Time Series Forecasting Model Based on SVM with Error Correction of Selected Parameter

Jiana Xu
Tianjin University of Technology, Tianjin 300384, China

Abstract
When application in time series forecasting, the standard SVM (support vector machine) algorithm has some defects, such as low accuracy in preferences and weak performance of anti-noise. In this paper, a time series forecasting model is proposed based on SVM with error correction of parameter selection. First, the historical data is for standardizing treatment to remove the smaller noise of signal within the prescribed scope. Then with space density clustering method reduces the influence of the noise data. And then the simulated annealing algorithm (SA) is introduced into the parameter selection of support vector machines, to modify the error of SVM algorithm. Simulation results show that the mean square error of the proposed SVM algorithm is smaller, greatly reduces the error of the parameter selection. And compared with standard SVM algorithm, forecasting Jingdong has significant improvement in the application of the short-term power load.

Key words: Time Series Forecasting, Error Correction, Parameter Selection, Spatial Density Clustering.

1. INTRODUCTION

Time series forecasting is the main research contents in the time series analysis, the basic idea is to use the current value and historical value of time series, establish the dynamic mathematical model which can accurately reflect the dependency contained in time sequence, and accurately find out the intrinsic characteristics of the system and forecast the future trend of sequence (Johannes et al., 2012). At present, the time series prediction has been widely used in space science, hydrological forecast, the weather forecast, the forecast of economic field, military science and industrial automation control and other fields (Bing et al., 2015; Thissen et al., 2013; Gunn et al., 2011).

Traditional time series forecasting method is a kind of forecasting method based on traditional statistics, the basic idea is to use the relationship between time series to establish the linear forecasting model, such as autoregressive model (AR), moving average (MA) model, autoregressive moving average (ARMA) model (Balaguer et al., 2008), etc. AR model and the ARMA model method are the most commonly used methods. The modeling mechanism of linear model method is simple, with fast speed of modeling and predicting. So it has been widely used in economic forecasting, agriculture, industrial control and other fields, such as: urban water consumption forecasting (Yang et al., 2012), the RMB exchange rate forecasting (Li, 2010), traffic flow forecasting (Huang et al., 2014), etc. Linear model based on traditional statistic is suitable for modeling of linear and stationary time series. In practice, however, acquired time series is generally nonlinear and non-stationary, the traditional linear model will no longer apply, established linear model and nonlinear system do not match, resulting in a decline in prediction accuracy, therefore the linear model needs be to improve. There are many scholars have studied how to improve the linear model to adapt to the nonlinear time series forecasting, such as differential treatment to remove the time series season trend and seasonal ARIMA model (Zhang et al., 2012), the autoregressive conditional heteroscedasticity (ARCH) model (Gao, 2007), the ARMA-ARCH model (Zhu, 2010), generalized autoregressive conditional heteroscedasticity (GARCH) model (Qiao, 2009), and local modeling methods of the combination of wavelet decomposition theory (Peng et al., 2011). The improved algorithm can solve the problem of nonlinear time series forecasting. But the strong pertinence of these methods may not use the different nonlinear time series. Poor generality of these methods are not generally applicable to the nonlinear time series prediction problems.

According to the defects of SVM in time series forecasting, a SVM time series forecasting model is proposed based on error correction of parameter selection, and simulations show the effectiveness of this model.

2. TIME SERIES FORECASTING MODEL BASED ON SVM

Because most of the time sequences are nonlinear, so to solve the nonlinear regression of samples set, the SVM is introducing kernel function \( K(x_i, x_j) = \phi(x_i) \phi(x_j) \) (\( \phi(x) \) for nonlinear function) to realize the mapping from low dimensional space of nonlinear samples set to high-dimensional feature space. Thus the nonlinear problem in low dimensional space into a high dimensional space linear problem, then the linear method is adopted to establish the SVM regression model. Kernel function is an important part of SVM. The SVM is
cleverly avoiding the time-consuming high-dimensional space of inner product operation with kernel function, and avoiding the dimension disaster, fundamentally solves the nonlinear problem.

The corresponding optimization problem of nonlinear regression is as follows.

$$\min \frac{1}{2} \|w\|^2$$  \hspace{1cm} (1)

The constraint is as the equation (2).

$$|w^T \phi(x_i) + b - y_i| \leq \varepsilon, i = 1, 2, ..., l$$  \hspace{1cm} (2)

In the equation (2), $\phi(x)$ is the nonlinear function. Converting into Lagrange dual problem is as follows.

$$\min_{\alpha, \alpha^*, \in \mathbb{R}} \frac{1}{2} \sum_{i,j} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j)$$
$$+ \varepsilon \sum_{i} (\alpha_i + \alpha_i^*) - \sum_{i} (\alpha_i - \alpha_i^*)$$  \hspace{1cm} (3)

The constraint is as the equation (4).

$$\sum_{i} (\alpha_i - \alpha_i^*) = 0$$  \hspace{1cm} (4)

After get the optimal solution $\alpha^* = (\alpha_1, \alpha_1^*, ..., \alpha_l, \alpha_l^*)^T$, can construct the regression function as follows.

$$f(x) = \sum_{i} (\alpha_i^* - \alpha_i)K(x_i, x) + b$$  \hspace{1cm} (5)

In order to verify the performance of the SVM algorithm in time series forecasting, the UCI data sets data Darwin is for simulating. Three parameters value of SVM are as follows: $C = 100$, $\varepsilon = 0.00054$, $\sigma = 0.0036$, the forecasting effect as shown in Figure 1.

![Figure 1](image_url)  \hspace{1cm} Figure 1. The forecasting results based on SVM algorithm

Seen from figure 1, time series forecasting model based on SVM algorithms also has some defects as follows: Standard SVM, in time series forecasting and applications in other fields, is same for error penalty parameter $C$ and error request parameter $\varepsilon$. That is, for different sample data, its accuracy, precision deviation punishment is non-discriminatory. However, in practice, it is often found that some date has high importance and small demand training error, while these dates have less importance and allow certain training error. For example, in reliability forecasting, the recent dates of the sample are often consistent with the environment of the dates to be predicted, thus it can provide more information.

3. PARAMETER SELECTED ERROR CORRECTION OF SVM ALGORITHM

3.1. Pseudo Data Filtering Based on Standardized Processing

The demand of nonlinear dynamics theory for the analysis of time series is high, because of various factors of interference, usually there are many false data in the original data. If the pseudo data is not for processing,
that will inevitably affect the precision and speed. Before forecasting, therefore, historical data need to deal, including the elimination of noise, filling of defect data and for the standardization of data appropriately.

If input data fluctuation is outside the range of specified threshold $D$, which produced a jump, then for the following processing.

1. If $|x(k) - \tilde{x}(k-1)| > D$, then made:

$$\tilde{x}(k) = \tilde{x}(k-1)$$  \hspace{1cm} (6)

In the equation (6), $x(k)$ is the real-time sampling data, $\tilde{x}(k-1)$ is for the data after pretreated in prior moment;

2. If $|x(k) - \tilde{x}(k-1)| < D$, the input data will adopt the method of moving average filtering algorithm for the second filter processing. The input data of a period of time are for $m$ arithmetic average operations:

$$\tilde{x}(k) = \frac{1}{m} \sum_{i=1}^{m} x(k-i+1)$$  \hspace{1cm} (7)

In the equation (7), $i = 1, 2, ..., m$, $m$ is the times of sampling, $\tilde{x}(k)$ is for preprocessing results, $x(k-i+1)$ is for the sampling data. The purpose is to filter the smaller noise of signal within the prescribed scope. In this way, the input data after the above processing becomes more smooth and real.

3.2. Denoising of Spatial Density Clustering

In the actual data, noise can affect the performance of the SVM forecasting model, so the method of spatial density clustering is put forward to reduce the influence of noise data.

Some data of high dimension and including noise points and the boundary point, the clustering effect becomes more obvious. For example, one dimensional characteristics of the data, as shown in figure 2, we can see a lot of discrete points, and usually we want to put these data points out.

![Figure 2. Boundary point or noise points of one-dimensional feature data](image)

According to the spatial noise density clustering method requires only one input parameter that can cover any shape of clustering. The shape of a neighbour is determined by the distance function of two points $p$ and $q$, represented as $d(p, q)$. Different distance functions can get different clustering.

Neighbourhood $\varepsilon$ of point $p$ is represented as $N_{\varepsilon}(p)$, defined as the equation (8).

$$N_{\varepsilon}(p) = \{q \in D | d(p, q) \leq \varepsilon\}$$  \hspace{1cm} (8)

Two class points are defined in the clustering algorithm. The internal clustering points are called nuclear data, and the clustering boundary points are called the boundary points. A minimum number within clustering points is expressed as $M$, point $p$ by point $q$ about $\varepsilon$ and $M$ called density can reach directly, if it meets equation (9),

$$\left[ q \in N_{\varepsilon}(p) \right] \left[ \left| N_{\varepsilon}(q) \right| \geq M \right]$$  \hspace{1cm} (9)
A point $p$ is the density reached by point $q$ about $\varepsilon$ and $M$, if there’s one point chain $p_1, \ldots, p_n$, $p_1 = q$ $p_n = p$ makes $p_{i+1}$ by $p$ density reach directly. Point $p$ and point $q$ about $\varepsilon$ and $M$ is density connection. If there is a point $o$ made $p$ and $q$ by $o$ about $\varepsilon$ and $M$, density can be reached.

Make $D$ denote a set of points, $C$ is a clustering, if meet the following conditions:
1. $C$ is the nonvoid subset of $D$;
2. For $\forall p, q$, if $p \in C$, and $q$ is $p$ about $\varepsilon$ and $M$ density reached and then $q \in C$;
3. For $\forall p, q \in C$, $p$ by $q$ about $\varepsilon$ and $M$ density connected.

![Figure 3. Density reached and density connected diagram](image3)

In Figure 3, $p$ by $q$ is density-reachable, and $q$ by $p$ is not density-reachable, $p$ and $q$ is density connected through $o$.

![Figure 4. Clustering results](image4)

In figure 4, through the space density clustering, the ring data with noise has been gathered to 2 classes, internal ring for a class, outer ring for a class, several of discrete noise point is marked out, do not belong to any class.
3.3. Parameter Selected Error Correction of the SVM Algorithm

After the above data preprocessing, simulated annealing algorithm (SA) is introduced to SVM parameter selection, to modify the error of SVM algorithm. The concrete steps are as follows:

(1) To set upper limit for the three parameters supporting vector machines and then give initial numbers and feedback to the SVM model. The absolute value of the prediction error is defined as system state $(E)$. Here it can reach the initial state $(E_0)$.

(2) Move randomly so that the system translates from initial state to critical state. There will be a collection of three parameters appear during this period.

(3) Decide whether accept or reject the critical state using the following formula (10):

\[
\begin{align*}
\text{accept, } E(S_{\text{new}}) > E(S_{\text{old}}) \& \ p < P(\text{accept } S_{\text{new}}) \\
\text{accept, } E(S_{\text{new}}) \leq E(S_{\text{old}}) \\
\text{refuse, other}
\end{align*}
\]

In the equation (10), $p$ is a random number deciding whether accept critical state or not. If accepting the critical state, set critical state as the current state.

(4) If not accept critical state, return to (2). If the current state is not better than system state, then recycle (2) and (3) until the current state is better than system state. In the end, set the current state as the new state. Previous study thought that it should set the maximum number system returned $N_{sa}$ should be $100d$ to avoid endless loop. $Na$ represents dimension of problems. When determine the SVM parameter, three parameters $(\sigma,C,\varepsilon)$ determine the system state, so set $N_{sa}$ as 300.

(5) After acquire the new parameter, reduce the temperature. The temperature reduced can be acquired from the formula:

\[
T_{\text{new}} = T_{\text{old}} \times \rho, 0 < \rho < 1
\]

In this paper, $\rho < 0.9$. If it reach the temperature determined previously, then the algorithm stops and the new state is approximate optimal solution. If not, return (2).

In the process of determining the optimal parameter of SA-SVM, we consider mean absolute percentage error (MAPE) as a criterion.

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{d_i - y_i}{d_i} \right| \times 100\%
\]

Where $n$ represents the number in forecast period; $d_i$ is the actual generated value in period I; $y_i$ is the predictive value of talent demand in period i.

Figure 5 is the process of parameters selected error correction of the SVM algorithm.
4. SIMULATION OF ALGORITHM

In order to verify the effectiveness of the proposed improved algorithm, we take simulation experiments on it. First, we simulate parameter selection of SVM algorithm and the mean square error \((MSE)\) is used as compared indicators of parameter optimization experiment, calculation formula is show in equation (13), and the result is shown in Figure 6.

\[
MSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left( \frac{y_i - \hat{y}_i}{\bar{y}_i} \right)^2} \times 100\% \tag{13}
\]

![Figure 6. Comparative results of preferences](image)

Then, we regard the history of a power system as input quantity, the improved SVM forecasting model is forecasting by electrical load. And the mean absolute percentage error as compared index to the precision, mean absolute percentage error calculation formula is shown as the equation (14).

\[
MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{|y_i|} \right| \times 100\% \tag{14}
\]

The results are shown in Figure 7.

![Figure 7. Comparison of forecasting error of power load](image)

Simulations show that the proposed SVM algorithm has smaller mean square error, greatly reduces the error of the parameter selection, and compared with the standard SVM algorithm, prediction in the application of the short-term power load forecasting Jingdong has significant improvement.
5. CONCLUSION

Time series forecasting method has been a hot topic in the research of scholars both at home and abroad. As prediction efficiency requirements advanced in many application fields, it is particularly important to research time series fast forecasting. The SVM has the advantages such simple model, high efficiency of training, strong learning ability, and it has a good application prospect for time series rapid prediction field. According to the defects of SVM in time series forecasting, a SVM time series forecasting model is proposed based on error correction of parameter selection. Simulations show that the proposed improved model has small error in the parameter selection, can accurately for time series with short-term forecasting.

REFERENCES