, kernel parameters for different modeling data sets with different background is still an open issue. Recently, many methods have been proposed to obtain multi-kernel instead of using a single one. They are called multi-kernel learning (MKL) approaches [1]. MKL method has been applied successful in biology information domain [2]. Recent applications have shown that using multi-kernel instead of a single one can enhance interpretability of the decision function and improve performances of the prediction model [3,4]. But MKL formulation is actually a convex but non-smooth minimization problem [5]. Thus, many research works have been focused on MKL optimization problem [6].

For most of the modeling data, there is co-linearity among input variables, especially for high dimensional spectral data, such as NIR and frequency spectral data. The popularly used feature extraction method is principal component analysis (PCA). However, PCA don't take into account correlation between inputs and outputs variables. Partial least squares (PLS) have been used widely on modeling such data due to its special characteristic. Namely, it can extract information from input and output data using less latent variables simultaneously [7]. In order to build more effective non-linear PLS model, kernel learning theory has been applied to PLS. Thus, kernel PLS (KPLS) was proposed in [8]. The idea of KPLS is based on a nonlinear mapping of the original data to a high-dimensional feature space, where a linear PLS model is constructed in that feature space. For solving optimal kernel parameters and features selection problem jointly for high dimensional data, adaptive genetic algorithm-kernel partial least squares (AGA-KPLS) based feature selection approach is proposed in [9]. However, how to selected effective kernel type cannot be solved, which is an open issue.

Therefore, multi-kernel partial least squares (MKPLS) regression modeling based on adaptive genetic algorithm (AGA) is proposed in this paper to address above problems.

Partial Least Squares

Partial Least Squares (PLS) can capture maximal covariance between the input and output variables using less latent variables (LVs). It is realized by decompose the input and output data simultaneously. At first, the predictor variables $\mathbf{X}$ and response variables $\mathbf{Y}$ are normalized as $\mathbf{E}_0$ and $\mathbf{F}_0$ wit zero mean and one variance. Let $\mathbf{t}_1$ be the first latent score vector of $\mathbf{E}_0$, $\mathbf{t}_1 = \mathbf{E}_0\mathbf{w}_1$, and $\mathbf{w}_1$ be the first axis of the $\mathbf{E}_0$, $\|\mathbf{w}_1\| = 1$. Similarly, let $\mathbf{u}_1$ be the first latent score vector of $\mathbf{F}_0$, $\mathbf{u}_1 = \mathbf{F}_0\mathbf{c}_1$, and $\mathbf{c}_1$ be the first axis of the $\mathbf{F}_0$, $\|\mathbf{c}_1\| = 1$. Then, we want to maximize the covariance between $\mathbf{t}_1$ and $\mathbf{u}_1$ with PLS algorithm. Thus, PLS algorithm needs to solve the following problem:

$$\mathrm{Max}(\mathbf{E}_0\mathbf{w}_1, \mathbf{F}_0\mathbf{c}_1)$$
$$\mathrm{s.t.}\begin{cases} \mathbf{w}_1{}^{\mathrm{T}}\mathbf{w}_1 = 1 \\ \mathbf{c}_1{}^{\mathrm{T}}\mathbf{c}_1 = 1 \end{cases} \tag{1}$$

The above problem can be solved with traditional Lagrange approach. At last, we obtain that $\mathbf{w}_1$ and $\mathbf{c}_1$ are the maximum eigenvector of matrix $\mathbf{E}_0^T\mathbf{F}_0\mathbf{F}_0{}^T\mathbf{E}_0$ and $\mathbf{F}_0{}^{\mathrm{T}}\mathbf{E}_0\mathbf{E}_0^{\mathrm{T}}\mathbf{F}_0$. Then, latent features $\mathbf{t}_1$ and $\mathbf{u}_1$ are obtained, which capture most of the information in $\mathbf{X}$ and $\mathbf{Y}$. Deflate $\mathbf{E}_0$ and $\mathbf{F}_0$ with $\mathbf{t}_1$ and $\mathbf{u}_1$ and their loading to obtain residual matrixes $\mathbf{E}_1$ and $\mathbf{F}_1$. Using the same method to $\mathbf{E}_1$ and $\mathbf{F}_1$, and repeat until the final residual matrixes $\mathbf{E} = \mathbf{F} = 0$. At last, the PLS outside model and inside model can be written as a multiple regression model:

$$\mathbf{Y} = \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{U}(\mathbf{T}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{U})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y} + \mathbf{G} \tag{2}$$

where, $\mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_{\mathrm{h}}]$ and $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_h]$ are score matrices; h is number of the latent variable (LV), and $\mathbf{G}$ is residual matrix.

Kernel Partial Least Squares

Kernel partial least squares (KPLS) constructs nonlinear modeling by extend nonlinear item to input matrix $\mathbf{X}$. At first, input data $\{(\mathbf{X})_l\}_{l=1}^{k}$ is nonlinear mapping to the high dimension features space, $\Phi : (\mathbf{X})_l \to \Phi((\mathbf{X})_l)$. Then, linear PLS algorithm is performed in this high dimension feature space. At last, nonlinear model in the original input space is obtained.

Kernel trick is used to realize the mapping from the original input space into the high dimensional feature space. Kernel matrix is calculated with:

$$\mathbf{K} = ((\mathbf{X})_l)^{\mathrm{T}}\Phi((\mathbf{X})_m) = \mathbf{K}(\mathbf{X}, \mathbf{X}'), \qquad l, m = 1, 2, \cdots k \tag{3}$$

The kernel matrix $\mathbf{K}$ is meaned as:

$$\tilde{\mathbf{K}} = (\mathbf{I} - \frac{1}{k}1_k 1_k^{\mathrm{T}})\mathbf{K}(\mathbf{I} - \frac{1}{k}1_k 1_k^{\mathrm{T}}) \tag{4}$$

where, $\mathbf{I}$ is unite matrix of k dimension; 1_k is a vector with value 1 and length k.

Prediction output of KPLS for the training samples can be represented as:

$$\hat{\mathbf{Y}} = \tilde{\mathbf{K}}\mathbf{U}(\mathbf{T}^{\mathrm{T}}\tilde{\mathbf{K}}\mathbf{U})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y} \tag{5}$$

For the testing samples $\{(\mathbf{X}_{\mathrm{t}})_l\}_{l=1}^{k_{\mathrm{t}}}$, it must be scaled as:

$$\tilde{\mathbf{K}}_{\mathrm{t},} = (\mathbf{K}_{\mathrm{t}}\mathbf{I} - \frac{1}{k}1_{kt}1_k^{\mathrm{T}})\mathbf{K}(\mathbf{I} - \frac{1}{k}1_k 1_k^{\mathrm{T}}) \tag{6}$$

where, $\mathbf{K}_{\mathrm{t}}$ is kernel matrix of the testing samples, $\mathbf{K}_t = \mathbf{K}((\mathbf{X}_{\mathrm{t}})_l, (\mathbf{X})_m)$; k_{t} is number of the testing sample; 1_{kt} is a vector with value 1 and length k_{t}.

Prediction output of KPLS for the testing samples can be represented as:

$$\hat{y}_{\mathrm{t}} = \tilde{\mathbf{K}}_{\mathrm{t}}\mathbf{U}(\mathbf{T}^{\mathrm{T}}\tilde{\mathbf{K}}\mathbf{U})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y} \tag{7}$$

Linear Multi-Kernel Learning

The main idea of linear MKL is to create a weighted linear combination kernel of different types. Moreover, it was shown that more effective soft sensor modeling can be obtained by using multiple kernels instead of one. In this paper, the following mix-kernel is constructed for KPLS model:

$$\mathbf{K}_{\text{mix}}(\mathbf{X},\mathbf{X}') = \lambda\mathbf{K}_{\text{rbf}}(\mathbf{X},\mathbf{X}') + (1-\lambda)\mathbf{K}_{\text{poly}}(\mathbf{X},\mathbf{X}') \tag{8}$$

where,

$$\mathbf{K}_{\text{rbf}}(\mathbf{X},\mathbf{X}') = \exp(\frac{-\|\mathbf{X}-\mathbf{X}'\|}{2\sigma^2}) \tag{9}$$

$$\mathbf{K}_{\text{poly}}(\mathbf{X},\mathbf{X}') = (\mathbf{X}\cdot\mathbf{X}'+1)^d, d=1,2,\cdots,N \tag{10}$$

where $\mathbf{K}_{\text{rbf}}$ and $\mathbf{K}_{\text{poly}}$ are polynomial and Gaussian kernel function respectively; $\lambda(0\le\lambda\le1)$ is weighing coefficient which used to adjust the local kernel $\mathbf{K}_{\text{rbf}}$ and global kernel $\mathbf{K}_{\text{poly}}$.

Adaptive Genetic Algorithm based Multi-Kernel Partial Least Squares

For MKPLS modeling algorithm, there are three parameters should be selected, which are denoted as $\{\sigma,d,\lambda\}$. We use adaptive genetic algorithm (AGA) to select them together, and normally used binary encoding system is used to encode and decode them. AGA can adjust probabilities of crossover and mutation according to fitness function automatically [10].

Objective of AGA algorithm is to minimize error between KPLS model's output and real value by select optimized parameters set $\{\sigma,d,\lambda\}$. The following fitness function is used:

$$\text{Fitness}(\sigma,d,\lambda) = \sqrt{\frac{1}{k}\sum_{l=1}^{k}[\mathbf{y}(l)-\hat{\mathbf{y}}(l)]^2} \tag{11}$$

where, k and $\hat{y}(l)$ are numbers and prediction value of the validating sample.

Experiments

Concrete Compressive Strength data set from UCI platform [11] is used to build MKPLS based soft sensor models. Studies show that concrete compressive strength is a highly nonlinear function of age and ingredients. These ingredients include cement, blast furnace slag, fly ash, water, superplasticizer, coarse aggregate, and fine aggregate.

In our experiments, the original data sets are parted into two parts with equal interval. AGA based optimization method is applied to select the optimal MKPLS model's parameters. Leave-one-out cross validating method is used to obtain the final model. The parameters of AGA algorithm are: Population size 80, Maximum generations 20, Percent at convergence 98 and Percent initial terms 40.

Root mean square error (RMSE) of testing samples is used to evaluate the model's prediction performance. The proposed method is compared with principal component analysis-support vector machines (PCA-SVM) and KPLS approach in [12].

Prediction curves and statistical results with the optimal modeling parameters are shown in Fig.3. Statistical results of different approaches are shown in Table .1.

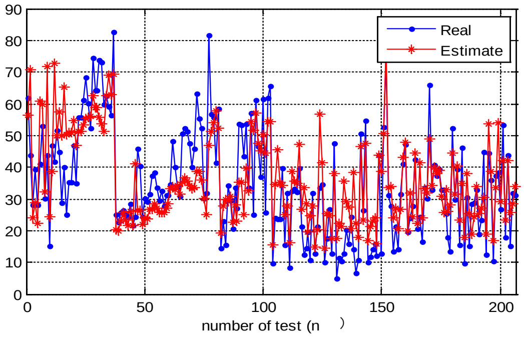

Fig. 1. Prediction curves of the proposed approach

Table. 1. Statistical results of different approaches

	Parameters			Weights		Testing error
	C,PCs/LVs	RBF	Poly	RBF	Poly	RMSE
PCA-SVM	21,5	1	---	1	0	9.597
KPLS	12	1	---	1	0	12.83
MKPLS	5	2.525	0	0.6359	0.3641	8.637

The result shows that the proposed method selects different kernel parameters and weighting coefficients for different kernel functions with the best prediction accuracy. However, only two kinds of kernels with linear combination method are used in this paper. Only one Benchmark data set is used to estimate the prediction performance of the proposed method. Therefore, more kernel types, more multi-kernel construction approaches and more Benchmark data sets would be addressed for the MKPLS algorithms in the future research.

Conclusions

This paper presents a novel multi-kernel partial squares modeling approach based on adaptive genetic algorithm. Popular poly kernel and radius basis function kernel are used to construct a multi-kernel function for kernel partial least squares algorithm. In order to solve the model learning parameters optimized selection problem, adaptive genetic algorithm is used. Namely, the linear combined multi-kernel's parameters and weighting coefficients are optimized select jointly. Data sets from UCI platform are used to validate the proposed modeling method. How to use more kernel types and more multi-kernel construction methods to improve the modeling performance would be research further. Moreover, how to reduce the training time and use more effective optimization algorithm would be addressed in the future study.

References

[1] M. G¨onen and E. Alpaydın. Journal of Machine Learning Research, Vol. 12 (2011), p. 2211.

[2] P. Pavlidis, J. Weston, J. Cai and W. N. Grundy, in Proceedings of the 5th Annual International Conference on Computational Biology, Montreal, Canada, (2001), p. 242.

[3] Y. Gu, C. Wang, D. You, Y. Zhang and S. Wang. IEEE Transactions on Geoscience and Remote Sensing, Vol. 50 (20127), p. 2852.

[4] P. Chen, D. Zhang, Z. Jiang and C.Wu. Journal of Computational Information Systems, Vol. 7 (2011), p. 4890.

[5] Z. Sun, C. Wang and Y. Huang. Journal of Computational Information Systems, Vol. 9 (2013), p. 3563.

[6] A. Rakotomamonjy, F. Bach, S. Canu and Y. Grandvalet, Journal of Machine Learning Research, Vol. 9 (2008), p. 2491.

[7] S. J. Qin, Computers & Chemical Engineering, Vol. 22 (1998), p. 503.

[8] R. Rosipal and L. J. Trejo. Journal of Machine Learning Research, Vol. 2 (2002), p. 97.

[9] Dong Yan, Shaowei Liu and Jian Tang. Adwance Material Research, Vol. 12 (2012), p. 2211.

[10] M. Srinivas and L. M. Patnaik. IEEE Transactions on SMC, Vol. 24 (1994), p. 656.

[11] Information on http://www.ics.uci.edu/~mlearn/ MLRepository.htm.

[12] J. Tang, W. Yu, T. Y. Chai and L. J. Zhao, Neurocomputing, Vol. 82 (2012), p. 167