Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (11): 2082-2086, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.2082.2086

Comparison of PCA and Model Optimization Algorithms for System Identification Using Limited Data

He Qinshu, Liu Xinen and Xiao Shifu
Institute of Systems Engineering, CAEP, Mianyang, 621999, China

Abstract: Support Vector Machine (SVM) is a novel modeling method that is valuable in regression and classification. Kernel parameters setting in the SVM training process, along with the feature selection, significantly affects system identification accuracy. The objective of this study is to obtain the better algorithms for better prediction accuracy. This study develops Principle Component Analysis (PCA) for feature selection and a grid searching and k-fold cross validation (GSCV) approach for parameter optimization in the SVM. Numerical and engineering results indicate that SVM based on PCA can be used for identification of nonlinear functions with related input variables, while the SVM based on GSCV is useful for complex system identification with limited number with kinds of uncertainties.

Key words: Support vector machine, principle component analysis, cross validation, system identification

INTRODUCTION

The system identification is a very difficult task which consists of determining the mathematical model of a system or predicting its behavior even if it is a nonlinear system. In the modeling, all available indicators can be used, but inputs data may be correlated and limited number with kinds of uncertainties.

At present, a new kind of regression method-SVM has been applied to regression for reliability analysis (Rocco and Moreno, 2002) and system identification and it has become a popular tool in solving the problems such as small samples, high dimensions, nonlinear and local minimum problem, due to its remarkable characteristics such as good generalization performance, the absence of local minima and the sparse representation of solution. Unlike most of the traditional methods which implement the Empirical Risk Minimization Principal, SVM implements the Structural Risk Minimization Principal which seeks to minimize an upper bound of the generalization error rather than minimize the training error (Vapnik, 2000). This eventually results in better generalization performance in SVM than other traditional methods.

In developing a SVM model, the first important step is feature selection (Lee *et al*., 2007). If the SVM is adopted without feature selection, then the dimension of the input space is large and non-clean, lowering the performance of the SVM. Thus, the SVM requires an efficient and robust feature selection method that discards noisy, irrelevant and redundant data, while still retaining the discriminating power of the data. Features extracted from the original data are adopted as inputs to the regression in the SVM.

However, the SVM embeds tuning parameters that control the training setting such as kernel parameters and the trade off variable C. These parameters have a regularization effect on the cost function minimized during the training process. If their values are not trained, these variables may diminish the overall performance of the regression if not well chosen. In fact, given a regression task, picking the best values for these variables is a nontrivial model selection problem that needs either an exhaustive search over the space of hyper-parameters or an optimization procedure that explores only a finite subset of the possible values (Chapelle *et al*., 2002). This procedure requires automatic grid search (Wang *et al*., 2005) over the space of parameter values and needs an algebraic criteria to estimate upper bound of the expected error (Ayat *et al*., 2005), especially when there are more than two hyper-parameters.

In this study, we considered PCA for feature selection and k-fold cross validation with grid searching for optimization of model parameter selection. Numerical examples show that SVM based on PCA can be used for identification of nonlinear functions with related input variables, while the SVM based on GSCV is useful for complex system identification with limited number with kinds of uncertainties.

Corresponding Author: He Qinshu, Institute of Systems Engineering, CAEP, Mianyang, 621999, China

PCA FOR FEATURE SELECTION

The regression accuracy rate of SVM is influenced by not only the kernel parameters, but also other factors including the quality of the feature's dataset. For instance, the correlation between features influences the regression result. Accidental elimination of important features might decrease the accuracy rate of prediction. Additionally, some features of the dataset may have no effect at all, or contain a high level of noise. Removal of such features can increase the search speed and the accuracy rate.

Feature selection methods can be categorized as filter models and wrapper models (Liu and Motoda 1998). Filter models utilize statistical techniques, such as Principal Component Analysis (PCA), factor analysis (FA) and so on. Principal Component Analysis (PCA) is a well-known method for feature extraction. By calculating the eigenvectors of the covariance matrix of the original inputs, PCA linearly transforms a high-dimensional input vector into a low-dimensional one whose components are uncorrelated (Cao *et al*., 2003).

Given a set of centered input vectors:

$$x_t(t = 1, \cdots 1, \sum_{t=1}^{1} x_t = 0)$$

each of which is of m dimension x_t $(x_t(1), x_t(2), \ldots, x_t(m))^T$ (usually m<l), PCA linearly transforms each vector x_t into a new one s_t by:

$$S_t = U^T x_t \quad (1)$$

where, U is the m×m orthogonal matrix whose ith column u_i is the ith eigenvector of the sample covariance matrix:

$$C = 1/1\sum_{t=1}^{1} x_t x_t^T$$

other words, PCA firstly solves the eigenvalue problem (2)

$$\lambda_i u_i = Cu_i,\ i = 1, \cdots, m \quad (2)$$

where, λ_i is one of the eigenvalues of C. u_i is the corresponding eigenvector. Based on the estimated u_i, the components of s_t are then calculated as the orthogonal transformations of x_t:

$$s_t(i) = u_i^T x_t,\ i = 1, \cdots, m \quad (3)$$

The new components are called principal components. By using only the first several eigenvectors sorted in descending order of the eigenvalues, the number of principal components in s_t can be reduced (chosen by the ratio of cumulative contribution, usually >85%). So PCA has the dimensional reduction characteristic.

SUPPORT VECTOR MACHINE FOR REGRESSION (SVR)

After feature extraction using PCA, the training data points can be expressed as (s_1, y_1), (s_2, y_2), …., (s_l, y_l), ($s_i \in R^n$, $n \le m$ is the transformed input vector, $y_i \in R$ is the target value). The so-called Lagrange multipliers α_i^* and β_i^*, i = 1 are obtained by maximizing the dual function of (4), which has the following form:

$$L_D(\alpha, \beta) = \sum_{i=1}^{1} y_i(\alpha_i + \beta_i) - \varepsilon\sum_{i=1}^{1}(\alpha_i + \beta_i) - \frac{1}{2}\sum_{i=1}^{1}\sum_{j=1}^{1}(\alpha_i + \beta_i)(\alpha_j + \beta_j)K(s_i, s_j) \quad (4)$$

with the following constraints:

$$\sum_{i=1}^{1}(\alpha_i - \beta_i) = 0, \quad 0 \le \alpha_i, \beta_i \le C, \quad i = 1, \cdots, 1 \quad (5)$$

given the training data (s_i, y_i), an inner product kernel K $(s_i, s_j) = \phi(s_i)\cdot\phi(s_j)$, an intensive zone ε and a regularization parameter C. The dual problem can be solved via standard quadratic programming. Once solutions α_i^* and β_i^* are calculated, the regression function is constructed in terms of these values as:

$$f(x) = \sum_{i=1}^{1}(\alpha_i^* - \beta_i^*)K(x_i, x) \quad (6)$$

In the regression problem, both ε and C play an important role for noise and smoothness of the function, respectively. However, obtaining the optimal parameters is still a challenge.

GSCV FOR MODEL PARAMETERS SELECTION

Model selection gained a particular interest when neural networks established effective classification models on many pattern recognition problems in the late 1980s. An incorrect choice of one of these elements, may lead to a high generalization error. Two major parameters in SVR, ε and C, have to be set appropriately.

Grid search (Hsu *et al*., 2003) is the popular way to determine the values for parameters ε and C. The k-fold cross validation strategy is one of the most widely used error estimation approaches. It enables us to compare the

predicted response values with the true values based on the information available in a given data set. The key applications of the k-fold cross validation strategy are model performance estimation, tuning learning model parameters and assessing various choices such as the search strategies.

In this approach, the N sample points are divided into k different data folds. In each data fold, N/k sample points are placed and they are stratified in order to properly explore and describe the design space. The process of stratification is very important at this stage of cross validation. Among the k data folds, k-1 data folds are used to train the SVM models while the remaining one data fold is used to validate. Subsequently the processes of training and validation are performed for k times with each time a different data fold is used for the validation. Subsequently, the cross validation error of the predictions which is called normalized mean square error (NMSE) can be calculated using the following expression:

$$\varepsilon_{cv} = \frac{1}{(N/k)\delta^2}\sum_{i=1}^{N/k}(\hat{y}_i - y_i)^2$$
$$\delta^2 = \frac{1}{(N/k)-1}\sum_{i=1}^{N/k}(\bar{y}_i - y_i)^2 \quad (7)$$

where, $\hat{y}_i$ represents the predictions, $\bar{y}$ denotes the mean of the actual value. The SVM models whose parameters chosen by grid search are tested by k-fold cross validation and the optimal parameters are which minimize the sum of ε_{cv} for each fold.

EXAMPLES

Numerical example: Let us consider the function as follows:

$$f(x, y) = xy+y \quad (8)$$

where the correlation coefficient of input variables is 0.7764. The 15 data points of x, y chosen by MCS and the corresponding value of f(x, y) are shown in Table 1. The data points at odd locations are used as the training set and the remaining data points are used as the test set. The comparison of predicted values for SVM, PCA-SVM, GSCV-SVM and PCA-GSCV-SVM algorithms are shown in Fig. 1 and Table 2. As seen from Fig. 1, the prediction effect is very bad for test set by traditional SVM algorithm because of the correlation between input variables and few samples, while the preprocessing of feature selection and the optimization algorithm can improve the regression effect a lot. As shown in Table 2, the minimum NMSE is produced by PCA-GSCV-SVM with 8.45×10^{-6} and SVM, PCA-SVM and GSCV-SVM algorithms provide NMSE of 0.744, 3.25×10^{-4} and 7.24×10^{-3}, respectively, which shows that both of SVM with feature selection by PCA and with model parameter optimization by GSCV can increase regression accuracy rate and SVM with feature selection together with parameter optimization is the most efficient algorithm in all of the algorithms presented here.

Engineering example: As shown in Fig. 2, the cantilever is connected to the base by three M8 bolts, there are some flexibility and uncertainty in the root of a cantilever

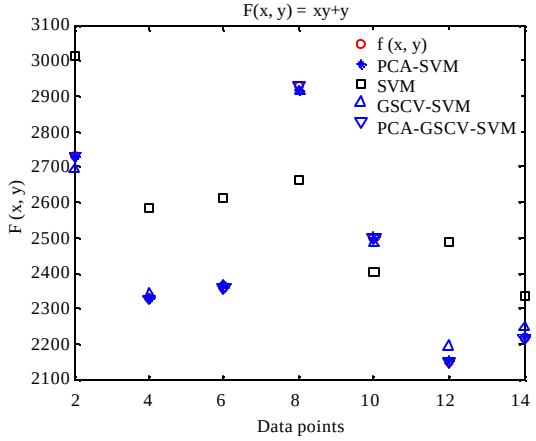

Fig. 1: Comparison of predicted values for different algorithms

Table 2: Model parameters with different algorithms

Parameters	SVM	PCA-SVM	GSCV-SVM	PCA-GSCV-SVM
λ	0.001	0.001	2×10^{-7}	2×10^{-9}
C	10	10	2×10^{-3}	20
NMSE	0.744	3.25×10^{-4}	7.24×10^{-3}	8.45×10^{-6}

Table 1: x,y and f(x,y)

Data	1	2	3	4	5	6	7	8
x	3.39	3.30	2.81	3.03	3.44	3.07	3.00	3.43
y	576	635	558	578	666	580	555	661
f(x, y)	2528.64	2730.5	2125.98	2329.34	2957.04	2360.6	2220.0	2928.23
Data	9	10	11	12	13	14	15	
x	3.36	3.13	3.12	2.74	2.76	2.88	2.96	
y	651	605	653	575	545	572	594	
f(x, y)	2838.36	2498.65	2690.36	2150.5	2049.2	2219.36	2352.24	

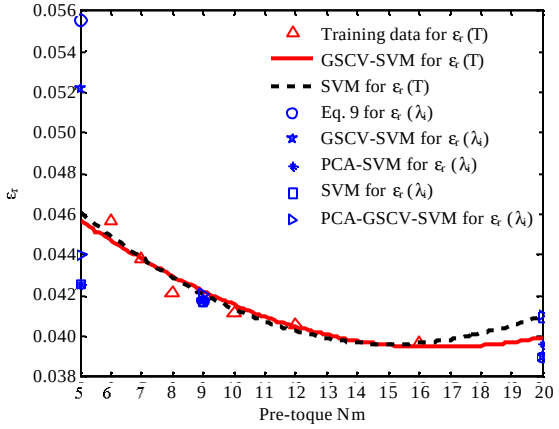

Fig. 2: A cantilever with bolt joints

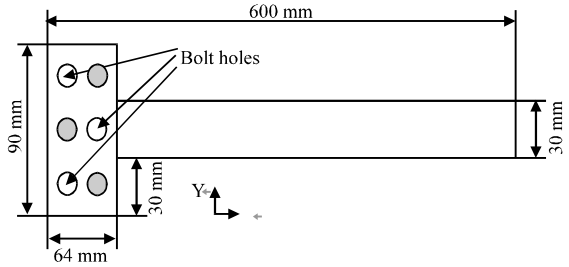

Fig. 3: Comparison of regression models for ε_γ (T) and predictions for ε_γ (λ_i)

with bolt joints. The rotation flexible parameter ε_γ is treated as an uncertain parameter, which is identified according to modal tests. Nine sets modal tests for nine pre-toques (5Nm, 6Nm, 7Nm, 8Nm, 9Nm, 10Nm, 12Nm, 16Nm, 20Nm) are carried out. The experiments are designed from 5Nm to 20Nm for five cycles and both bolts and gaskets are replaced for each cycle, in which experiments are repeated five times for each pre-toque. There are 25 experimental data for each order nature frequency with every pre-toque.

The relation of the first three order nature frequencies with the rotation flexible parameter is built as follows (Xiao and Chen, 2005):

$$\varepsilon_r = 1 - 0.4028\lambda_1 - 0.0908\lambda_2 + 0.0231\lambda_3 \quad (9)$$

Given the first three nature frequencies identified by modal tests, the rotation flexible parameter ε_γ is achieved with Eq. 9, which is the regression aim. The means of λ_i^j (for i = 5Nm, 9Nm, 20Nm, j = 1, 2, 3) and the corresponding ε_γ^i are used as the test set and the remaining six groups of λ_j^i with ε_γ^i are used as the training set. The comparison of the predicted values for SVM, PCA-SVM, GSCV-SVM and PCA-GSCV-SVM algorithms are shown in Table 3. As seen from it, the three predictions represent two extrapolations and one interpolation and the effects of interpolation for all of the algorithms are very well, while the effects of extrapolations are different a lot for different algorithms. For the largest pre-toque with 20Nm, all algorithms predictions are approximate with the calculated value as Eq. 9, while the effect of GSCV-SVM is the best. For the smallest pre-toque with 5Nm, none algorithms can predict approximately except GSCV-SVM. Furthermore, Six groups of T_i (for T_i = 6Nm, 7Nm, 8Nm, 10Nm, 12Nm, 16Nm) with the corresponding ε_γ^i are used as training set and the regression models ε_γ (T) built by GSCV-SVM and SVM respectively are shown in Fig. 3, in which predicted values for 5Nm, 9Nm and 20Nm are also drawn as scatter points with different marks.

Comparing with Eq. 8, the effects of feature selection before SVM model building are totally different, it is because that the relation of output with coherent input variables is nonlinear in Eq. 8, which is the most effective factor for regression, while the relation of output and input variables is linear in Eq. 9 and the uncertainties in experimental data is another important factor, which means the model is identified by input and output both with uncertainties.

CONCLUSIONS

In this study, PCA is applied to SVM for feature selection and GSCV is applied to SVM for model parameters optimization. Both of them improve the prediction ability of SVM for regression. The numerical and engineering examples show that feature selection as input data fore treatment more fits for nonlinear coherent multivariable system identification, while SVM model parameters optimization more fits for complex system with kinds of uncertainties using limited data and it has effect on the former situation too. Furthermore, although all of the algorithms presented here are good enough for interpolation, model built by GSCV-SVM has better extrapolation ability than other algorithms. Future work needs to explore ways of reducing the large computation cost in model parameters optimization because the high generalization performance of GSCV is obtained at the cost of a large amount of computation time.

ACKNOWLEDGMENT

The authors would like to thank the Project 2013A0203007 supported by CAEP.

REFERENCES

Ayat, N.E., M. Cheriet and C.Y. Suen, 2005. Automatic model selection for the optimization of SVM kernels. Pattern Recognit., 38: 1733-1745.

Cao, L.J. K.S. Chua, W.K. Chong, H.P. Lee and Q.M. Gu, 2003. A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine. Neurocomputing, 55: 321 -336.

Chapelle, O., V. Vapnik, O. Bousquet and S. Mukherjee, 2002. Choosing multiple parameters for support vector machines. Mach. Learn., 46: 131-159.

Hsu, C.W., C.C. Chang and C.J. Lin, 2003. A practical guide to support vector classification. Technical Report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei.

Lee, K. and V. Estivill-Castro, 2007. Feature extraction and gating techniques for ultrasonic shaft signal classification. Applied Soft Comput., 7: 156-165.

Liu, H. and H. Motoda, 1998. Feature Selection for Knowledge Discovery and Data Mining. Kluwer Academic, Boston, USA.

Rocco, C.M. and J.A. Moreno, 2002. Fast monte carlo reliability evaluation using support vector machine. Reliab. Eng. Syst. Saf., 76: 237-243.

Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. 2nd Edn., Springer, New York, USA., ISBN: 9780387987804, Pages: 314.

Wang, J., X.D. Wu and C.Q. Zhang, 2005. Support vector machines based on k-means clustering for real-time business intelligence systems. Int. J. Bus. Intell. Data Min., 1: 54-64.

Xiao, S.F. and B. Chen, 2005. Dynamic modeling of beams with flexible support. Mech. Eng., 27: 21-25.