Forecasting Satellite Power System Parameter Interval Based on Relevance Vector Machine with Modified Particle Swarm Optimization

Xu Kang, Dechang Pi

Abstract—**Power system is an essential system in satellite, which ensures the security and stability of energy in the whole satellite system. This paper presents a mixed relevance vector machine with modified particle swarm optimization (MPSO-RVM) algorithm to forecast parameters intervals of satellite power system involved the main bus load current and the main bus voltage. First, RVM with radial basis kernel function is established to solve the regression problems of the data in satellite power system. Next, modified PSO algorithm is utilized to find out the optimal parameters of RVM to enhance the generalization capability. In addition, the self-adaptive parameter setting mechanisms is conceived to avoid the MPSO algorithm trapping into the local optima. Moreover, MPSO-RVM model can obtain desirable prediction intervals rather than prediction values. Experimental results demonstrate that MPSO-RVM model can achieve better prediction accuracy, sparser solution and shorter test-time than RVM model and PSO-SVR model. Meanwhile, the majority of samples are located into the prediction interval obtained at higher confidence level. Therefore, the proposed MPSO-RVM model vividly depicts the variation tendency of parameters in satellite power system, which is conducive to adopt available measures for avoiding satellite accidents and faults initiatively.**

Index Terms—**Satellite Power System, Prediction Interval, Relevance Vector Machine (RVM), Modified Particle Swarm Optimization (MPSO)**

I. Introduction

Power system is an important subsystem in multi-functional and complicated satellite system [1]. The satellite power system mainly includes solar array, storage battery, power cable, power regulation circuit and corresponding control system [2]. Fig. 1 shows the schematic structure of satellite power system with the process of supplying and discharging. As an important equipment in satellite power system, the main bus has two crucial parameters included the load current (denoted by IN1) and the load voltage (denoted by VN1). The main bus load current represents the current inflowing into the load through main bus filter capacitor. Similarly, the main bus voltage expresses the voltage among main bus. Up to now, there are many catastrophic accidents or temporary malfunctions in the satellite resulting from power system failures, see http://www.sat-nd.com/failures/ [3]. Therefore, it is one of the most significant subjects for researchers to prevent satellite accidents by detecting anomalous states of power system.

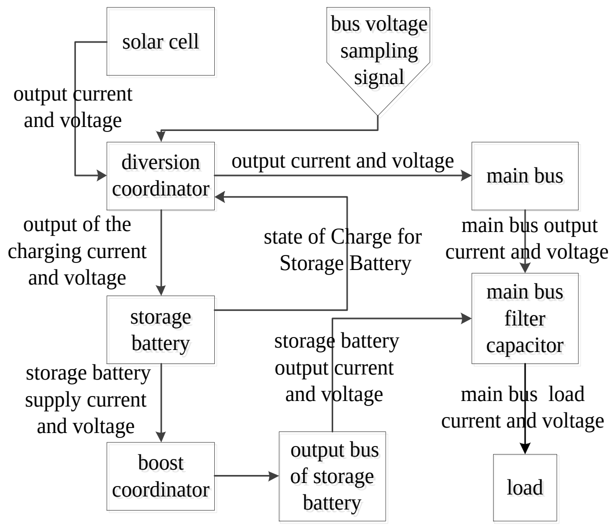

Fig. 1 The schematic structure of satellite power system

Recent years, many methods are proposed to predict anomalies of satellite power system [4]. Pan et al. [5] presented a data-driven method to monitor satellite power system anomalies using kernel principal component analysis (KPCA) and association rule mining. This method achieves better performance on distinguishing anomalies, but it is not suitable for on-line anomaly detection system. Wang et al. [6] proposed an approach to detect satellite power system's faults based on wavelet which reduces the effects of the noise data. Xie [7] et al. researched the fault detection of satellite power system using Bayesian Network, the results indicate the model deals with the uncertainties of fault diagnosis efficiently. Fang et al. [8] researched the health state evaluation on component-level and system-level satellite power system respectively and proposed a method based on SVM [9] to find health degradation and hidden danger of satellite power system. This method achieves the automation management of satellite power system and improves the system's accuracy. But SVM has problems of worse sparse capability, a deterministic output rather than a probability

Manuscript received March 8, 2016; revised August 9, 2016. This work was supported by National Natural Science Foundation of China (U1433116), Aviation Science Foundation of China (20145752033), Fundation of Graduate Innovation Center in NUAA (kfjj20161604).

X. Kang is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106 China. (E-mail: kx2014vip@163.com).

D. C. Pi is with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, 211106 China. (E-mail: dc.pi@nuaa.edu.cn).

distribution, and high time-consuming [10] [11]. After researching to solve this problems, a new thought called relevance vector machine (RVM) [12] is proposed. Wang and Liu [13] studied on forecasting the short-term load in electric power systems using RVM model with PSO algorithm. The results show that PSO-RVM model is sparser and can obtain higher forecasting accuracy compared with conventional models. Jin [14] presented a fault recognition method for automobile engine using improved PSO-RVM, and the method could obtain a sparser solution and the time spent on testing is smaller. In addition, the RVM relaxes the restrictions on the kernel function, and obtains the probability distribution of output [15]. Similarly, PSO-RVM model could also be applied to forecast satellite power parameter interval.

However, the difficulty in selecting proper RVM hyper-parameters slows down the velocity to resolve the practical problems significantly [16]. There are many efficient algorithms proposed in handling optimization problems in recent researches, such as particle swarm optimization (PSO) [17], genetic algorithms (GA) [18], and so on. Among these intelligent algorithms, PSO algorithm displays some significant features, like excellent performance and easy-to-implement virtue [19]. Therefore, PSO algorithm is an attractive option to optimize RVM parameter. To improve original PSO, a modified PSO (MPSO) is proposed by modifying accelerating factors, random variables and inertia weight.

In this paper, we propose a mixed MPSO-RVM model for forecasting the satellite power system parameters (IN1 and VN1 respectively) intervals. In the model, a modified PSO algorithm is utilized to find out optimal parameters of RVM to enhance the generalization capability. In order to validate the superiority of the model, we carry out research on the performance of MPSO-RVM compared with RVM and PSO-SVR. The remainder of the paper is organized as follows: The fundamental principle of RVM is introduced in Section II briefly and then section III elaborates the basic theory and algorithm of MPSO. A mixed MPSO-RVM model to forecast the satellite power system parameter interval is proposed in Section IV. Section V discusses the experimental analysis of MPSO-RVM model. Finally, this paper draws some conclusions in Section VI.

II. Relevance Vector Machine

Relevance Vector Machine (RVM) [12] proposed by Doctor Michael E Tipping is a new supervised learning algorithm based on Bayesian Theorem, Markov Property, automatic relevance determination (ARD), maximum likelihood and many kinds of theories [12][20][21]. RVM has similar structure of SVM, which can solve the nonlinear regression problems with small and high dimension samples, and also overcomes some problems of large number of free parameters, the difficulty in determining the parameters, and the kernel function abided by Mercer conditions [22]. In addition, the training model of RVM is sparser, and it can obtain the probability distribution of output. Because the test-time is shorter, thus it is more suitable for real-time prediction or online prediction [23] [24].

In the process of general regression prediction modeling, firstly give a training sample set including the input vector $\{x_n\}_{n=1}^{N}$ and corresponding target value $\{t_n\}_{n=1}^{N}$, where t_n is the actual value. According to RVM theory [12], the function of RVM obeys the following form:

$$y(x;w)=\sum_{i=1}^{N} w_i K(x,x_i)+w_0=w^T\phi(x) \tag{1}$$

where w represents weight vector whose value is $w=(w_0,w_1,\cdots,w_n)^T$, K denotes kernel function, and $\phi(x)$ is the linear combination of kernel functions defined as $\phi(x)=[1,K(x,x_1),K(x,x_2),\cdots,K(x,x_N)]^T$. Thereby, in order to estimate the specific form of model, we need to select appropriate kernel function and determine the value of each weight vector. The target value of each sample is independent and has Gaussian error, so the form is

$$t_i=y(x_i;w)+\varepsilon_i \tag{2}$$

In terms of equation (2), the target value of each sample satisfies the Gaussian distribution with $y(x_n;w)$ as average value and σ^2 as variance, so it can be regarded as

$$p(t_i\mid x_i)\sim N(t_i\mid y(x_i;w),\sigma^2) \tag{3}$$

In equation (3), the Gaussian distribution could be denoted as $p(t_i\mid w,\phi(x_i),\sigma^2)$. Thus, for each individual target value, the likelihood function of the whole sample set can be expressed as

$$\begin{aligned} p(t\mid w,\sigma^2)&=\prod_{i=1}^{N} N(t_i\mid y(x_i;w),\sigma^2)\\ &=(2\pi\sigma^2)^{-N/2}\exp\{-\frac{1}{2\sigma^2}\|t-\Phi w\|^2\} \end{aligned} \tag{4}$$

where $t=[t_1,t_2,\cdots,t_N]^T$,

$$\Phi=[\phi(x_1),\phi(x_2),\cdots,\phi(x_N)]^T =\begin{bmatrix} 1 & K(x_1,x_1) & K(x_1,x_2) & \cdots & K(x_1,x_N)\\ 1 & K(x_2,x_1) & K(x_2,x_2) & \cdots & K(x_2,x_N)\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & K(x_N,x_1) & K(x_N,x_2) & \cdots & K(x_N,x_N) \end{bmatrix}$$

In equation (4), adopting maximum-likelihood estimation to estimate w and σ^2 might lead to severe over-fitting problem. In order to avoid this problem, the exponential term $\|t-\Phi w\|^2$ in equation (4) can be briefly expressed as $\|t-y\|^2$ which can be regarded as the errors between the actual value and the measured ones. Let $\|\varepsilon\|^2$ express the errors, where $\varepsilon=[\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_N]^T$, so the number of the exponential term is the square of errors. Maximizing the likelihood function is equivalent to minimize the square of errors, and the maximum likelihood function is inclined to obtain the model which fits the training set best. However, overcomplicated model will lead to poor generalization capability, it doesn't have practical significance. To avoid this, RVM defines a Gaussian prior distribution with zero as average value of the weight vector w, as shown in equation (5),

$$p(w\mid\alpha)=\prod_{i=1}^{N} N(w_i\mid 0,\alpha_i^{-1}) \tag{5}$$

in which each α_i^{-1} monitors the velocity of the corresponding weight component w_i tending to zero. According to the hypothesis of RVM, the super prior distributions of α_i and σ^2 satisfy Gamma distribution which are shown in equation (6) and (7),

$$p(\alpha)=\prod_{i=0}^{N} Gamma(\alpha_i \mid a,b) \quad (6)$$

$$p(\sigma^2)=Gamma(\sigma^{-2} \mid c,d) \quad (7)$$

and where

$$Gamma(\alpha \mid a,b)=\Gamma(a)^{-1} b^a \alpha^{a-1} e^{-b\alpha} \quad (8)$$

with the "gamma function" $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$, and a=b=c=d=0. According to the priori theory of ARD, the vast majority of α_i approach to infinity and the corresponding α_i^{-1} concentrate at zero after enough iterating and updating. The corresponding weight components not equal to zero are relevance vectors, therefore the model becomes sparser.

In terms of Sparse Bayesian Theory, the posterior probability distribution can be expressed as follows:

$$p(w,\alpha,\sigma^2)=\frac{p(t \mid w,\alpha,\sigma^2)p(w,\alpha,\sigma^2)}{p(t)} \quad (9)$$

For a new input vector x_* , in terms of prediction distribution, the corresponding target t_* can be predicted as follows:

$$\begin{aligned} p(t_* \mid t) &= \int p(t_*,w,\alpha,\sigma^2 \mid t)dwd\alpha d\sigma^2 \\ &= \int p(t_* \mid w,\sigma^2)p(w,\alpha,\sigma^2 \mid t)dwd\alpha d\sigma^2 \end{aligned} \quad (10)$$

In equation (10), because we could not compute the integral $p(t)=\int p(t \mid w,\sigma^2)p(w,\alpha,\sigma^2)dwd\alpha d\sigma^2$, so we could not compute the posterior $p(w,\alpha,\sigma^2 \mid t)$ directly. Instead, we decompose the posterior $p(w,\alpha,\sigma^2 \mid t)$ as follows:

$$p(w,\alpha,\sigma^2 \mid t)=p(w \mid t,\alpha,\sigma^2)p(\alpha,\sigma^2 \mid t) \quad (11)$$

$$p(w \mid t,\alpha,\sigma^2)=\frac{p(t \mid w,\sigma^2)p(w \mid \alpha)}{p(t \mid \alpha,\sigma^2)} \quad (12)$$

Since $p(t \mid w,\sigma^2)$ and $p(w \mid \alpha)$ have been notified, then $p(t \mid \alpha,\sigma^2)$ could be computed by equation (13).

$$\begin{aligned} p(t \mid \alpha,\sigma^2) &= \int p(t \mid w,\sigma^2)p(w \mid \alpha)dw \\ &= (2\pi)^{-N/2} \mid \Omega \mid^{-1/2} \exp\{\frac{t^T\Omega^{-1}t}{2}\} \end{aligned} \quad (13)$$

where $\Omega=\sigma^2 I+\Phi A^{-1}\Phi^T$, $A=diag(a_0,a_1,\cdots,a_N)$.

Thus, the posterior probability distribution $p(w \mid t,\alpha,\sigma^2)$ of the weight ω is given via equation (14).

$$p(w \mid t,\alpha,\sigma^2)=(2\pi)^{-(N+1)/2} \mid \Sigma \mid^{-1/2} \exp\{-\frac{(w-u)^T\Sigma^{-1}(w-u)}{2}\} \quad (14)$$

where the posterior covariance and average are as follows respectively:

$$\Sigma=(\sigma^{-2}\Phi^T\Phi+A)^{-1} \quad (15)$$

$$u=\sigma^{-2}\Sigma\Phi^T t \quad (16)$$

where $A=diag(a_0,a_1,\cdots,a_N)$.

On the basis of delta function, the remainder $p(\alpha,\sigma^2 \mid t)$ can be accurately approximated by $p(\alpha,\sigma^2 \mid t)=\delta\left(\alpha_{MP},\sigma_{MP}^2\right)$ where α_{MP} is the most probably value of α , similarly σ_{MP}^2 is the most probably value of σ^2 .

Finally, $p(t_*,t)$ is approximated by equation (17).

$$\begin{cases} p(t_* \mid t) \approx \int p(t_* \mid w,\alpha_{MP},\sigma_{MP}^2)dw \\ (\alpha_{MP},\sigma_{MP}^2)=\arg\max_{\alpha,\sigma^2} p(\alpha,\sigma^2 \mid t) \end{cases} \quad (17)$$

According to equation (17), the aim of RVM model is converted into attaining the solution of α_{MP} and σ_{MP}^2 to maximize $p(\alpha,\sigma^2 \mid t)$. However, $p(\alpha,\sigma^2 \mid t)$ is proportional to $p(t \mid \alpha,\sigma^2)p(\alpha)p(\sigma^2)$, where $p(\alpha)$ and $p(\sigma^2)$ obey uniform distribution. Thus maximizing $p(\alpha,\sigma^2 \mid t)$ is equivalent to maximize $p(t \mid \alpha,\sigma^2)$.

In equation (13), the maximum values of α and σ^2 can't be obtained easily, thus we summarize the formula of their iterative re-estimation. Calculate the derivatives of equation (13), and make them equal to zero then rearrange them.

$$\alpha_i^{new}=\frac{\gamma_i}{\mu_i^2} \quad (18)$$

$$(\sigma^2)^{new}=\frac{\| t-\Phi\mu \|^2}{N-\sum_i \gamma_i} \quad (19)$$

where $\gamma_i=1-\alpha_i\Sigma_{ii}$, Σ_{ii} is the *i*-th diagonal element in Σ and N refers to the number of data examples.

Thus, the algorithm calculates Σ and μ iteratively, and updates the posterior probability distribution $p(w \mid t,\alpha,\sigma^2)$ until the convergence criteria has been satisfied.

During the hyper parameter estimation procedure, we should set up a proper iterative convergence condition in advance. The iterative convergence condition is set up as reaching the maximum number of iteration or determining whether α_i is convergent, that is whether $|\alpha_i^{new}-\alpha_i^{old}|<10^{-5}$. Then, we initialize α and σ^2 and make predictions based on the posterior distribution of the weight. Ultimately, we attain the optimal solution of α_{MP} and σ_{MP}^2 via the above iteration process.

For a new input sample x_i , we can calculate the prediction distribution from equation (10) and it can be given by applying equation (20).

$$p(t_* \mid t,\alpha_{MP},\sigma_{MP}^2)=\int p(t_* \mid w,\sigma_{MP}^2)p(w \mid t,\alpha_{MP},\sigma_{MP}^2)dw \quad (20)$$

In equation (20), because both terms in the integrand satisfy Gaussian distribution, so the result still obeys Gaussian distribution, giving:

$$p(t_* \mid t,\alpha_{MP},\sigma_{MP}^2) \sim N(t_* \mid y_*,\sigma_*^2) \quad (21)$$

where $y_*=\mu^T\phi(x_*)$ and $\sigma_*^2=\sigma_{MP}^2+\phi(x_*)^T\Sigma\phi(x_*)$.

In equation (1), kernel function $K(x,x_i)$ is defined as $K(x,x_i)=\phi(x)\phi(x_i)$. Introducing kernel functions could not only avoids the problem of "Curse of Dimensionality" effectively, but also solves the question of nonlinear regression efficiently. The three common kernel functions are shown in Table I.

Table I
Three common Kernel Functions

polynomial function	$K(x,x_i)=a((x\cdot x_i)+b)^d$
Radial basis function (Gauss kernel function)	$K(x,x_i)=\exp(-\| x-x_i \|^2/2\sigma^2)$
Cauchy function	$K(x,x_i)=1/\exp(1+\frac{\| x-x_i \|}{\sigma^2})$

In the process of general regression prediction modeling, RBF usually achieves more satisfactory performance than other mentioned functions, thus we choose RBF as the kernel function of RVM model.

III. Modified Particle Swarm Optimization

Particle Swarm Optimization (PSO) [25] algorithm is a stochastic optimization technique developed by Kennedy and Eberhart. It has attracted worldwide attention in various optimization problems and becomes the hotspot in field of evolutionary computation owing to its excellent performance and easy-to-implement virtue [26].

Suppose there is a swarm formed by *m* particles in an *S* dimensional search space. The *i*-th particle is initialized with a position vector $\vec{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iS})$ and a velocity vector $\vec{v} = (v_{i1}, v_{i2}, \cdots, v_{iS})$ where i is from 1 to *m*. The current optimal position of the *i*-th particle is denoted as $\overrightarrow{p_{iS}} = (p_{i1}, p_{i2}, \cdots, p_{is})$ and the global optimal position determined by the whole swarm is denoted by $\overrightarrow{p_{gS}} = (p_{g1}, p_{g2}, \cdots, p_{gS})$. During the searching procedure, each particle updates its position vector and velocity vector according to the following equations [25] [27] [28]:

$$v_{is}(t+1) = wv_{is}(t) + c_1 r_1 (p_{is}(t) - x_{is}(t)) + c_2 r_2 (P_{gs}(t) - x_{is}(t)) \quad (22)$$

$$x_{is}(t+1) = x_{is}(t) + v_{is}(t+1) \quad (23)$$

where i is from 1 to m, s is from 1 to S, t represents the current iteration, accelerating factors c_1 and c_2 are nonnegative constants, r_1 and r_2 are mutually independent random number, $w(t)$ stands for inertia weight, $x_{is}(t)$ ranged $[-x_{\max}, x_{\max}]$, $v_{is}(t)$ ranged $[-v_{\max}, v_{\max}]$, $p_{is}(t)$ and $p_{gs}(t)$ are the position, velocity, current optimal position and global optimal position of particle i on dimension s at iteration t respectively.

In order to accelerate convergent performance and enhance global optimization capability, two accelerating factors c_1 and c_2 adopt continuously decreasing function and continuously increasing function respectively in modified PSO algorithm. Thus the two factors are illustrated as follows:

$$c_1(t) = (c_{1,start} - c_{1,end}) \cdot (t_{max} - t) / t_{max} + c_{1,end} \quad (24)$$

$$c_2(t) = (c_{2,start} - c_{2,end}) \cdot (t_{max} - t) / t_{max} + c_{2,end} \quad (25)$$

where *t* represents the current iteration and t_{max} represents the maximum number of iteration, $c_{1,start}$ and $c_{2,start}$ stand for the initial values of c_1 and c_2, $c_{1,end}$ and $c_{2,end}$ stand for the final values of c_1 and c_2.

In modified PSO algorithm, random variables obey Gaussian distribution in equation (26) rather than uniform distribution on [0, 1]. The distributions of r_1 and r_2 are as follows:

$$r_1 \sim N(0, \sigma_1^2),\ r_2 \sim N(0, \sigma_2^2) \quad (26)$$

$$\sigma_1 = \frac{Fitness(x_i)}{Fitness(p_{iS})} \quad (27)$$

$$\sigma_2 = \frac{Fitness(x_i)}{Fitness(p_{gS})} \quad (28)$$

where $Fitness(X)$ represents the fitness of vector *X*.

In order to obtain a reliable and stable model, we select *k*-fold cross-validation to evaluate the fitness of each particle. The average of errors in *k*-subset is approximated as the fitness of each particle. It is estimated by the equation (29):

$$fitness = \frac{1}{k} \sum_{i=1}^{k} \sqrt{\frac{\sum_{j=1}^{l_k} \left(y_j - \hat{y}_j \right)^2}{l_k}} \quad (29)$$

where l_k represents the length of the subset s_k, $\hat{y}$ is the prediction value, *y* is the actual value. Eventually, store the optimal position and fitness of each particle.

The Gaussian random variables dominate the increment of velocity vector [29]. Thus, MPSO maintains the diversity of the swarm by escaping from local optima and attains the global optima with a high possibility.

The inertia weight *w* controls the influence of the previous velocity on the current velocity [30]. Therefore, the adjustment of *w* adopts the linearly decreasing weight strategy ranging from 0.9 to 0.4, that is expressed as follows:

$$w(t) = (w_{start} - w_{end}) \cdot (t_{max} - t) / t_{max} + w_{end} \quad (30)$$

where *t* represents the current iteration and t_{max} represents the maximum number of iteration. w_{start} stands for the initial weight and w_{end} denotes the final weight.

In the procedure of iteration, termination condition is reaching the maximum number of iteration or satisfying the predetermined minimum adaptive threshold.

The evolution procedure of searching global optimal solution with modified PSO is elaborated in Algorithm 1.

Algorithm 1: modified Particle Swarm Optimization

Input: Dimension of searching space S, number of particles in the swarm *m*, maximum iteration $t_{\max}$, initial accelerating factors $c_{1,start}$ and $c_{2,start}$, final accelerating factors $c_{1,end}$ and $c_{2,end}$, random numbers r_1 and r_2, initial weight w_{start}, final weight w_{end}.

Output: Global optimal position $\overrightarrow{p_{gS}}$.

// Initialize the parameters of MPSO

1. Initialize the original iteration: $t = 0$, accelerating factors: $c_{1,start} = 2.5$, $c_{1,end} = 0.5$, $c_{2,start} = 0.5$ and $c_{2,end} = 2.5$, initial weight: $w_{start} = 0.9$, final weight: $w_{end} = 0.4$
2. Set up initial accelerating factors, random numbers and inertia weight: $c_1(0) = c_{1,start}$, $c_2(0) = c_{2,start}$. $r_1(0) = rand(0,1)$, $r_2(0) = rand(0,1)$, $w(0) = w_{start}$

// Initialize the particle's position, velocity, current optimal position and global optimal position

3. For each particle $i = 1, 2, \cdots m$ do:
4. Initialize the particle's position with a uniformly distributed random vector : x_{iS}
5. Initialize the particle's current optimal position: $p_{iS} \leftarrow x_{iS}$
6. If ($Fitness(p_{iS}) < Fitness(p_{gS})$)
7. Update the swarm's global optimal position:

$$p_{gS} \leftarrow p_{iS}$$

8. End if
9. Initialize the particle's velocity with a uniformly distributed random vector: v_{iS}
10. End for
// Search the global optimal solution in searching-space
11. Until termination condition is satisfied, loop:
12. For each particle $i = 1, 2, \cdots m$ do:
13. For each dimension $s = 1, 2, \cdots S$ do:
14. Update the particle's velocity by equation (22).
15. End for
16. Update the particle's position by equation (23).
17. If ($Fitness(x_{iS}) < Fitness(p_{iS})$)
18. Update the particle's current optimal position:
$$p_{iS} \leftarrow x_{iS}$$
19. If ($Fitness(p_{iS}) < Fitness(p_{gS})$)
20. Update the swarm's global optimal position:
$$p_{gS} \leftarrow p_{iS}$$
21. End if
22. End if
23. End for
24. $t \leftarrow t + 1$
25. Update the accelerating factors, random variables and inertia weight by equation (24)-(28), and (30).
26. End loop
27. Output p_{gS} holds the global optimal solution

Notes:
(1) The position vector of *i*-th particle $x_i \sim U(-x_{\max}, x_{\max})$ where $x_{\max}$ is the maximum position; the velocity vector of *i*-th particle $v_i \sim U(-v_{\max}, v_{\max})$ where $v_{\max}$ is the maximum velocity.
(2) The termination condition is achieving the maximum number of iteration, or finding a solution corresponding with the predetermined minimum adaptive threshold.

IV. MPSO-RVM Model in Forecasting Satellite Power System Parameter Interval

Procedure of establishing the prediction model includes four steps: data preprocessing, parameter optimization of RVM with modified PSO, model establishing and prediction. The specific process of MPSO-RVM model is shown in Fig. 2. The circumstantial procedure of optimizing parameter of RVM with modified PSO and establishing MPSO-RVM model are summarized as an important part in Fig. 2.

A. Data Preprocessing

The experimental dataset in this work is some telemetry data comes from the power system of an anonymous satellite in space. First, we carry out experiments on satellite power system parameter interval prediction with telemetry dataset between June 12, 2011 and June 13, 2011, approximately 50 thousand records. Then we select the important parameters the main bus load current (denoted by IN1) and the main bus voltage (denoted by VN1) in satellite power system as the predictive objects in accordance with the suggestions proposed by space experts. Next, we preprocess the satellite telemetry data with the following steps.

Step 1: data cleaning

Based on expert knowledge, outliers are the data which are beyond 30 times range of the normal data while appearing less than three times per minute. Find out them and then delete.

Step 2: date conversion

Transform the initial data which are in equivalent time interval into standard deviation sequences.

Step 3: data transformation

Transform each parameter sequence x_t into data pattern $T = \{(X_1, Y_1), \ldots, (X_i, Y_i), \ldots, (X_{n-m+1}, Y_{n-m+1})\} \in (X \times Y)^{n-m+1}$

where *m* denotes the embedding dimension.

Step 4: data normalization

Normalize the experimental data ranged in [0, 1]. Then divide the processed data into two non-overlapping and independent parts with the ratio 90% and 10%, the former as training data is applied for RVM parameter optimization and model establishment and the latter as testing data is used to evaluate the model prediction efficiency and robustness.

B. Performance Criterion

At present, the effective method to judge the performance of the prediction model is mainly based on the accuracy of the prediction model. The following three common methods are used in this paper.

(1) Mean Absolute Percentage Error (MAPE)

Mean absolute percentage error reflects the overall credibility of the measured data. The computational formula is as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{31}$$

(2) Root Mean Square Error (RMSE)

Root mean square error is relatively sensitive to the measured data in large or small error, which reflects the measurement precision. The computational formula is as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{32}$$

(3) Normalized Mean Square Error (NMSE)

Normalized mean square error focuses on the relationship between the deviation among the prediction value with the actual value and the fluctuation intensity of the measured data. The computational formula is as follows:

$$NMSE = \frac{1}{\delta^2 n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{33}$$

where $\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y}_i)^2$, $\hat{y}$ represents the prediction value, *y* stands for the actual value, $\overline{y}$ indicates the average of the actual value, *n* is the number of samples.

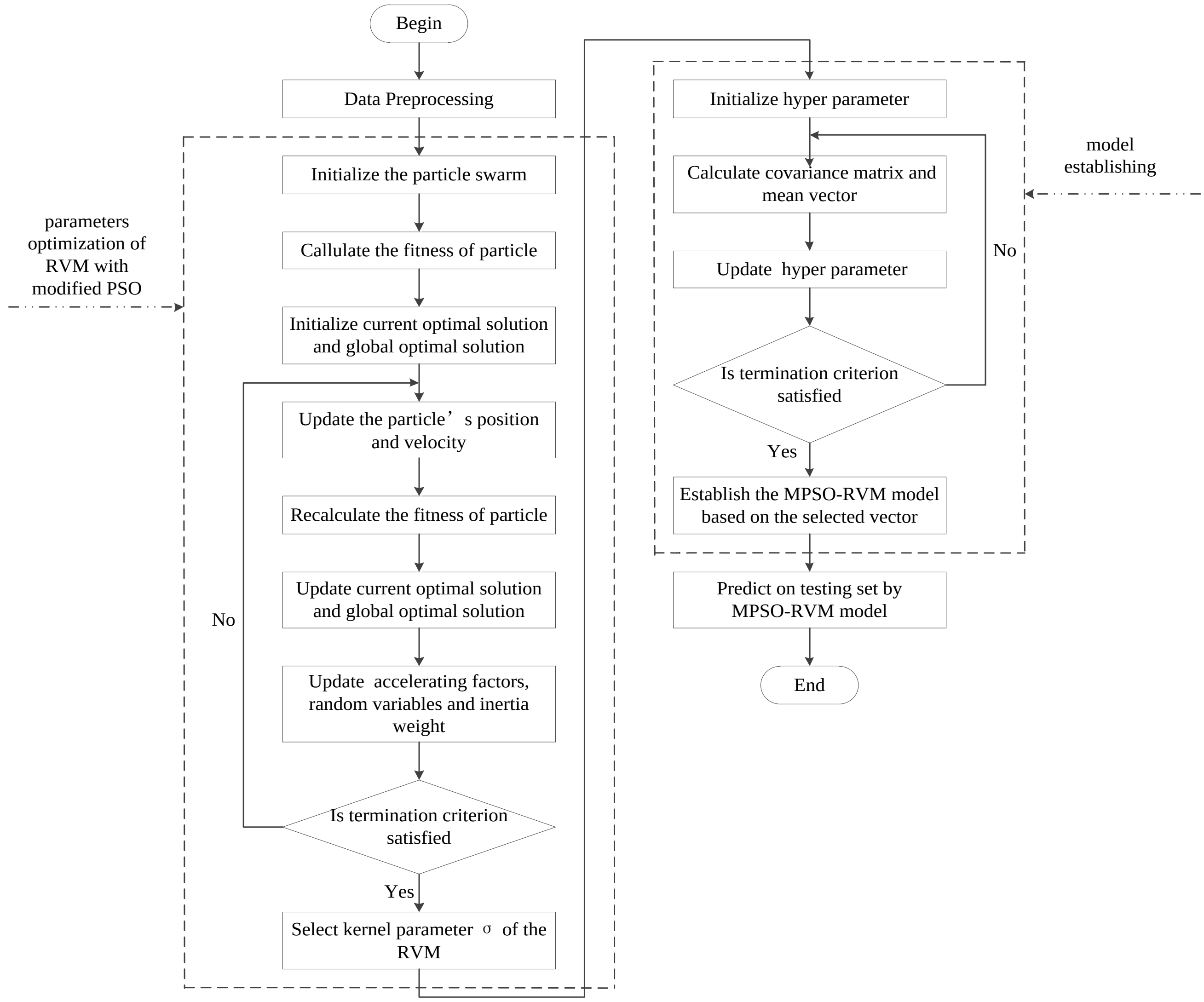

Fig. 2 The flow chart of MPSO-RVM.

V. EXPERIMENTAL ANALYSIS

In this work, we propose a mixed MPSO-RVM model to forecast satellite power system parameter interval. We develop a MPSO-RVM model based on Sparse Bayesian toolbox (http://www.miketipping.com/downloads.htm), to achieve the purpose of monitoring the operation state of satellite power system.

A. Setting of Experimental Parameters

(1) Embedding Dimension

In the procedure of data preprocessing, the data transformation stage needs to determine the value of the embedding dimension in advance. Therefore, we carry out 10 different experiments on crucial parameters IN1 and VN1 in satellite power system with the value of embedding dimension ranged from 1 to 10 and the parameter of kernel function set to 2. The influence of the embedding dimension *m* and the forecasting performance is shown in Fig. 3. Then the optimal embedding dimension is the one which is minimizing the MAPE on the testing set. Thus the optimal dimensions of crucial parameters IN1 and VN1 in satellite power system are all 4 respectively.

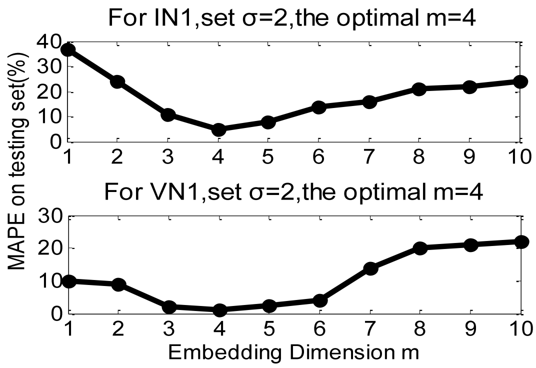

Fig. 3 Influence of embedded dimension *m* and forecasting performance

(2) Parameter of MPSO algorithm

In the process of parameter optimization of RVM with modified PSO, we set the dimension of searching space $S=19$, the number of particles *m=30*, maximum iteration $t_{\max}=100$, accelerating factors $c_{1,start}=2.5$, $c_{1,end}=0.5$, $c_{2,start}=0.5$ and $c_{2,end}=2.5$. Random numbers r_1 and r_2 obey Gaussian distribution on equation (26). The adjustment of *w* adopts the linearly decreasing weight strategy ranging from 0.9 to 0.4 via analyzing the time spent in optimizing the search procedure and the accuracy of the solution comprehensively.

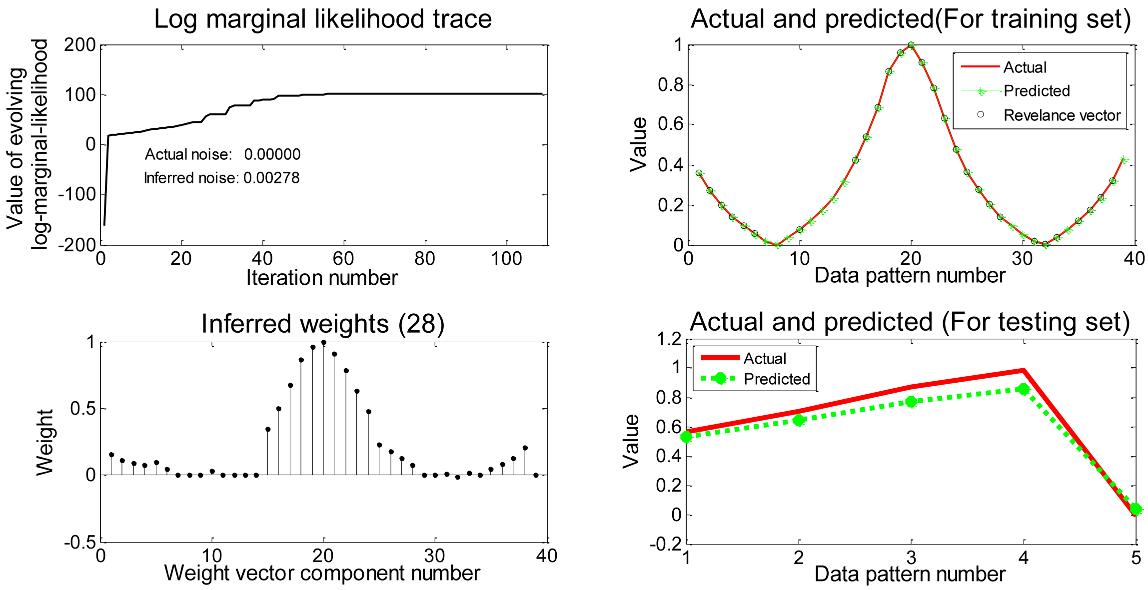

Fig. 4 Inferred weights and forecasting performance when $\sigma = 0.1$

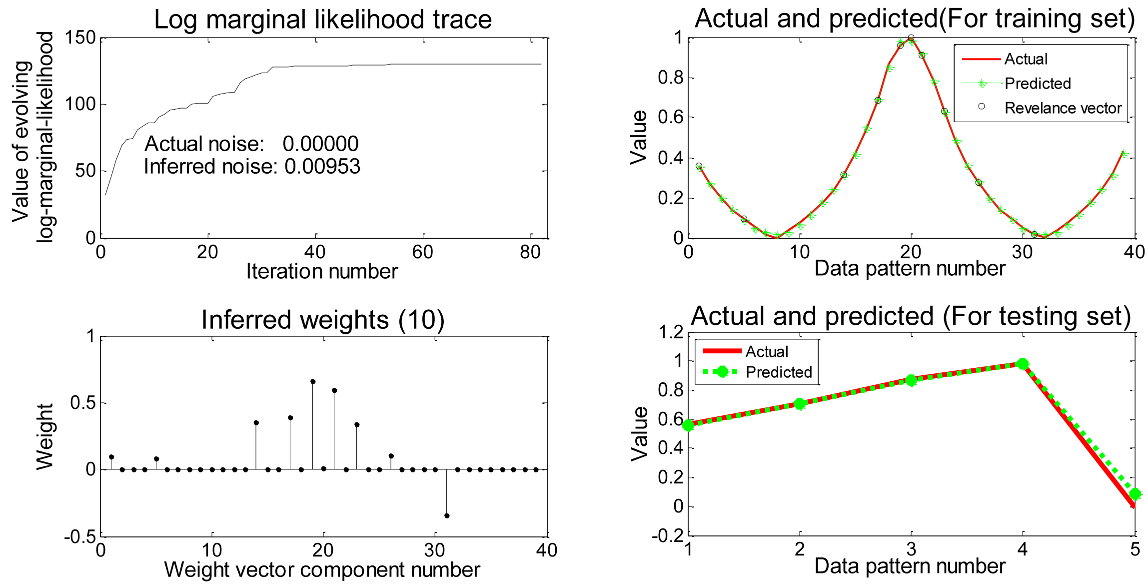

Fig. 5 Inferred weights and forecasting performance when $\sigma = 0.5$

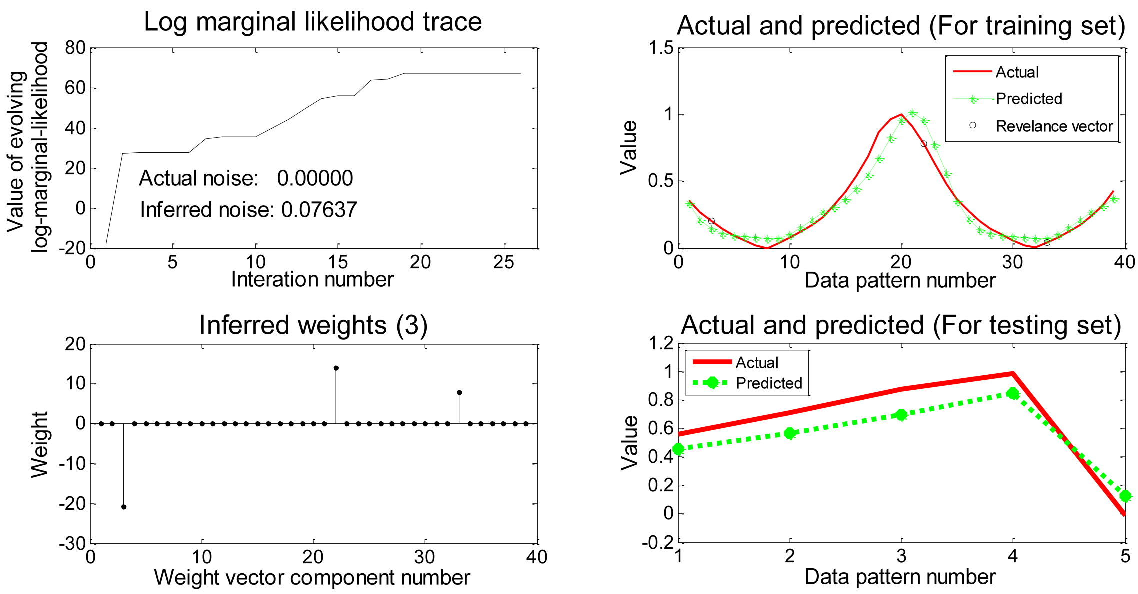

Fig. 6 Inferred weights and forecasting performance when $\sigma = 5$

(3) Parameter of Kernel Function

In this paper, we choose RBF as the kernel function, owing to it is suitable to solve the nonlinear mapping problems. There is only one parameter to be determined, namely, the kernel width parameter σ of RBF.

As shown in Fig. 4 - Fig. 6, with the increasing of σ, the iteration number decreases gradually but the inferred noise increases correspondingly. The number of relevance vectors reduces significantly because the Gaussian distribution becomes smoother. Although the structure of model becomes sparser with the increasing of σ, the generalization ability and fitting performance degrade due to a small number of relevance vectors does not satisfy to express the actual result accurately.

When $\sigma = 0.1$ as shown in Fig. 4, although RVM model trains very well, the generalization ability is not good, and the relevance vectors are too many, so the model is not sparse. When $\sigma = 0.5$ as shown in Fig. 5, RVM model not only obtains better fitting result, but also has better generalization and prediction performance, the number of relevance vectors reduces significantly, so the model is sparser. However, when the value of σ is large enough, see $\sigma = 5$ in Fig. 6, although the model is sparse enough, the generalization ability and fitting performance are poor, the prediction values diverge from the actual values, thus the result is not reasonable.

In practice, kernel width σ has a great influence on the performance of the RVM model. Thus we apply MPSO algorithm with 10-fold cross-validation to optimize the parameter of RBF for RVM model. The obtained optimal parameters of MPSO-RVM model are achieved in TABLE II.

Table II
Parameter of Kernel Function Optimized by MPSO

Parameter in Power System	kernel width parameter σ
IN1	1.202
VN1	1.027

B. Experiment Result and Discussion

In this paper, we propose a mixed MPSO-RVM model to predict the crucial parameters intervals in satellite power system. Based on Sparse Bayesian toolbox (http://www.miketipping.com/downloads.htm), we develop a MPSO-RVM model to achieve the purpose of monitoring the operation state of satellite power system.

In the modeling stage, we set the optimal solution obtained by MPSO algorithm as kernel parameter of RBF. Then establish RVM model and attain the prediction interval. Eventually, calculate the average of the prediction interval as prediction value, and calculated the accuracy criteria to evaluate the prediction model. In order to analyze the performance of MPSO-RVM model, we compare it with RVM model and PSO-SVR model. RVM model only adopts the theory of RVM to predict the telemetry data in satellite power system without having the stage of parameter optimization, in which the RBF kernel parameter σ=2. PSO-SVR model adopts PSO algorithm to optimize the hyper parameter of SVR model, in which the penalty factor C=89.85, the RBF kernel parameter σ=1.51 and the ε-insensitive loss function parameter ε=5.56E-4. Then establish SVR model to predict the values of crucial parameters in satellite power system.

Comparisons of performance among the MPSO-RVM model, RVM model and PSO-SVR model are shown in Table III and Table IV. The prediction accuracies of three models established in the experimental data are summarized in Table III. As shown in Table III, both the prediction results and prediction accuracies of MPSO-RVM model and PSO-SVR model are more excellent than RVM model, and the performance of MPSO-RVM model is similar to PSO-SVR model in the IN1 and VN1 sequences. However, the fitting accuracy, generalization ability and prediction accuracy of MPSO-RVM model is slightly higher than PSO-SVR model and the MPSO-RVM model plays better performance in solving the nonlinear mapping problem than PSO-SVR model. Moreover, comparing with RVM model, the employment of MPSO algorithm avoids the problem of over-fitting, and it also accelerates convergent performance and enhances global optimization capability. In addition, the prediction error in MPSO-RVM model is smaller than that of other two models. It can be seen that root mean squared error (RMSE), mean average percentage error (MAPE) and normalized mean square error (NMSE) are 0.0397, 0.0734 and 0.9841 in VN1 sequence, thus the prediction accuracy of MPSO-RVM is better than other two models. The results indicate the MPSO-RVM model is more suitable to avoid satellite accidences and faults via predicting the parameters in the satellite power system.

Table III
Comparisons of Prediction Accuracies among MPSO-RVM, RVM and PSO-SVR

Parameter in Power System	Prediction Model	Training set			Testing set		
		RMSE	MAPE (%)	NMSE	RMSE	MAPE (%)	NMSE
IN1	MPSO-RVM	0.2379	1.0367	0.8165	0.2439	1.2159	0.8354
	RVM	0.7876	3.170	8.1084	8.4474	6.0512	8.6821
	PSO-SVR	0.2405	1.0486	0.8179	0.3229	1.4432	0.8805
VN1	MSPO-RVM	0.0246	0.0501	0.9511	0.0397	0.0734	0.9841
	RVM	0.0122	0.0248	0.8140	18.6935	8.0334	9.5292
	PSO-SVR	0.0283	0.0523	1.2614	0.0421	0.0798	1.3542

Table IV
Comparisons of MPSO-RVM, RVM and PSO-SVR

Parameter in Power System	Prediction Model	Train-time(s)	Test-time(s)	Number of relevance (support) vectors
	MPSO-RVM	95.2554	0.2322	9
IN1	RVM	1.5109	0.1410	3
	PSO-SVR	56.1248	1.2356	21
	MPSO-RVM	82.3715	0.1537	5
VN1	RVM	1.2106	0.0898	1
	PSO-SVR	53.5683	1.1379	18

The train-time, test-time and the number of relevance (support) vectors of three models established in the experimental data are summarized in Table IV, where the number of relevance (support) vectors reflects the sparse properties of the model. As shown in Table IV, the train-time of RVM model is smallest owing to the training procedure of it does not have the stage of parameter optimization. While the train-time of MPSO-RVM model is longer than PSO-SVR model, because the procedure of MPSO-RVM model searching iteratively the hyper parameter with maximum marginalized likelihood function is more complex than the procedure of PSO-SVR model solving convex quadratic programming problem. However, the number of relevance vectors of MPSO-RVM model is smaller than that of PSO-SVR model. Thus the MPSO-RVM model obtains sparser structure, which makes the time spent in the testing procedure less. Although the number of relevance vectors of RVM model is least and the train-time and test-time of it are smallest, the prediction performance is worse due to too small number of relevance vectors does not make the prediction result reasonably. As a consequence, MPSO-RVM model plays an excellent role in the actual problems which require more stringent response time, such as online prediction system and real-time testing system.

In experiment analysis, Fig. 7 and Fig. 8 demonstrate the weight component of IN1 and VN1 respectively. Based on Fig. 7 and Fig. 8, the vast majority of weight components tend to zero, the number of relevance vectors is small correspondingly, therefore the structure of the MPSO-RVM model is sparser.

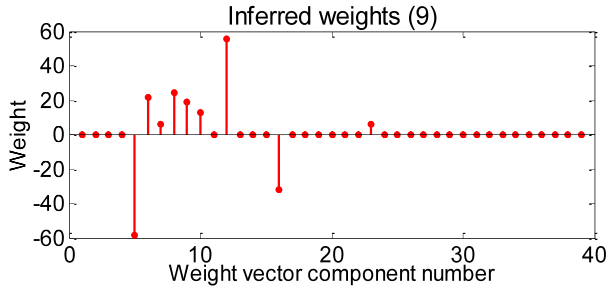

Fig. 7 The schematic diagram of weight component in IN1

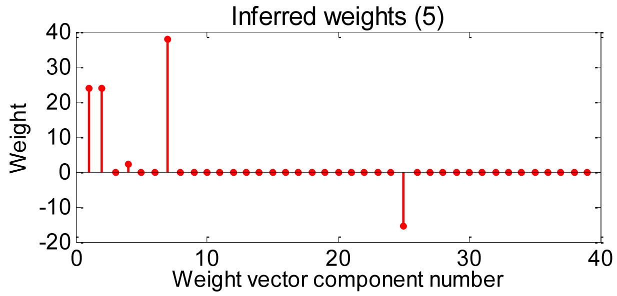

Fig. 8 The schematic diagram of weight component in VN1

According to the regression principle of RVM introduced in Section II, the final prediction model obtained by MPSO-RVM is a normal probability distribution. Therefore, the MPSO-RVM model could obtain the prediction interval under a certain confidence level according to the knowledge of probability theory. In general, we select the average of the prediction interval as the prediction value of the input sample. In order to present a visualized performance of MPSO-RVM model, Fig. 9 and Fig. 10 depict the prediction results of IN1 and VN1 at the confidence level of 95.45%, respectively. The prediction results not only display the prediction values, but also show the prediction intervals of volatility. As shown in Fig. 9 and Fig. 10, only one experimental sample out of the prediction interval of IN1, and all experimental samples fall into the prediction interval of VN1. The results indicate the MPSO-RVM model has higher reliability and reference in parameter interval prediction of satellite power system.

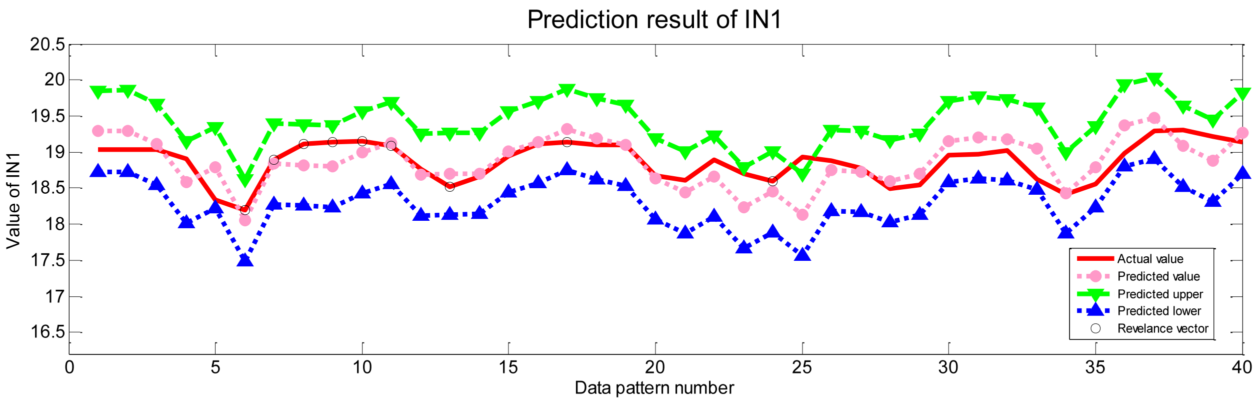

Fig. 9 Prediction result of IN1 at the confidence level of 95.45%.

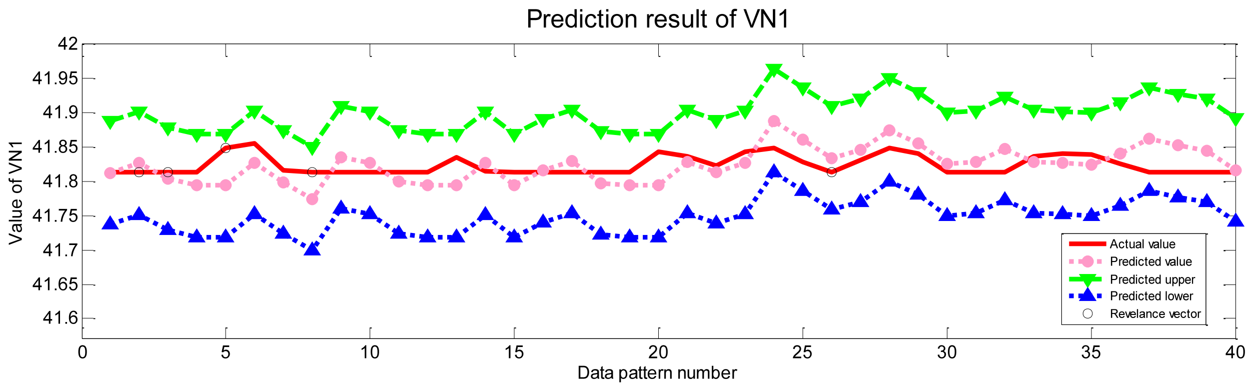

Fig. 10 Prediction result of VN1 at the confidence level of 95.45%.

Comparing with existing prediction models, the model we proposed has several advantages:

1) The existing prediction models usually apply traditional regression methods such as Artificial Neural Network, SVM and so on whose results are just prediction values. Unfortunately the prediction value could not estimate the actual value without bias estimation, so they could not obtain creditable prediction results. However, the prediction result of the MPSO-RVM model is an interval which obtains the range of satellite power system parameter under a certain confidence level, and avoids the problem of unbiased estimation.

2) Procedure of MPSO-RVM model optimizing the hyper parameter is more complex than that of PSO-SVR model, but the test-time spent in MPSO-RVM model is much less than that of PSO-SVR model. At the same time, the prediction efficiency and performance of MPSO-RVM model is much higher than that of PSO-SVR due to a sparser solution obtained by MPSO-RVM model.

3) For complex satellite power system data, although PSO-SVR has a better performance on short-term prediction, once the prediction value deviates from the actual value, the prediction result also deviates from the actual result correspondingly. However, the prediction result of the MPSO-RVM is a desirable prediction interval rather than prediction value. The prediction interval describes the probability distribution of the parameter in satellite power system.

In addition, the performance of the method we proposed on forecasting satellite power system parameter interval can be further improved as well. Here are some guidelines for future studies in work.

1) The prediction intervals are symmetric with the prediction values. Although the method could depict the variation range of parameters vividly, the prediction values may diverge from the actual values in some situations.

2) In order to reduce the train-time cost, other optimization algorithms can be applied to find out the optimal parameter instead of MPSO algorithm.

VI. Conclusions

The power system is one of the important subsystems in satellite, which has a direct influence on the working state, reliability and operational life span of the satellite. In this paper, a new thought is presented to improve the accuracy of parameter forecasting. It believes that the key solution is to establish a RVM model and optimize its kernel parameter. Based on this, a mixed MPSO-RVM model to forecast the satellite power system parameter interval is proposed. The experimental results display that RMSE, MAPE and NMSE are respectively 0.0397, 0.0734 and 0.9841 in VN1 sequence. It demonstrates MPSO-RVM model has higher prediction accuracy than RVM model and PSO-SVR model. The number of relevance vectors is 5 in VN1 sequence which is less than that of support vectors. Thus, the structure of the model is sparser which makes the test-speed faster, the test-time shorter and the generalization ability higher. As a consequence, MPSO-RVM model is more suitable for the practical requirements. Similarly, the shortcomings of the MPSO-RVM model are also obvious. The training process is more complex and the train-time is slightly longer.

Significantly, the prediction results of MPSO-RVM model are normal distributions, thus the prediction intervals in a certain confidence level can be obtained by the theory of normal distribution. The forecasting results show that the majority of the samples are located in the prediction intervals obtained at the confidence level of 95.45%. The results also indicate the MPSO-RVM model simulates the dynamic trend of satellite power system parameter well. Thus it is more suitable for the practical requirements to avoid satellite accidences and faults. More importantly, the proposed method could be further applied to forecast the crucial parameters of other components in satellite.

References

[1] Mahdi M C, Jaafer S, Shehab A A R, "Design and Implementation of an Effective Electrical Power System for Nano-Satellite," International Journal of Scientific & Engineering Research, vol. 5, no. 5, pp. 29-35, 2014.

[2] L. Jinkyu, E. Kim, K.G. Shin, "Design and management of satellite power systems," *IEEE34th Real-Time Systems Symposium (RTSS)*, pp. 97–106, 2013.

[3] Zhong Z, Pi D, "Forecasting Satellite Attitude Volatility Using Support Vector Regression with Particle Swarm Optimization," *IAENG International Journal of Computer Science*, vol. 41, no. 3, pp. 153-162, 2014

[4] Dong X, Pi D, "An Effective Method for Mining Quantitative Association Rules with Clustering Partition in Satellite Telemetry Data," *2014 Second International Conference on Advanced Cloud and Big Data (CBD). IEEE Computer Society*, pp. 26-33, 2014.

[5] Pan D W, Liu D T, Zhou J, Zhang G Y "Anomaly detection for satellite power subsystem with associated rules based on Kernel Principal

Component Analysis," *Microelectronics Reliability*, vol. 55, no. 9, pp. 2082-2086, 2015.
[6] Wang D, Zhang S, Shuang W, et al, "A fault detection method based on wavelet theory for satellite power systems," *Control and Decision Conference (CCDC), 2015 27th Chinese. IEEE*, pp. 581-586, 2015.
[7] Xie S, Peng X, Zhong X, "Fault Diagnosis of the Satellite Power System Based on the Bayesian Network," *The 8th International Conference on Computer Science & Education (ICCSE 2013)*, pp. 1004-1008, 2013.
[8] Fang H, Shi H, Xiong Y, et al, "The component-level and system-level satellite power system health state evaluation method," *2014 Prognostics and System Health Management Conference (PHM-2014 Hunan)*, pp. 683-688, 2014.
[9] Cortes C, Vapnik V, "Support-vector networks," Machine learning, vol. 20, no. 3, pp. 273-297, 1995.
[10] Yang Q, Yang Q, Yan W J, et al "PSO based LS-SVM approach for fault prediction of primary air fan," Chinese Automation Congress (CAC), 2015. IEEE, pp. 75-80, 2015.
[11] Chen T T, Lee S J, "A weighted LS-SVM based learning system for time series forecasting," *Information Sciences*, vol. 299, no. 1, pp.99-116, 2015.
[12] Tipping M E, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, no. 3, pp. 211-244, 2001.
[13] Wang L, Liu Z, "RVM with wavelet kernel combined with PSO for short-term load forecasting in electric power systems," Przegląd Elektrotechniczny, vol. 88, no. 1B, pp. 146-149, 2012.
[14] Jin H, "Research on Automobile Engine Failure Recognition Technology Based on Improved PSO-RVM Algorithm," *Applied Mechanics & Materials*, vol. 727-728, pp. 757-760, 2015.
[15] Liu Y, Yan J, Han S, et al, "An optimized short-term wind power interval prediction method considering NWP accuracy," *Chinese Science Bulletin*, vol. 59, no. 59, pp. 1-9, 2014
[16] Bai S. "Import and Export Trade Prediction Algorithm Based on Multi-kernel Weighting RVM Model Optimized by PSO," *Computer & Modernization*, 2014.
[17] Fei S, He Y, Ma X, et al, "A Hybrid Model of RVM and PSO for Dissolved Gases Content Forecasting in Transformer Oil," *Recent Patents on Electrical & Electronic Engineering*, vol. 6, no. 6, pp. 183-189, 2013.
[18] Wang J, Duan X H, Li Y, et al, "Prediction of Aero Engine Fault by Relative Vector Machine and Genetic Algorithm Model," *Advanced Materials Research*, 2014, vol. 998-999, pp. 1033-1036, 2014.
[19] Gharib A, Benhra J, Chaouqi M, "A performance comparison of PSO and GA applied to TSP," International Journal of Computer Applications, no. 15, pp. 975-8887, 2015.
[20] Tipping M E, "The Relevance Vector Machine," *Advances in Neural Information Processing Systems*, pp. 652-658, vol. 12, no. 3, 1999.
[21] Bishop C M, Tipping M E, "Variational relevance vector machines," *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 46-53, vol. 28, no. 3, 2002.
[22] Wang G, Yang Y, Xie Q, et al, "Force based tool wear monitoring system for milling process based on relevance vector machine" Advances in Engineering Software, vol. 71, no. 3, pp. 46-51, 2014.
[23] Jiang H, Xiao Y, Li J, et al, "Prediction of the Melt Index Based on the Relevance Vector Machine with Modified Particle Swarm Optimization," *Chemical Engineering & Technology*, vol. 35, no. 5, pp. 819–826, 2012.
[24] Yan J, Liu Y, Han S, et al, "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine," *Renewable & Sustainable Energy Reviews*, vol. 27, no. 6, pp. 613–621, 2013.
[25] Kennedy J, Eberhart R, "Particle swarm optimization" in 1995 Proc. proceedings of IEEE International conference on neural networks, pp. 1942-1948, 1995.
[26] Kachitvichyanukul V, "Comparison of Three Evolutionary Algorithms: GA, PSO, and DE," *Journal of Electrical Engineering & Technology*, vol. 12, no. 3, pp. 215-223, 2012.
[27] Shi Y, Eberhart R. A, "modified particle swarm optimizer" in 1998 *Proc. Proceedings of IEEE International Conference on Evolutionary Computation*, pp. 69-73, 1998.
[28] Abdulkadir M, Yatim A H M, Yusuf S T, "An Improved PSO-Based MPPT Control Strategy for Photovoltaic Systems," *International Journal of Photoenergy*, vol. 16, no. 27, pp. 14298-14308, 2014.
[29] Basu M, "Modified Particle Swarm Optimization for Non-smooth Non-convex Combined Heat and Power Economic Dispatch," *Electric Power Components & Systems*, vol. 43, no. 19, pp. 2146–2155, 2015.
[30] Jiang S, Ji Z, Wang Y, "A novel gravitational acceleration enhanced particle swarm optimization algorithm for wind–thermal economic emission dispatch problem considering wind power availability," *International Journal of Electrical Power & Energy Systems*, vol. 73, no. 2015, pp. 1035-1050, 2015.