Journal of Universal Computer Science, vol. 15, no. 13 (2009), 2726-2745
submitted: 31/10/08, accepted: 13/6/09, appeared: 1/7/09 © J.UCS

A New Short-term Power Load Forecasting Model Based on Chaotic Time Series and SVM

Dongxiao Niu
(North China Electric Power University, 102206, Beijing, China
niudx@126.com)

Yongli Wang
(North China Electric Power University, 102206, Beijing, China
wyl_2001_ren@163.com)

Chunming Duan
(North China Electric Power University, 102206, Beijing, China
dchm7093@yahoo.com.cn)

Mian Xing
(North China Electric Power University, 071003, Baoding, China
35507887@qq.com)

Abstract: This paper presents a model for power load forecasting using support vector machine and chaotic time series. The new model can make more accurate prediction. In the past few years, along with power system privatization and deregulation, accurate forecast of electricity load has received increasing attention. According to the chaotic and non-linear characters of power load data, the model of support vector machines (SVM) based on chaotic time series has been established. The time series matrix has also been established according to the theory of phase-space reconstruction. The Lyapunov exponents, one important component of chaotic time series, are used to determine time delay and embedding dimension, the decisive parameters for SVM. Then support vector machines algorithm is used to predict power load. In order to prove the rationality of chosen dimension, another two random dimensions are selected to compare with the calculated dimension. And to prove the effectiveness of the model, BP algorithm is used to compare with the results of SVM. Findings show that the model is effective and highly accurate in the forecasting of short-term power load. It means that the model combined with SVM and chaotic time series learning system have more advantage than other models.

Keywords: Support vector machine, Chaotic time series, Lyapunov exponents, Parameter selection, Load forecasting
Categories: F.2.1, H.1.1, I.1.2, I.1.6

1 Introduction

1.1 Electric Load Forecasting Approaches

As short-term power load prediction is of crucial importance to the reliability and economic utilization of electric networks, it is drawing more and more attention from both the practice and academia. Making the best use of electric energy and relieving

the conflict between supply and demand is the aim of load forecasting. Load forecasting has become a crucial issue for the operational planners and researchers of electric power systems. For electricity load reliance, electricity providers face increasing competition in the demand market and must pay more attention to electricity quality, including unit commitment, hydrothermal coordination, short-term maintenance, interchange and transaction evaluation, network power flow dispatched optimization and security strategies [Niu et al., 98].

Basic operating functions such as unit commitment, economic dispatch, security assessment, fuel scheduling and unit maintenance etc. can be performed efficiently with an accurate load forecasting. Because of the importance of load forecasting, a wide variety of models have been proposed in the last two decades, such as the exponential smoothing model, state estimation model, multiple linear regression model and stochastic time series model .

Generally speaking, these techniques are based on statistic methods and extrapolated past load behaviour while allowing for the effect of other influence factors such as the weather and day of week. However, the techniques employed for those models use a large number of complex and non-linear relationships between the load and these factors. A great amount of computational time is required and may result in numerical instabilities. Some deficiencies in the presence of an abrupt change in environment or weather variables are also believed to affect the load patterns. Therefore, some new forecasting models have been recently introduced such as artificial intelligence (AI), artificial neural network (ANN), and support vector machines (SVM).

Proposed by Vapnik [Vapnik, 95], support vector machines (SVMs) are one of the significant developments in overcoming shortcomings of ANNs mentioned above. Rather Than by implementing the empirical risk minimization (ERM) principle to minimize the training error, SVMs apply the structural risk minimization (SRM) principle to minimize an upper bound on the generalization error. SVMs could theoretically guarantee to achieve the global optimum, instead of local optimum like ANNs models. Thus, the solution of a nonlinear problem in the original lower dimensional input space could find its linear solution in the higher dimensional feature space.

SVMs have been widely applied in the field of pattern recognition, bio-informatics, and other artificial intelligence relevant areas. Particularly, along with Vapnik's ε-insensitive loss function, SVMs also have been extended to solve nonlinear regression estimation problems, which are so-called support vector regression (SVR). SVR has been successfully employed to solve forecasting problems in many fields, such as financial time series (stocks index and exchange rate) forecasting [Cao, 03; Huang et al, 05; Pai and Lin, 05], engineering and software field (production values and reliability) forecasting [Pai and Hong, 06], atmospheric science forecasting [Mohandes et al., 04] and so on. Meanwhile, SVR model has also been successfully applied to forecast electric load [Amari and Wu, 99; Pai and Lin, 05]. The empirical results indicate that the selection of the three parameters (C, ε and σ) in a SVR model influences the forecasting accuracy significantly. Although numerous publications have given recommendations on appropriate setting of SVR parameters [Cherkassky and Ma, 04], those approaches do not simultaneously consider the interaction effects among the three parameters. There is no general

consensus, thus, intelligent algorithms are employed to determine appropriate parameter values.

The last twenty years has witnessed the increasing maturity of chaotic theory and statistic learning theory. The application of chaotic theory to the load prediction has also come to mature. According to the characters of chaotic time series, based on the phase space delay coordinate reconstruction theory, combined with statistical theory the SVM prediction model based on Lyapunov exponents is established. Then the model is put into use in short-term power load forecasting of some electric grids to verify its effectiveness. The forecasting outcome from the model with the embedding dimension obtained through Lyapunov method, shows that the model is more accurate than BP neural network and SVM with random embedding dimension. This new model can be utilized by power companies for electric power load forecast, or by universities for scientific research.

Firstly, the paper explores the chaotic time methods. Secondly, it studies the technology of reconstruction phase space in chaotic time series. In the process, chaotic time series calculates Lyapunov exponents. Thirdly, it probes into the calculation of Lyapunov exponents such as reconstructing m-dimensional phase space, choosing minimal time delay, search a small neighborhood point, and finds out the embedding dimension. Fourthly, in the first part the time series matrix is established according to the theory of phase-space reconstruction, and then Lyapunov exponents is computed to determine time delay and embedding dimension. The support vector machines algorithm is used to predict power load, taking the above computed embedding dimension as one parameter.. It also has put forward some methods to determine other parameters. Lastly, it proves the accuracy of new method and compares the new model with other methods. The result shows that the new model is effective and highly accurate in the forecasting of short-term power load.

1.2 Chaotic Time Methods

Chaotic processes are characterized by irregular, unpredictable behaviours that nonetheless is deterministic. One century ago, Henri Poincare put forward chaotic properties while he investigated the dynamics of the three-body problem a century ago. In the early 1960s, during the course of his investigation of simple convection models, Edward Lorenz, the meteorologist [Lorenz, 63] noticed the sensitive dependence on initial conditions and realized its significance to the long-range weather forecasting.

Chaos refers to the phenomenon which comes forth in the certainty system with a seemingly without rules, similar to a random. Chaotic time series exist in many natural economic phenomena such as the price change in stock markets [Ma, 03; Su, 06] and some other economic Chaos [Katarzyna et al, 04]. The chaotic state with stochastic quality in nonlinear systems is predictable in the short term but unpredictable in the long term. At present, chaotic time series forecasting methods include: global prediction [Nv, 02], Local Forecast [James, 02], Lyapunov exponent forecast [Chen and Wang, 04] and neural network prediction method [Bunn, 00]. Based on the dynamic theory, the local forecast method is very suitable for chaotic time series, which is better than the global prediction. The result of neural network prediction is similar to that of the local forecast method [Douglas et al, 98]. The

commonly used methods it is depend on the largest Lyapunov exponent to identify whether the curve of history data is chaotic or not..

On the one hand, In the process by the method of computering the largest Lyapunov exponent. there are two reasons would lead to a stable flat curve. One reason is that by using the phase points of neighbourhood centre point and then using the least-squares method, the relationship between these phase points and the centre point can be determined. for the relationship is a linear. The more points there are in the neighbourhood, the more smooth the relationship between the centre point and other points will be. But the more separation space between the specific points, the weaker the impact of specific points to central point will be. The dynamic behaviours of centre points will be mostly impacted by nearest specific points Thus a stable flat curve can be obtained in this situation . Another reason is, when embedding dimension is determined, the number of elements which every point contains is decided too. When the relationship between each phase point and other phase points is examined, every element is of the same importance. It is to say that every element of each phase point produces different effects on the prediction. In fact each point's effects on prediction can generate some exponential decay with the passage of time. The result is most affected by the current time element . This equal treatment to elements in all phase points makes the forecast results to be smoothing changes in the trend. So if chaotic time series changes are relatively relaxed when the adding-weight one-rank local-region method is used to forecast, and the predict result is good enough, otherwise the prediction accuracy will be significantly affected. On the other hand, there are two reasons which would lead to predict results showing drastic change trend [Amari and Wu, 99]. In addition to these shortcomings, the above two methods are sensitive to changes in the embedding dimension. The determination of embedding dimension is current highly subjective, so there is a negative impact on forecasting accuracy.

In the last twenty years, chaotic theory and statistic learning theory have become been gradually mature and the usage of load prediction has become more and more mature. According to the characters of chaotic time series, the SVM prediction model based on Lyapunov exponents is established, which combines the phase space delaying coordinate reconstruction theory with SVM theory. Then it is used in short-term power load forecasting of some electric networks to verify its effectiveness. As a result, the model with the embedding dimension, which is got through Lyapunov method, shows this model is more accurate than BP neural network and SVM with random embedding dimension.

2 Chaotic Time Series and Lyapunov Exponents

According to the Chaos theory, the drive factors have influenced each other in chaotic system. Therefore the digital points occurring in time order are also related. At present, the phase space delay coordinate reconstruction method is commonly employed to analyze the factors of serial dynamics. Generally, the dimension is very great even infinite in the phase space in system, but in most cases the number of dimensions are unknown..In fact, the phase space delay coordinate reconstruction method can expand the given time series $x_1, x_2, \cdots, x_{n-1}, x_n, \cdots$ to three-dimensional

and even higher dimensional space so that information potentially stored in time series can be fully demonstrated, classified and extracted [Wang, 03; Sun et al, 04; Wen et al, 01; Wolf et al, 85; Li et al,, 03].

2.1 Reconstruction of Phase Space

The technology of reconstruction phase space is the premise to calculate Lyapunov exponents. In electric power system, actual loading series of single argument { $x(t_j)$ =1, 2, 3… n} can be obtained with the gap Δt. The structural characters of system attractors are contained in this time series. The specific method, which can estimate the information of phase space reconstruction in single argument time series, is

$$\begin{matrix} x(t_1) & x(t_2) & \cdots & x(t_j) & \cdots & x(t_n-(m-1)\tau) \\ x(t_1+\tau) & x(t_2+\tau) & \cdots & x(t_j+\tau) & \cdots & x(t_n-(m-2)\tau) \\ x(t_1+2\tau) & x(t_2+2\tau) & \cdots & x(t_j+2\tau) & \cdots & x(t_n-(m-3)\tau) \\ \cdots & \cdots & & \cdots & & \cdots \\ x(t_1+(m-1)) & x(t_2+(m-1)) & \cdots & x(t_j+(m-1)) & \cdots & x(t_n) \\ y(t_1) & y(t_2) & \cdots & y(t_j) & \cdots & y(t_n-(m-1)\tau) \end{matrix}.$$

In this method, the time series can be extended to m-dimensional phase space. Here, $\tau = k \Delta t$ (Δt =1, 2….) refers to the time delay is. In previous permutation, every column makes up a phase point of m -dimensional phase space. And each phase point has m components. These $n_p = n-(m-1)\tau$ phase points { $x(t_j)$, j =1, 2… n_p } make up a facades pattern in m-dimensional phase space, and the continuation of these phase points describes the evolutionary trace of a system in the phase space.

2.2 Calculation of Lyapunov Exponents

A. Wolf advanced a method by which a maximal Lyapunov exponents in single argument time series can be extracted. The process is:

1) Reconstructing m-dimensional phase space with time series.

2) Choosing minimal τ which marks the correlation among phase space.

3) In the continuation m-dimensional phase space, the initial phase point $A(t_1)$ is chosen as a reference point. There are m components in the phase space, they are $x(t_1), x(t_1+\tau), x(t_1+2\tau), \cdots x(t_1+(m-1))$.

According to the following formula:

$$L_{nbt} = \min\left[\left\|Y_i - Y_j\right\|\right] \qquad i \neq j, \tag{1}$$

$B(t_1)$,the nearest neighborhood point to $A(t_1)$ can be obtained. Assume L_{nbt}, referring to the distance between $A(t_1)$ and its nearest neighborhood point in Euclidean meaning, as $L(t_1)$. Suppose $t_2 = t_1 + k\Delta t$ with $k\Delta t$ as the step length, and $A(t_1)$ evolving into $A(t_2)$,meanwhile $B(t_1)$ evolving into $B(t_2)$, then the

distance $A(t_2)B(t_2) = l(t_2)$ is obtained. Let λ_1 represent the rate of exponential growth and $l(t_2) = L(t_1)2^{\lambda_2}$, then the following equation can be obtained.

$$\lambda_1 = \frac{1}{k(t_2 - t_1)} \log_2 (l(t_2) / L(t_1)) \quad (\Delta t = 1) . \tag{2}$$

λ is the Lyapunov exponent that can be used to judge the stability of time behavior of the system.

4) Find a small neighborhood point $C(t_2)$ which subjects to the angle θ_1 in the nearest neighborhood point to $A(t_2)$ (If it can't meet the two conditions: small θ_1 and neighborhood, it should still choose $B(t_1)$). Suppose $t_3 = t_2 + k\Delta t$, and suppose that $A(t_2)$ evolves into $A(t_3)$ and $C(t_2)$ evolves into $C(t_3)$, $A(t_2)C(t_2) = L(t_2)$ and $A(t_2)B(t_2) = l(t_2)$, then:

$$\lambda_2 = \frac{1}{k} \log_2 (l(t_3) / L(t_2)) . \tag{3}$$

The above process goes on until it reaches the end of point-group $\{X(t_j), j = 1,2,\cdots,n_p\}$. Then adopt the average of the calculated rates of exponential growth as the maximal estimated value of Lyapunov exponent. That is

$$LE_1(m) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{k} \log_2 \frac{l(t_i - 1)}{L(t_i - 1)} \quad . \tag{4}$$

$N = n_p / k$ means total steps of step length.

5) Increase embedding dimension m in turn and repeat the steps(3)-(4)until the estimated value $LE_1(m)$ of the exponent keeps stable and $LE_1(m_0) = LE_1(m_0 + 1) + LE_1(m_0 + 2) = \cdots = LE_1$. LE_1 is just the maximal Lyapunov exponent.

3 SVM Regression Theory

This section presents the fundamental knowledge of SVM regression. Suppose a set of data $\left(x_i, y_i\right)$, $i = 1,2\cdots n$, $x_i \in \boldsymbol{R}^n$ are given as inputs , and $y_i \in \boldsymbol{R}$ is the corresponding output. SVM regression theory is to find a nonlinear map from input space to output space and map the data to a higher dimensional feature space through the nonlinear map, and then the following estimated function is used to make linear regression [Tan, 01; Vapnik, 95; Du and Wu, 03].

$$f(x) = \left[\omega \cdot \phi(x)\right] + b \qquad \phi : \boldsymbol{R}^m \to \boldsymbol{F} \quad , \quad \omega \in \boldsymbol{F} \tag{5}$$

$f(x)$ is the regression estimated function which is constructed through learning of the sample set. ω is weight vector, b is the threshold value, $\phi(x)$ is the nonlinear mapping from input space to high-dimensional feature space which is the only hidden space. The problem of the function approximate is equivalent with the minimizing the following problem.

$$R = \frac{1}{2}\|\omega\|^2 + C\frac{1}{l}\sum_{i=1}^{l}|y_i - f(x_i)|_{\varepsilon} \tag{6}$$

$\|\omega\|^2$ is the weight vector norm, used to constrain the model structure capacity in order to obtain better generalization performance. C is the regularized constant determining the trade-off between the empirical error and the regularization term. In addition, in (6), Vapnik's linear loss function with ε-intensive zone is adopted as a measure for empirical error as shown in (7).

$$|y - f(x)|_{\varepsilon} = \begin{cases} 0 & if\ |y_i - f(x_i)| \le \varepsilon \\ |y_i - f(x_i)| - \varepsilon & otherwise \end{cases} \tag{7}$$

After we introduce positive slack variables ε_i and ε_i^*, minimizing the risk function R in (8) is equivalent to minimizing the objective function shown as follows:

$$R = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{n}(\xi_i^* + \xi_i) \tag{8}$$

Subject to:

$$\begin{cases} (\omega \cdot \phi(x)) + b - y_i \le \varepsilon + \xi_i^* & i = 1,2,\cdots,l \\ y_i - (\omega \cdot \phi(x)) - b \le \varepsilon + \xi_i & i = 1,2,\cdots,l \\ \xi_i, \xi_i^* \ge 0 & i = 1,2,\cdots,l \end{cases} \tag{9}$$

where ξ_i is the upper training error (ε_i^* is the lower) subject to the ε-intensive tube $y_i - (\omega \cdot \phi(x)) + b \le \varepsilon$. With the Lagrange multipliers introduced, the decision function in (9) can be expressed as the following explicit form:

$$f(x) = \sum_{i=1}^{l}(a_i - a_i^*)k(X_i, X) + b \tag{10}$$

where a_i and a_i^* are Lagrange multipliers with $a_i \times a_i^* = 0$ and $a_i, a_i^* \ge 0$ for any $i = 1,2,\cdots,l$. Using Mercer's theorem, the regression is obtained by solving a finite dimensional QP problem in the dual space avoiding explicit knowledge of the high dimensional mapping and using only the related kernel function. In (10), a kernel function $K(x_i, x_j) = \phi(x_i) \times \phi(x_j)$ is introduced, which is the inner product of two vectors in feature space $\phi(x_i)$ and $\phi(x_j)$. It can be shown that any symmetric kernel function K satisfying Mercer's condition corresponds to adopt product in some feature space. A common kernel is RBF kernel adopted in the paper as follows.

$$K(x, y) = \exp(-\|x - y\|^2 / (2\sigma^2)) \tag{11}$$

Thus, the Lagrange multipliers can be obtained by maximizing the following form:

$$R(a_i, a_i^*) = -\frac{1}{2}\sum_{i,j=1}^{n}\left(a_i - a_i^*\right)\left(a_j - a_j^*\right)K(x_i, x_j) \\ -\varepsilon\sum_{i=1}^{l}\left(a_i + a_i^*\right) + \sum_{i=1}^{l} y_i\left(a_i - a_i^*\right)] \tag{12}$$

Subject to:

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i^* \\ 0 \le a_i^* \le \boldsymbol{C}, i = 1, 2, \cdots, l \\ 0 \le a_i \le \boldsymbol{C}, i = 1, 2, \cdots, l \tag{13}$$

Through adjusting the two parameters C and ε, the generalized performance can be controlled in high-dimension space. According to Karush-Kuhn-Tucker(KKT) conditions, only some of coefficients $\left(a_i - a_i^*\right)$ are different from zero, and the corresponding training data are referred to as support vector, which can be regarded as the number of neurons in hidden layer of the network structure.

4 SVM based on Lyapunov Exponents

4.1 SVM Based on Time Series

If time series $\{x_1, x_2 \cdots x_N\}$ is given and the previous actual values of t time which are $x(1)$, $x(2)$ $\cdots$ $x(t)$ are known. Then the forecasting value of $t+1$ time point can be got through the following map:

$$f : \boldsymbol{R}^m \to \boldsymbol{R}. \tag{14}$$

The following equation can be obtained:

$$x(t \hat{+} 1) = f(x(t), x(t-1), \cdots x(t-(m-1))) \tag{15}$$

$x(t \hat{+} 1)$ is the predicted value of the $t+1$ time point and m is the embedding dimension. Then use SVM to make prediction. When the SVM kernel function and other parameters are determined, the model can be used to make prediction. Compared with BP neutral network, SVM has the following priorities: BP algorithm needs to select the number of interlayer while SVM just needs to search optimal parameters. BP network is based on empirical risk minimization and easy to create a local optimum. SVM is based on structural risk minimization, and it considers the sample error and the model complexity. Local optimum is global optimum in SVM. SVM exhibits better generalization ability than BP network.

4.2 Determination of Chaotic Time Series and Embedding Dimension

Lyapunov exponents is used to express average speed of neighborhoods in system. A positive Lyapunov exponent is to judge the segregation degree of average index number about two adjacent tracks while a negative Lyapunov exponent is to express the degree of closeness.. If a discrete nonlinear system is dissipative, a relatively stable and positive Lyapunov exponent can be computed to judge whether the time series is chaotic or not. The operation procedure is as follows:

Firstly, it finds Lyapunov exponents to judge whether the data is chaotic or not. Lyapunov exponents is a genealogy $L_i(i=1,2,3,\cdots,m)$. $L_1 > L_2 > \cdots > L_m$. L_i is the degree of evolution separation of the adjacent track with the time changing in m-dimensional space. If the max $L_1 > 0$ then it can be sure that the data series is chaotic.

Secondly, Judge whether Lyapunov exponents is smooth or not. The Lyapunov exponents will be varied with the different embedding dimensions. It finds the embedding dimension when the Lyapunov exponents has become smooth.

According to the characters of chaotic time series, the embedding dimension and Lyapunov exponents can be computed by Visual.Basic.6.0.with.SP6. The dimensional value is determined when Lyapunov exponents tend to be stationary. Then the model can be established by combining embedding dimension with time series theory.

The main source code is:

```
'Find out Lyapunov exponents
Private Function LyapunovLE(ByRef currM As Integer) As Double
  Dim m As Integer
  Dim tempLE(10) As Double
  Dim i As Integer
  Dim k As Integer
  k = 1
  m = 4
  For i = 1 To 20
      tempLE(i) = pingJunLE(m + i - 1)
  Next i
  Do
      If blnPingHua(tempLE(), k) = True Then
          LyapunovLE = tempLE(5)
          currM = m
          Exit Function
      Else
          m = m + k
          For i = 1 To 10 - k
              tempLE(i) = tempLE(i + k)
          Next i
          For i = 1 To k
              tempLE(10 - k + i) = pingJunLE(m + i + 9 - k)
          Next i
      End If
  Loop While (oldDateNum - m) > 11
  LyapunovLE = 0
End Function

'Judge Lyapunov exponents whether smooth or not
Private Function blnPingHua(tempLE() As Double, ByRef k As Integer) As Boolean
  Dim temp As Double
      Dim i As Integer

  temp = tempLE(1)
  For i = 2 To 10
```

```
            If Abs(temp - tempLE(i)) > mJingDu Then
                blnPingHua = False
                k = i - 1
                Exit Function
            End If
            'temp = tempLE(i)
        Next i
        blnPingHua = True
        k = 1
End Function

'Compute the average of    LE
Private Function pingJunLE(ByVal m As Integer) As Double
        Dim Ai As Integer
        Dim MinC As Integer
        Dim tempSumLE As Double
        tempSumLE = 0
        For Ai = 1 To oldDateNum - m
            DoEvents
            MinC = findMinC(m, Ai)
            tempSumLE = tempSumLE + sumLE(m, Ai, MinC)
        Next Ai

        pingJunLE = tempSumLE / (oldDateNum - m)
End Function
```

4.3 Proposed approach for parameter selection

Selection of parameter C. According to Mattera and Haykin (1999), standard parameterization of SVM solution given by Eq.(10) need to be considered, assuming that the ε-insensitive zone parameter has been chosen. Also suppose, without loss of generality, that the SVM kernel function is bounded in the input domain [Vladimir and Ma, 04; Deng and Tian, 04; Zhang et al, 04].

$$K(x_i, x) = \exp(-\frac{\|x - x_i\|^2}{2p^2}) \tag{16}$$

where p is the width parameter.

Under these assumptions, one can relate the value of C to the range on response values of the training data.

Specifically, referring to Eq.(10), it is noted that the regularization parameter C defines the range of values $0 \le a_i$, $a_i^* \le C$ assumed by dual variables used as linear coefficients in SVM solution (10). Hence, a "good" value for C can be chosen equal to the range of output values of training data. However, such a selection of C is quite sensitive to possible outliers (in the training data), so we propose instead the following prescription for regularization parameter:

$$C = \max(\left|\bar{y} + 3\sigma_y\right|, \left|\bar{y} - 3\sigma_y\right|) \tag{17}$$

where $\bar{y}$ and σ_y are the mean and the standard deviation of the y values of training data. Proposed selection of C given by Eq.(17) coincides with prescription suggested by Mattera and Haykin (1999) when the data has no outliers, but yields better C -values (in our experience) if the data contains outliers.

Selection of ε . It is well-known that the value of ε should be proportional to the input noise level, which is $\varepsilon \propto \sigma$. Assume that the standard deviation of noise σ is known or can be estimated from data. However, the choice of ε should also depend on the number of training samples. larger sample sizes should yield smaller ε -values. Precise nature of such a dependency can be derived using a combination of simple statistical arguments followed by empirical tuning/verification, as discussed next. First, try to relate the value of ε to an empirical distribution of "errors" $\delta_i = \hat{y} - y_i, (i = 1, \cdots, n)$ observed for a given training data set of size n . Consider the sample mean of these errors:

$$\hat{\delta} = \frac{1}{n}(\delta_1 + \delta_2 + \cdots + \delta_n) \tag{18}$$

Random variable $\hat{\delta}$ can be interpreted as empirical estimate of noise observed from available training data set of size n , Hence, the choice of ε should depend on the variance of $\hat{\delta}$, In order to estimate the variance of $\hat{\delta}$, recall that component errors δ_i in (18) all have zero mean and variances σ^2 . According to the Central Limit Theorem, the sample mean (18) is Gaussian with zero mean and variance σ^2 / n . Hence, it seems reasonable to set the value of ε proportional to the "width" of the distribution of $\hat{\delta}$:

$$\varepsilon \sim \frac{\sigma}{\sqrt{n}} \tag{19}$$

Based on a number of empirical comparisons, we found that Eq. (19) works well when the number of samples is small. However, large values of n prescription (19) yield ε -values that are too small (practically zero). Hence, we propose the following (empirical) dependency:

$$\varepsilon \sim \sigma \sqrt{\frac{\ln n}{n}} \tag{20}$$

We do not have specific theoretical justification for factor $\ln n$ in (20). other that this factor typically appears in analytical bounds used in VC theory [Vapnik, 2001]. Based on the empirical tuning, the following practical prescription for ε is found:

$$\varepsilon = 3\sigma \sqrt{\frac{\ln n}{n}} \tag{21}$$

This equation provides good performance for various data set sizes, noise levels and target functions for SVM regression.

4.4 SVM Prediction Model

The given time series $\{x_1, x_2 \cdots x_N\}$ is separated into two parts. The previous n_{tr} data are used as training sample to compute parameter values and the rest data as testing sample to prove the effectiveness of the model. The one-dimensional time series is converted into poly-dimensional matrix by reconstructing phase-space. Time delay is 1. The m-dimensional matrix is established as follows.

$$\boldsymbol{X} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ x_2 & x_3 & \cdots & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n_{tr}-m} & x_{n_{tr}-m+1} & \cdots & x_{n_{tr}-1} \end{bmatrix}, \quad \boldsymbol{Y} = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_{n_{tr}} \end{bmatrix}.$$

$\boldsymbol{X}$ is input matrix and $\boldsymbol{Y}$ is output matrix. $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy (15). When the basic structure of the time series is determined, SVM can be used to train and predict the samples. The regression equation is

$$y_t = \sum_{i=1}^{n_{tr}-m} (a_i - a_i^*) k(X_i, X_t) + b \;\; t = m+1, m+2 \cdots n_{t_r} . \tag{22}$$

The prediction model of the next time point is

$$y_{n_{tr}+1} = \sum_{i=1}^{n_{tr}-m} (a_i - a_i^*) k(X_i, X_{n_{tr}-m+1}) + b , \tag{23}$$

and $X_{n_{tr}-m-1} = \{x_{n_{tr}-m+1}, x_{n_{tr}-m+2}, \cdots x_{n_{tr}}\}$.

4.5 Steps of SVM Model Based on Lyapunov Exponents

The whole process including four steps is shown as the follows. It can be seen in Fig 1:

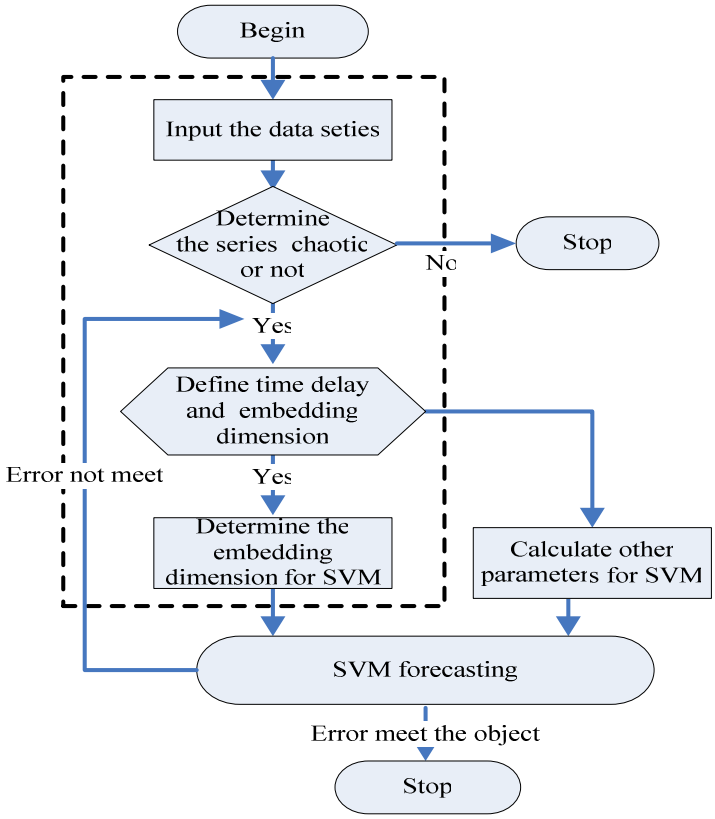

Figure 1: The SVM forecasting process based on chaotic time series

1) According to the historical power load data, a single-variable time series is established. Then wolf method is used to determine whether the series is chaotic or not by computing Lyapunov exponents.

2) Define time delay and work out the embedding dimension, then make data preprocessing. The embedding dimension and the number of phase points determined by Lyapunov method are used to restrict phase space.

3) Parameters in SVM model should be initialized. α_i α_i^* and b are assigned random values.

4) Proposed approach for parameter selection, the parameters C, ε should be calculated by the method proposed in the above, in which ε should be obtained by the equation (21).

5) The objective functions like (8)-(9) are established by using training sample. Then previous functions are converted into their dual problems like (12)- (13), thus α_i α_i^* and b will be worked out. And then they are substituted into (23), by this way the predicted value of the subsequent time points will be worked out.

5 Application and Analysis

5.1 Choosing Samples

The data used in this study were provided by the Shaanxi electric power corporation, a power transmission company in Shaanxi province. The data consist of quarter-by-quarter observations of electricity demand in Shaanxi province for 156 weeks ranging from 13 June 2003 to 15 June 2006. Power load data in Shaanxi province is used to prove the effectiveness of the model. The power load data from 0:00 at 6/13/2003 to 12:00 at 6/12/2006 are as training sample and used to establish the single-variable time series $\{x(t_1), x(t_2), \cdots x(t_{948})\}$. And the power load data from 13:00 to 24:00 at 6/15/2006 as testing sample.

Shaanxi Power Grid transmitted 1.836 billion KW to the north of China in 2006. Shaanxi province has become a important energy terminal of the economic development of north of China. As the Shaanxi Power Grid has to supply power for regional and several outside markets, Shaanxi Power Grid has paid more and more attention to the load forecasting to improve the security and stability of its electric network.

Selection of the Shaanxi province for the load forecasting is thus very representative in China due to the follwoing three reasons. Firstly, the Shaanxi province has distinct seasons, which is quite common all over China; Secondly, this area belongs to the land-locked region in the northwest and is seldom affected by extreme climate conditions. This is important to study the changing pattern of the load forecasting; Thirdly, the Shaanxi Power Grid has double tasks: satisfying regional power load demand and the demand of North China power grid, the main power grid in North China.

In order to investigate the performance of the forecasting system thoroughly, the Shaanxi power system with different typical load characteristics and weather conditions is considered in this research. Shaanxi province is a big power supply system in China, meeting the electrical needs of the North of China. The electrical demand of the Shaanxi province is mainly industrial and residential in a large area.

5.2 Chaos Analysis

For the training sample, $\tau = 1$ is chosen and Wolf method is used to compute Lyapunov exponents and embedding dimension. According to the theory, Lyapunov exponents λ begins to show stationary trend when the embedding dimension is 11. The power load time series shows chaotic characters because $\lambda > 0$. The embedding dimension is 11 and the number of phase points is 1136. The above parameters are used to reconstruct the phase-space. The results are shown in Fig. 2.

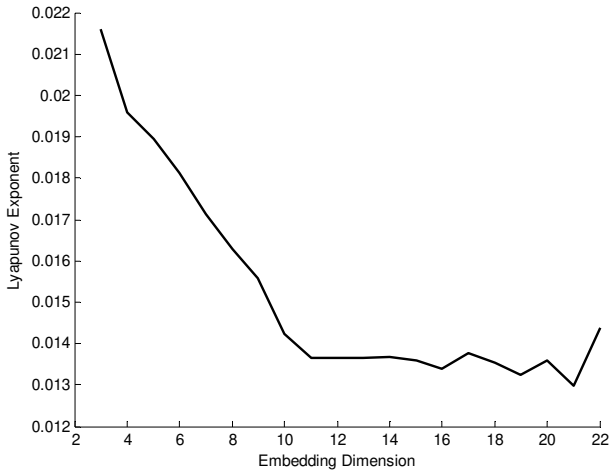

Figure 2: λ (Lyapunov exponent) changes with m (embedding dimension). When embedding dimension is 11, Lyapunov exponents begin to show stable tendency.

5.3 Prediction Process

SVM is used to make prediction after the samples are normalized. Libsvm toolbox is used to compute the results and radial basis function is chosen as the kernel function [9]. The parameters are chosen as follows: m=11,C=86.23, ε=0.016, σ^2=5.16. The other two comparing matrixes are established by choosing the parameters as follows: m = 9,C=78.86, ε=0.024, σ^2=3.58 and m = 13,C=49.68, ε=0.013, σ^2=2.11. The results of each matrix are shown in Table I.

BP algorithm is used to make prediction with sigmoid function. The parameters are chosen as follows: the node number of input layer is 11 and the node number of output layer is 1. The node number of interlayer is 8 according to the experience. The system error is 0.001 and the maximal interactive time is 5000. The results are shown in table 1.

5.4 Predicted Values and Evaluating Indicator

Relative error and root-mean-square relative error are used as the final evaluating indicators:

$$E_r = \frac{x_t - y_t}{x_t} \times 100\% \quad , \tag{24}$$

$$RMSRE = \sqrt{\frac{1}{n}\sum_{t=n}^{n}\left(\frac{x_t - y_t}{x_t}\right)^2} \quad . \tag{25}$$

The results show:

1) To see whether the approach of the chosen embedding dimension is reasonable or not, three cases are chosen as follows: 11-dimension, less than 9-dimension and larger than 13-dimension. Compaed with the performance of other three models, it can be seen that the relative errors of the proposed SVM(11) model in this paper are around 0, 9 out of 11 points. 75% of the forecasting points are in the scope of [-0.025, 0.025], and the root-mean-square relative error is only 2.03%,.The accuracy of load forecasting is satisfied. We used SVM model with different dimensions to check the result, there being 9-dimension and 13-dimension. They also can not calculate accurately which is wanted, and the RMSRE is 2.90% and 3.15% bigger than the RMSRE calculated with SVM(11).

2) In general, the expected scope of relative error is [-0.03, 0.03], so if the results are measured by the standard which is less than or equal to 3%, there is only one forecasting point exceeding 3% with SVM(11), 4 points with SVM(9), 5 points with SVM(9) and 8 points with BP model. The acceptable results in the condition of 11-dimension covers 91.67% while 9-dimension covers 66.67% and 13-dimension covers 58.33%, but the BP model only cover s33.33%. It can be seen from the above analysis that the predicting effectiveness of 11-dimension is better than other dimensions when SVM is used to make prediction.

Time point	**Actual load**	E_r **SVM(11)**	E_r **SVM(9)**	E_r **SVM(13)**	E_r **BP(11)**
13:00	413.5241	-1.56%	-3.31%	-2.58%	3.89%
14:00	438.4254	-2.31%	-2.98%	-3.87%	-5.63%
15:00	441.2467	1.28%	2.21%	2.65%	-6.31%
16:00	427.1403	-0.47%	-1.87%	2.11%	1.53%
17:00	480.7445	-2.73%	-2.33%	-2.24%	-3.08%
18:00	667.5127	-1.38%	2.26%	-3.61%	4.13%
19:00	834.5315	-1.57%	-3.24%	1.89%	1.69%
20:00	868.3869	0.93%	2.84%	2.66%	-2.57%
21:00	840.1742	2.85%	3.65%	-4.99%	5.87%
22:00	639.2995	3.28%	-2.55%	2.42%	-2.58%
23:00	455.3530	-1.61%	4.24%	3.71%	3.03%
24:00	368.5361	2.35%	2.47%	3.58%	-3.86%
RMSRE		2.03%	2.90%	3.15%	3.98%

Table 1: Comparison of The Predicted Values and Evaluating Indicators

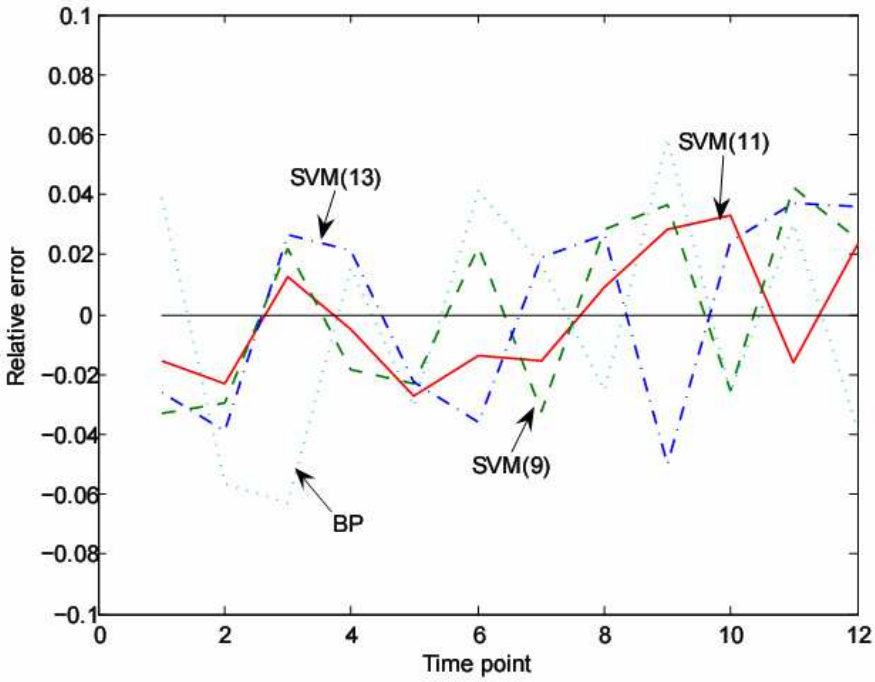

Figure 3: Error analysis of the proposed method

3) The comparison between SVM and BP is shown as follows. The relative error of predicted results by SVM has small rangeability. The maximal relative error is 3.28% and the value from the maximal relative error to the minimal relative error is 6.01%. On the contrary, the relative error of predicted results by BP has large rangeability. The maximal relative error is 6.31% and the maximal relative error to the minimal relative error is 12.18%. If the results are measured by the standard which is less than or equal to 3%, the acceptable results of SVM are 91.67% while BP are 33.33%. If the results are measured by root-mean-square relative error, RMSRE of SVM is less than BP. It can be seen from the above analysis that the predicting effectiveness of SVM is better than BP when the embedding dimension is determined.

6 Conclusions

The results show that SVM based on Lyapunov exponents has great effectiveness for short-term power load forecasting. And the conclusions are shown in the following:

1) The power load data show apparent chaotic characters. Chaotic time series is established and chaotic parameters are computed, then SVM prediction model is established to make prediction. The real load data prediction shows that the model is effective in short-term power load forecasting.

2) The embedding dimension is chosen through Lyapunov method. The predicted results with chosen dimension and other random dimension are compared.

The comparison shows that the approach is scientific and rational. The results show there is a suitable embedding dimension which is used to predict the power load effectively. The predicted values by the model with chosen dimension are highly accurate.

3) In the condition of the same dimension, SVM is much higher than BP in accuracy. The reason is that the differences between BP and SVM. BP network are based on empirical risk minimization and easy to lead to local optimum. SVM is based on structure risk minimization. Local optimum is global optimum. SVM exhibits better generalization ability than BP network.

Because the data of weather and temperature are hard to acquire while the data of power load are easy to acquire. In fact, the model of SVM based on Lyapunov exponents is more significant in the application than the models which need more power load data or the models which need the data of weather or temperature.

In near future we should do some research somewhere else. Firstly, the influential factors should be considered when making prediction, such as the weather, temperature, wind and other factors; Secondly, the parameter should be calculated with some intelligence methods so as to improve the accuracy of forecasting results.

Acknowledgements

Natural Science Foundation of China(70671039), The Ministry of Education to support the new century talents plan(NCET-07-0281). Beijing Municipal Commission of Education disciplinary construction and Graduate Education construction projects.

References

[Amari and Wu, 99] Amari S., Wu S. (1999). Improving support vector machine classiers by modifying kernel functions. *Neural Networks,* 12(6): 783-789.

[Amari and Wu, 99] Amari S., Wu S. (1999). Improving support vector machine classiers by modifying kernel functions. *Neural Networks,* 12(6): 783-789.

[Bunn, 00] Bunn DW. (2000). Forecasting loads and prices in competitive power markets. *Proc IEEE.,* 88(2): 163-169.

[Cao, 03] Cao L. (2003). Support vector machines experts for time series forecasting. *Neurocomputing,* 51: 321-339.

[Chen et al, 95] Chen JF, Wang WM, Huang CM. (1995). Analysis of an adaptive time-series auto regressive moving-average(ARMA) model for short-term load forecasting. *Electric Power System Research,* 34(3): 187-196.

[Chen and Wang, 04] Chen Suyan, Wang Wei. (2004). Chaos Forecasting for Traffic Flow Based on Lyapunov exponent. *China Civil Engineering Journal,* 37(9): 96-99.

[Cherkassky and Ma, 04] Cherkassky V., Ma Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks,* 17(1): 113-126.

[Cortes and Vapnik, 95] Cortes C, Vapnik. V. (1995). Support vector networks. *Mach Learn*, 20(3): 273-297.

[Deng and Tian, 04] Naiyan Deng and Yingjie Tian. (2004). *the New Approach in Data Mining- Support Vector Machines.* Science Press.

[Douglas et al, 98] Douglas AP, Breipohl AM, Lee FN, Adapa R. (1998). The impact of temperature forecast uncertainty on Bayesian load forecasting. *IEEE Trans Pow Syst.,* 13(4): 1507-1513.

[Douglas et al, 98] Douglas AP, Breipohl AM, Lee FN, Adapa R. (1998). Risk due to load forecast uncertainty in short term power system planning. *IEEE Transactions on Power Systems,* 13(4): 1493-1499.

[Du and Wu, 03] Shuxing Du and Tiejun Wu. (2003). Support vector machines for regression. *Journal of System Simulation,* 15(11): 1580-1586.

[Huang et al., 05] Huang W, Nakamori Y, Wang SY. (2005). Forecasting stock market movement direction with support vector machine. *Computers and Operations Research,* 32(10): 2513-2522.

[James, 02] James McNames. (2002). Local Averaging Optimization for Chaotic Time Series Prediction. *Neurocomputing,* 48(2): 279-297.

[Katarzyna et al, 04] Katarzyna Brzozowska-Rup, Arkadiusz Orowski. (2004). Application of Bootstrap to Detecting Chaos in Financial Time Series. *Physical A: Statistical Mechanics and its Applications,* 344(2): 317-321.

[Liang et al, 98] Zhishan Liang, Liming Wang, and Dapeng Fu. (1998). Short-term power load forecasting based on lyapunov exponents. *Proceeding of the CSEE,* 18: 368-472.

[Li et al, 03] Yuanchen Li, Tingjian Fang, and Erkeng Yu. (2003). Study of support vector machines for short-term power load forecasting. *Proceeding of the CSEE.,* 25: 55-59.

[Li et al,, 03] Guohui Li, Shiping Zhou, and Deming Xu. (2003). Computing the largest lyapunov exponents from time series. *Journal of Applied Sciences,* 21(2): 127-131.

[Li and Liu, 00] Tianyun Li and Zifa Liu. (2000). The chaotic property of power load and its forecasting. *Proceeding of the CSEE,* 20: 36-40.

[Lorenz, 63] E.N. Lorenz, (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.,*20: 130-141.

[Lv et al, 02] Jinghu Lv, Junan Lu and Shihua Chen. (2002). *Analysis and Application of Chaotic Time Series.* Wuhan University Press.

[Ma, 03] Ma Haijun. (2003). Application Study on Reconstruction of Chaotic Time Series and Prediction of Shanghai Stock Index. *Systems Engineering-theory & Practice,* 8(12): 86-94.

[Mbamalu et al, 93] Mbamalu GAN, El-Hawary ME. (1993). Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. *IEEE Trans Pow Syst.,* 8(1): 343-348.

[Mohandes et al., 04] Mohandes MA, Halawani TO., Rehman S., Hussain AA. (2004). Support vector machines for wind speed prediction. *Renew Energy,* 29(6): 939-947.

[Niu et al., 98] Dongxiao Niu, Shuhua Cao, and Yue Zhao. (1998). *Technology and Application of Power Load Forecasting*. China Power Press.

[Nv, 02] Nv Jinhu. (2002). Chaos time series analysis and its application. *Wu Han university press.*

[Pai and Lin, 05] Pai PF, Lin CS. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega,* 33(6): 497-505.

[Pai and Hong, 05] Pai PF, Hong WC. (2005). Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Conversion and Management,* 46(17): 2669-2688.

[Pai and Hong, 06] Pai PF, Hong WC. (2006). Software reliability forecasting by support vector machines with simulated annealing algorithms. *J Syst Software.,* 79(6): 747-755.

[Park et al, 91] Park JH, Park YM, Lee KY. (1991). Composite modelling for adaptive short-term load forecasting. *IEEE Trans Pow Syst.,* 6(2): 450-457.

[Su, 06] Su Chengjian. (2006). Study of Nonlinear Behaviour on Price and Volatility of Chinese Stock Markets. *Mathematics In Practice and Theory,* 36(2): 141-148.

[Sun et al, 04] Kehui Sun, Guoqiang Tan, and Liyuan Sheng. (2004). Design and implementation of lyapunov exponents calculating algorithm. *Computer Engineering and Application,* 35: 12-14.

[Tan, 01] Dongning Tan and Donghan Tan. (2001). Small-sample machine learning theory-statistical learning theory. *Journal of Nanjing University of Science and Technology,* 25(1): 108-112.

[Vapnik, 95] Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory.* Springer-Verlag Heidelberg New York.

[Vladimir and Ma, 04] Vladimir Cherkassky, Yunqian Ma. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Network,* 17: 113-126.

[Vladimir et al, 00] Vladimir N., Vapnik and Xuegong Zhang. (2000). *Nature of Statistics Theory.* Tsinghua University Press.

[Wang, 03] Xiaodan Wang and Jiqing Wang. (2003). A survey on support vector machine training and testing algorithms. *Computer Engineering and Application,* 13: 75-79.

[Wen et al, 01] Quan Wen, Yangchuan Zhang, and Shijie Chen. (2001). The analysis approach based on chaotic time series for load forecasting. *Power System Technology,* 25(10): 13-16.

[Wolf et al, 85] Wolf A, Swift J B, and Swinney H L. (1985). Determining lyapunov exponents from a time series. *Physics D.,* 16: 285-317.

[Yang and Cheng, 04] Yang, J.F., Cheng, H.Z. Application of SVM to Power System Short-term Load Forecast. *Electric Power Automation Equipment,* 24: 30-32.

[Zhan and Stan, 07] Justin Zhan, Stan Matwin. (2007). Privacy-preserving support vector machine classification. *International Journal of Intelligent Information and Database Systems,* 1(3/4): 356-385.

[Zhang and Han, 03] Zhang, H.R., Han, Z.Z. (2003). An Improved Sequential Minimal Optimization Learning Algorithm for Regression Support Vector Machine. *Journal of Software,* 14: 2006-2013.

[Zhang et al, 04] Lin Zhang, Xianshan Liu, and Hejun Yin. (2004). Application of support vector machines based on time sequence in power system load forecasting. *Power System Technology,* 28(19): 38-41.