Yong-liang LIANG[1], Ke-jun LI[1]， Bao-guang ZHAO[2], Jian-guo Zhao[1,3], Lin Niu[3]

Key Laboratory of Power System Intelligent Dispatch and Control of Ministry of Education (Shandong University) (1)
Laiwu Power Supply Company of Shandong Electric Power Corporation (2)
State Grid of China Technology Colleague (3)

Particle swarm optimization-based Fast Relevance Vector Machine for forecasting dissolved gases content in power transformer oil

Abstract. *Forecasting of dissolved gases concentration in power transformer is very significant to detect incipient failures of transformer early and ensure hassle free operation of entire power system. A forecasting model based on Particle Swarm Optimization –Fast Relevance Vector Machine (PSO-FRVM) is proposed in this paper. PSO is utilized to optimize the free parameter of the Gaussian kernel function to improve the forecasting performance. The Matlab program testify the correctness and validity of the model.*

Streszczenie. *W artykule przedstawiono metodę prognozowania rozpływu gazów w transformatorze elektrycznym, opartą na zbudowanym modelu. W tworzeniu modelu wykorzystano Optymalizację Stadną Cząsteczek z maszyną opartą na wektorach istotnych (ang. PSO-FRVM). Metoda PSO wykorzystana została do optymalizacji doboru parametru wolnego w funkcji jądra Gaussa dla polepszenia jakości prognozowania. Weryfikację przeprowadzono w programie Matlab (**Prognozowanie zawartości gazów rozpuszczonych w oleju transformatorowym przy wykorzystaniu metody RVM – Optymalizacja Stadna Cząsteczek**).*

Keywords: Relevance vector machine; Particle swarm optimization; Dissolved gas; Gaussian kernel function.
Słowa kluczowe: Maszyna oparta na wektorach istotnych, Optymalizacja Stadna Cząsteczek, gazy rozpuszczone, funkcja jądra Gaussa

Introduction

Transformer is a major apparatus in power system. The smooth operation of transformers influences the security and reliability of the power grid greatly, so it makes significant sense to detect the incipient fault of transformer. The gas generated when insulating oil understands heat or electricity stress can reflect the type and severity of transformer fault, thus dissolved gas analysis (DGA) becomes the most efficient fault diagnosis technique. If future dissolved gas concentration can be forecasted according to historical data, transformer fault will be detected earlier, and the loss will be reduced minimally.

In recent years, different techniques are applied in transformer dissolved gas concentration forecasting, such as grey model (GM) [1, 2], artificial neural networks (ANN) [3, 4], support vector machine (SVM) [5, 6]and so on. ANN is one of the most commonest techniques in nonlinear forecasting, however, ANN has its inherent defects, one of which is the need of large amount of data. While the long historical data can decrease the forecasting accuracy, so ANN technique doesn't show satisfying performance in this area. GM can get superior forecasting accuracy with a small quantity of historical data. However, GM depicts a monotonously increasing or decreasing process, the dissolved gas concentration in transformer doesn't conform to this process because it is influenced by outside environment. So there is always a certain bias in GM method. SVM has excellent performance in dealing with a small amount of data and applied in dissolved gas forecasting with better effects than ANN and GM. But SVM has several free parameters, which increases the computing complexity.

Relevance vector machine (RVM) is a new machine learning method [7]. RVM learning approach is based on Bayesian estimation theory, which can be applied for both classification and regression problems. Compared with SVM, RVM can obtain sparser model and need less free parameters, besides kernel functions doesn't need to satisfy Mercer rule. Right now RVM is widely used in many areas and get satisfying performance [8-10].

The structure of this paper is as follows: in section II, theory of traditional RVM and its fast algorithm is introduced; section III introduces the principle of particle swarm optimization (PSO); in section IV, particle swarm optimization-based fast relevance vector machine for forecasting dissolved gases content in power transformer oil model is created; section V testifies the feasibility of relevant theory and model; section VI concludes this paper..

Relevance Vector Machine

Traditional RVM

RVM aims at solving the following problem: we are given a set of examples of input vectors $\{\boldsymbol{x}_i\}^{\mathrm{N}}{}_{i=1}$ along with corresponding targets $\{t_i\}^{\mathrm{N}}{}_{i=1}$. The former $\boldsymbol{x}_i=[\xi_1, \xi_2, ..., \xi_m]$ and the latter t_i might be real values (in regression) or class label (in classification). From the training process, we wish to learn a model of the dependency of the target on the input vector with the purpose of making accurate predictions t for unseen result of $\boldsymbol{x}$. Dissolved gas concentration forecasting belongs to regression problem, so in this paper we just introduce the regression application of RVM. Input vectors and targets have the following relation [7]:

$$t_n = \sum_{i=1}^{M} \omega_i K(\boldsymbol{x}_n, \boldsymbol{x}_i) + \varepsilon_n \tag{1}$$

Where $K(\boldsymbol{x}_n, \boldsymbol{x}_i)$ is the kernel function, Gaussian function is used in this paper; ω_i is the weight of basis function $\phi_i=[K(\boldsymbol{x}_1, \boldsymbol{x}_i), K(\boldsymbol{x}_2, \boldsymbol{x}_i)..., K(\boldsymbol{x}_N, \boldsymbol{x}_i)]^{\mathrm{T}}$, $\boldsymbol{\omega}=[\omega_1, \omega_2,...\omega_M]$, ε_n is noise subjected to mean-zero Gaussian $\mathrm{N}(0, \sigma^2)$. Thus, the likelihood of the complete data set can be written as:

$$\begin{aligned} p(\mathbf{t} \mid \sigma^2, \boldsymbol{\omega}) &= \prod_{i=1}^{N} N(t_i \mid y(\boldsymbol{x}_i, \boldsymbol{\omega}), \sigma^2) \\ &= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{\|\mathbf{t}-\boldsymbol{\Phi}\boldsymbol{\omega}\|}{2\sigma^2}\right) \end{aligned} \tag{2}$$

$\boldsymbol{\Phi}$ is the design matrix composed of basis functions ϕ:

$$\boldsymbol{\Phi}=\begin{bmatrix} 1 & K(\boldsymbol{x}_1, \boldsymbol{x}_1) & K(\boldsymbol{x}_1, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_1, \boldsymbol{x}_N) \\ 1 & K(\boldsymbol{x}_2, \boldsymbol{x}_1) & K(\boldsymbol{x}_2, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_2, \boldsymbol{x}_N) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & K(\boldsymbol{x}_N, \boldsymbol{x}_1) & K(\boldsymbol{x}_N, \boldsymbol{x}_2) & \cdots & K(\boldsymbol{x}_N, \boldsymbol{x}_N) \end{bmatrix} \tag{3}$$

To overcome the problem of over-fitting, $\boldsymbol{\omega}$ is assumed to $\mathrm{N}(0, \boldsymbol{\alpha}^{-1})$.. Let $\beta=\sigma^{-2}$, make $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ conform to Gamma distribution:

(4) $p(\boldsymbol{\alpha}) = \prod_{i=1}^{N} Gamma(\boldsymbol{\alpha} \mid a, b)$

(5) $p(\beta) = \prod_{i=1}^{N} Gamma(\beta \mid c, d)$

To make these priors non-informative, we might set their parameters to small values, here a=b=c=d=10^{-4}. Then, given a new test point, $\boldsymbol{x}$*, predictions are made for the corresponding target t*:

(6) $$p(t^* \mid \mathbf{t}) = \int p(t^* \mid \boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2) * p(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2 \mid \mathbf{t}) d\boldsymbol{\omega} d\boldsymbol{\alpha} d\sigma^2$$

From Bayes' rule, we can get:

(7) $p(\boldsymbol{\omega}, \boldsymbol{\alpha}, \sigma^2 \mid \mathbf{t}) = p(\boldsymbol{\omega} \mid \mathbf{t}, \boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}, \sigma^2 \mid \mathbf{t})$

We make approximation to equation 7, the right first part could be expressed in algebra form, and the second part is approximated with delta function. Thus, the learning process equals to maximization of $p(\boldsymbol{\alpha}, \sigma^2|\boldsymbol{t}) \propto p(\boldsymbol{t}|\boldsymbol{\alpha}, \sigma^2) p(\boldsymbol{\alpha}) p(\sigma^2)$. we need find α_{MP}、σ^2_{MP} to satisfy:

(8) $(\boldsymbol{\alpha}_{MP}, \ \sigma^2_{MP}) = \arg\max_{\alpha, \sigma^2} p(\boldsymbol{\alpha}, \sigma^2 \mid \mathbf{t})$

We maximize the marginal likelihood function $p(\mathbf{t}|\boldsymbol{\alpha},\sigma^2)$or its logarithmic form (equation 9) with the objectivity to find the answer of equation 8:

(9) $$L(\boldsymbol{\alpha}) = \log p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2) = \log \int_{-\infty}^{+\infty} p(\mathbf{t}|\boldsymbol{\omega}, \sigma^2) p(\boldsymbol{\omega}|\boldsymbol{\alpha}) d\boldsymbol{\omega} = -\frac{1}{2}\left[N \log 2\pi + \log|\boldsymbol{C}| + \mathbf{t}^{\mathrm{T}} \boldsymbol{C}^{-1} \mathbf{t} \right]$$

where: $\mathbf{C}=\sigma^2\mathbf{I}+\mathbf{\Phi A^{-1}\Phi^T}$.

The iteration process and detailed expression of symbols in $\mathbf{C}$ can be seen in reference [7].

Fast algorithm

The defect of traditional RVM is its long training process which restricts the ability of dealing with large amount of data and real-time calculation. Thus, the fast algorithm came into being [11].

Considering the dependency of L(α) on a single α_i, $i \in \{1,2,...M\}$. we can decompose $\mathbf{C}$ in 9 as:

(10) $\mathbf{C} = \sigma^2\mathbf{I} + \sum_{m \neq i} \alpha_m^{-1} \phi_m \phi_m^T + \alpha_i^{-1} \phi_i \phi_i^T = \mathbf{C}_{-i} + \alpha_i^{-1} \phi_i \phi_i^T$

where $\mathbf{C}_{-i}$ is the $\mathbf{C}$ with contribution of basis vector i removed.

From equation 10, equation 9 can be written as:

(11) $$L(\boldsymbol{\alpha}) = -\frac{1}{2}[N \log 2\pi + \log|\mathbf{C}_{-i}| + \mathbf{t}^{\mathrm{T}} C_{-i}^{-1} \mathbf{t} - \log \alpha_i + \log(\alpha_i + \phi_i^{\mathrm{T}} C_{-i}^{-1} \phi_i) - \frac{(\phi_i^{\mathrm{T}} C_{-i}^{-1} \mathbf{t})^2}{\alpha_i + \phi_i^{\mathrm{T}} \mathbf{C}_{-i}^{-1} \phi_i}] = L(\boldsymbol{\alpha}_{-i}) + \frac{1}{2}[\log \alpha_i - \log(\alpha_i + s_i) + \frac{q_i^2}{\alpha_i + s_i}] = L(\boldsymbol{\alpha}_{-i}) + \ell(\alpha_i)$$

For simplification of description, we define sparse factor s_i: $s_i = \phi_i^T \mathbf{C}_{-i}^{-1} \phi_i$; quality factor q_i: $q_i = \phi_i^T \mathbf{C}_{-i}^{-1} \mathbf{t}$. Analysis of $l(\alpha_i)$ shows that L(α) has a unique maximization with respect to α_i:

(12) $\alpha_i = \dfrac{s_i^2}{q_i^2 - s_i}$ $\qquad q_i^2 > s_i$

(13) $\alpha_i = \infty$ $\qquad q_i^2 \leq s_i$

To simplify calculation process, we define: $S_i = \phi_i^T \mathbf{C}^{-1} \phi_i$; $Q_i = \phi_i^T \mathbf{C}^{-1} \mathbf{t}$. Also the inversion of can be avoided by using the Woodbury identity [11] to write:

(14) $S_i = \phi_i^T B \phi_i - \phi_i^T B \Phi \Sigma \Phi^T B \phi_i$

(15) $Q_i = \phi_i^{\mathrm{T}} B \mathbf{t} - \phi_i^{\mathrm{T}} B \Phi \Sigma \Phi^{\mathrm{T}} B \mathbf{t}$

Thus we can compute s_i and q_i:

(16) $s_i = \dfrac{\alpha_i S_i}{\alpha_i - S_i}$

(17) $q_i = \dfrac{\alpha_i Q_i}{\alpha_i - S_i}$

Based on the analysis above, we conclude that: if $q_i^2 > s_i$, we add ϕ_i to $\mathbf{\Phi}$ and update α_i in next iteration when ϕ_i is not in the model, or update α_i directly when ϕ_i is in the model; if $q_i^2 < s_i$, we delete the corresponding basis function ϕ_i.

In regression, $B=\sigma^{-2}\mathbf{I}$, $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ just have the basis functions in model [7], thus computation complexity is decreased a lot。

The steps of Fast RVM algorithm is as follows [11].

(1) Initialize σ^2 to some sensible value;

(2) Initialize one basis function and compute:

$$\alpha_i = \frac{\|\phi_i\|^2}{\|\phi_i^{\mathrm{T}} \mathbf{t}\|^2 / \|\phi_i\|^2 - \sigma^2}$$

All other α_m are set to infinity;

(3) Explicitly compute $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$ for all M base functions along with s_i and q_i.

(4) Select a candidate basis vector ϕ_i from all M basis ones;

(5) Conpute $\theta i = q_i^2 - s_i$;

(6) If θi >0 and $\alpha_i < \infty$ (ϕ_i is in the model), re-estimate α_i;

(7) If θi >0 and $\alpha_i < \infty$, add ϕ_i to the model and update α_i;

(8) If $\theta i \leqslant 0$ and $\alpha_i < \infty$, delete ϕ_i from the model, and set α_i to infinity;

(9) Update σ^2;

(10) Recompute $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$, and all s_i and q_i, $i \in \{1,2,...M\}$;

(11) If converged terminate, otherwise goto 4.

Particle Swarm Optimization

Particle Swarm Optimization is a populated search method and has high generality. PSO performs searches using a population (called swarm) of individuals (called particles) that are updated from iteration to iteration [5]. The basic principle is: every particle is a candidate solution of the problem with a fitness value, and each particle moves in the direction of its previous best position (*pbest*) and global best position (*gbest*) with the restriction of the velocity limit. Compared with genetic algorithm, PSO doesn't have the complex operation such as cross and mutation and have excellent optimization performance, thus PSO is widely used [12-15].

In a D dimension search space, m particles constitute a swarm, the position of ith particle is $\boldsymbol{x_i}=(x_{i1}, x_{i2}, \ldots, x_{iD})$, and the velocity of which is $\boldsymbol{v_i}=(v_{i1}, v_{i2}, \ldots, v_{iD})$, $i \in \{1,2,...M\}$. The

velocity and position of particles can be updated by the following equations:

(18)
$$v_{ij}(t+1)=\omega v_{ij}(t)+c_1r_1(p_{ij}-x_{ij}(t))+c_2r_2(p_{gj}-x_{ij}(t))$$

(19) $$x_{ij}(t+1)=x_{ij}(t)+v_{ij}(t+1)$$

In iteration, velocity is limited to:

(20) $$\begin{cases} v_{ij}=V_{\max} & if \quad v_{ij}>V_{\max} \\ v_{ij}=-V_{\max} & if \quad v_{ij}<-V_{\max} \end{cases}$$

where: $i\in\{1,2,...M\}$, $j\in\{1,2,...D\}$, t is the evolutionary generation. v_{ij} is the velocity of particle i on dimension j; x_{ij} is the position of particle i on dimension j. Positive constant c_1 and c_2 are personal and social learning factors, whose values are usually set to 2. The r_1 and r_2 are random function in the range [0,1]. The inertia weight ω is used to balance the global exploration and local exploitation. The value of ω influence the local and global search ability a lot, so linear decline srategy is adopted in this paper:

(21) $$\omega=\omega_{\max}-\frac{\omega_{\max}-\omega_{\min}}{T_{\max}}t$$

$\omega_{\max}$、$\omega_{\min}$ is the maximum and minimum respectively; t、$T_{\max}$ is current and maximum iteration. When iteration generation is above a certain value, such as 500，$\omega=\omega_{\min}$.

PSO-FRVM based dissolved gas concentration forecasting model

4.1 Constructing training sample sets

After essential processing, we get dissolved gas data with the same time interval between each other, $A=[a_1, a_2, \ldots, a_n]$. Then we can get training samples, input matrix and output matrix :

(22) $$\boldsymbol{X}=\begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ a_2 & a_3 & \cdots & a_{m+1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n-m} & a_{n-m+1} & \cdots & a_{n-1} \end{bmatrix}=\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_{n-m} \end{bmatrix}$$

(23) $$\boldsymbol{Y}=\begin{bmatrix} a_{m+1} \\ a_{m+2} \\ \vdots \\ a_n \end{bmatrix}=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-m} \end{bmatrix}$$

where: m is the dimension of input vector. After the training process in fast RVM model, we can get the relevance vectors and corresponding weights and make prediction of the gas concentration in next time interval.

4.2 Optimizing the FRVM parameters with PSO

Compared with SVM, RVM has less parameter which can effect the prediction performance and it is essential to optimize the value of the parameters, where in this paper, is the width of Gaussian kernel function. The optimization process is as follows:

(1) Initialize a population of random positions and velocities;

(2) Fitness evaluation. The leave-one-out cross-validation method is used to evaluate fitness， which is testified to obtain balance between computation complexity and stability. leave-one-out is the method which one of the training samples is taken as validation set in turn, others are taken as training set, then each training sample is validated once. After k times validation, the mean absolute percentage error (MAPE) is taken as the fitness function：

(24) $$MAPE=\frac{1}{k}\sum_{i=1}^{k}\left|\frac{y_i-y_{it}}{y_i}\right|\times 100\%$$

where: y_i is the true value, y_{it} is the prediction value. In this study, the better particle is whose fitness value is less；

(3) Update the global and personal best according to the fitness evaluation results;

(4) Update velocity according to equation 18 and update position according to equation 19;

(5) The same procedures from Step 2 to Step 4 are repeated until stop conditions are satisfied.

PSO-FRVM based forecasting model of dissolved gas concentration is shown in Figure 1:

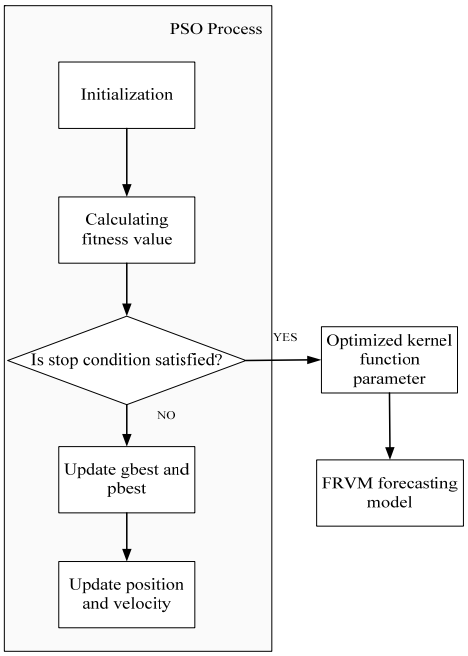

Fig.1. PSO-FRVM based forecasting model of dissolved gas concentration

Empirical analysis for forecasting dissolved gases content in power transformer oil

Dissolved gas in oil include hydrogen(H_2), Methane(CH_4), Ethane(C_2H_6), Ethylene(C_2H_4) and Acetylene(C_2H_2). The gas concentration prediction involves the five kinds of gas.

The model proposed in this paper is validated in Matlab numerical simulation environment. Dissolved gas data of one transformer is listed in Table 1. It shows that concentrations of all kinds of gas increased in fluctuating. As no heat or discharge faults had occur, content of C_2H_2 is zero. Data from 2005-1-22 to 2005-1-31 is chosen as training samples, and data of 2005-2-1 is taken as the validation ones. Dimension of input vector is 4. Optimized value of parameter of kernel function and forecasting results are listed in Table 2(concentration of C_2H_2 is zero, so it is not shown).

Table 2 shows that different kinds of gas have different kernel function parameters which illustrates the importance of the parameter. Besides, the forecasting results are very close to real value with little error, which testifies the effectiveness of the PSO-FRVM model.

The forecasting results and comparison with other methods(including PSO-SVM, GM and ANN) [5] are listed in Table 3. φ_t represents the real value. From the comparison, we can see that the forecasting results of the model proposed in this paper are very close to real value and more accurate than other models except that of C_2H_4 is

somewhat less than PSO-SVM model; PSO-SVM needs to optimize three parameters while the model in this paper just have one to optimize, so PSO-FRVM model not only improve the forecasting performance but also decrease the computation complexity.

Table 1. Data of dissolved gas (μL/L)

Date	H_2	CH_4	C_2H_6	C_2H_4	C_2H_2
2005-1-22	28.6	125.0	38.5	133.0	0
2005-1-23	30.4	117.0	44.2	138.0	0
2005-1-24	27.7	118.0	38.6	132.0	0
2005-1-25	29.9	127.0	37.1	135.0	0
2005-1-26	32.3	134.0	41.9	143.0	0
2005-1-27	29.4	125.0	40.3	149.0	0
2005-1-28	28.0	135.0	41.6	152.0	0
2005-1-29	31.3	136.0	46.4	147.0	0
2005-1-30	30.4	147.0	51.9	151.0	0
2005-1-31	31.8	141.0	45.2	157.0	0
2005-2-01	30.8	149.0	47.9	146.0	0

Table 2. Predicted Data of dissolved gas based on PSO-FRVM (μL/L)

Gas	Parameter of kermel function	Forecasting results
H_2	1.2560	30.7237
CH_4	0.926	149.0096
C_2H_6	0.3956	47.89
C_2H_4	1.4528	151.3526

Table 3. Comparison of predicted results

Gas	φ_t	PSO-FRVM	PSO-SVM	GM	ANN
H_2	30.8	30.7237 0.25%	30.0462 2.4%	31.1708 1.2%	26.8851 12.7%
CH_4	149	149.0096 0.01%	149.7130 0.48%	148.6605 0.23%	143.6773 3.6%
C_2H_6	47.9	47.89 0.02%	49.5109 3.4%	48.5993 1.5%	44.9137 6.2%
C_2H_4	146	151.3526 3.7%	146.3962 0.27%	159.1980 9.0%	152.4498 4.4%

Conclusion

Via ameliorating the parameters' iteration process thus overcoming the weakness of traditional RVM, PSO-FRVM can improve the computation speed substantially. This paper used the PSO method to optimize the parameters of the kernel function, which is the key to determine the predictive ability of RVM, and founded the dissolved gas concentration forecasting model in transformer oil based on PSO-FRVM. An example testified the accuracy of this model on air density prediction in transformer oil and that comparing to SVM, this model could increase the predictive ability while decrease the number of parameters to be optimized.

It should be pointed out that although the overall performance of the model proposed in this paper is superior to others, it can not be guaranteed that it can perform better to all cases, e.g., the prediction on C_2H_4 in this paper, which also indicates that each single predictive method has its defects. Therefore, the next work is to synthesize various methods to do combined prediction on dissolved gas concentration in transformer oil, expecting better prediction result.

REFERENCES

[1] Wang M. H., Hung C.P., Novel grey model for prediction of trend of dissolved gases in oil-filled power appratus, *Electric Power Systems Research*, 67 (2003), No. 1, 53-58

[2] Wang W., Liu Y., Pedrycz W., A Two-Stage Online Prediction Method for a Blast Furnace Gas System and Its Application, *IEEE Trans. on Control Systems Technology*, 19 (2011), No. 3, 507-520

[3] Elragal, H., Lu T.-L., Combination of artificial neural-network forecasters for prediction of natural gas consumption, *IEEE Trans. on Neural Networks*, 11(2000), No. 2, 464-473

[4] Azadeh A, Ghaderi S. F., Sohrabkhani S., A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. *Energy Policy*, 36(2008), No. 7, 2637–2644

[5] Shengwei Fei, Mingjun Wang, Yubin Miao, et al, Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Conversion and Management*, 50(2009), 1604-1609

[6] Shengwei Fei, YuSun, Forecasting dissolved gases content in power transformer oil based on support vector machine with genetic algorithm, *Electric Power Systems Research*, 78(2008), 507-514

[7] Michael E. Tipping. Sparse Bayesian Learning and the Relevance Vector Machine. *Journal of Machine Learning Research*, 1(2001), 211-244

[8] Fereidoun A. Mianji, Ye Zhang, Robust Hyperspectral Classification Using Relevance Vector Machine, *IEEE Trans. on Geosc. and Remote Sensing*, 49(2011), No. 6, 2100-2112

[9] Ioannis Psorakis, Theodoros Damoulas, Mark A., Multiclass Relevance Vector Machines: Sparsity and Accuracy, *IEEE Trans. on Neural Networks*, 21(2000), No. 10, 1588-1598

[10] Liyang Wei, Yongyi Yang, Robert M. Nishikawa, et al. Relevance Vector Machine for Automatic Detection of Clustered Microcalcifications, *IEEE Trans. on Medical Imaging*, 24(2005), No. 10, 1278-1285

[11] Michael E. Tipping, Anita Faul, Fast Marginal Likelihood Maximisation for Sparse Bayesian Models. *Proc. 9th Int.* Workshop Artif. Intell. Statist., Key West, FL, 2003

[12] Fernandez-Martinez, J.L., Garcia-Gonzalo E., Stochastic Stability Analysis of the Linear Continuous and Discrete PSO Models, *IEEE Trans. on Evolutional Computing*, 15(2011), No. 3, 405-423

[13] Das Sharma, K.; Chatterjee, A.; Rakshit, A., A Random Spatial Ibest PSO-Based Hybrid Strategy for Designing Adaptive Fuzzy Controllers for a Class of Nonlinear Systems, *IEEE Trans. on Instrumentation and Measurement*, 61(2012), No. 6, 1605-1612

[14] Bo Liu, Ling Wang, Yihui Jin, An Effective PSO-Based Memetic Algorithm for Flow Shop Scheduling, *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, 24(2007), No. 1, 18-27

[15] Chia-Nan Ko; Ying-Pin Chang; Chia-Ju Wu, A PSO Method With Nonlinear Time-Varying Evolution for Optimal Design of Harmonic Filters, *IEEE Trans. on Power Systems*, 24(2009), No. 1, 437-444

***Authors**: Yongliang Liang, Phd, Shandong University, Institute of Electrical Engeering, jingshi road 17923 Jinan of China, E-mail: windgod0112@163.com; Ke-kun Li, Professor, Phd, Shandong University, Institute of Electrical Engeering, jingshi road 17923 Jinan of China, E-mail: lkjun@sdu.edu.cn; Bao-guang ZHAO, general manager of Laiwu Power Supply Company of Shandong Electric Power Corporation, E-mail:zhaobaoguang-lw@sd.sgcc.com.cn ;Jian-guo Zhao, Professor, Phd, Shandong University, Institute of Electrical Engeering, jingshi road 17923 Jinan of China, E-mail: zhaojianguo@sgtc.sgcc.com.cn; Lin-Niu, Phd, State Grid of China Technology Colleague, south erhuan road 500 Jinan of China, E-mail:niulin2008@126.com*