Optimal Support Vector Regression Algorithms for Multifunctional Sensor Signal Reconstruction

Xin Liu[1], Dan Liu*[2], Yan Zhang[3], Qisong Wang[4], Shen Zhang[5], Hua Wang[6]
[1,5,6]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China
[2,3,4]School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China
*Corresponding author, e-mail: xinliu@hit.edu.cn[1], liudan@hit.edu.cn[2], zyhit@hit.edu.cn[3],
wangqisong@hit.edu.cn[4], shenzhang@hit.edu.cn[5], wanghua@hit.edu.cn[6]

Abstract
The empirical risk minimization methods were often used to estimate the multifunctional sensor regression function in signal reconstruction. The small size of sample data would lead to the problem of poor generalization capability and overfitting. Support vector machine (SVM) is a novel machine learning method based on structural risk minimization, and it can improve generalization capability and restrain overfitting. In this paper, an optimal v-Support Vector Regression (v-SVR) algorithm has been proposed for multifunctional sensor reconstruction, which combined v-SVR with particle swarm optimization (PSO), achieving accurate estimation of both the hyperparameters and reconstruction function. The results of emulation and theory analysis indicate that the proposed algorithm is more accurate and reliable for signal reconstruction.

Keywords: v-SVR, PSO, hyperparameters, multifunctional sensor, signal reconstruction

1. Introduction

In the last decades, multifunctional sensors have received more attention due to the development of microelectronics, micromachining and other related technologies, which can simultaneously detect several different electric or non-electric signals, greatly reduce the size and consumption of the measurement system, and it could be applied into the field of environmental perception and industry measurement [1, 2], also naturally applied into the region of aeronautics, astronautics and micro-mechanical technology. In general, the multifunctional sensing technique can be studied from two related aspects [3]: the physical structure design of the multifunctional sensor for multiple variables sensing usually by exploiting the crossing sensitivity of sensitive components and the development of corresponding algorithm for reconstructing the measured variables. The schematic structure of multifunctional sensing technique is shown in Figure 1, where x_1, …, x_n are the physical quantities under measurement, y_1, …, y_n are the sensor output signals, while $\hat{x}_1$, …, $\hat{x}_n$ are the estimation of the measured quantities that can be obtained through the signal reconstruction algorithm, and this process is also called multifunction sensor signal reconstruction.
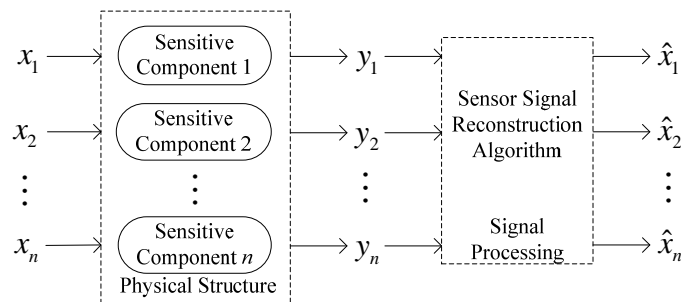

Figure 1. Schematic Structure of Multifunctional Sensing Technique

By now, the study of reconstruction algorithm is becoming more interesting and many signal reconstruction algorithms have been proposed [4-6], while these methods are based on the empirical risk minimization (ERM) principle, which ensures the actual risk close to the value of empirical risk when the sample data set is large. The signal reconstruction is usually a high-dimensional signal processing problem, however, the sample data set obtained from the experiment is small compared to whole measurement range of the multifunctional sensor. In this case, minimizing the empirical risk can not guarantee a small value of actual risk, and thus lead to the overfitting and poor generalization capabilities [7-8]. Support vector machine (SVM) and its modified algorithms could provide powerful and efficient tools that are capable of dealing with the small sample size problem and theoretical bounds on the generalization error through replacing ERM principle with structural risk minimization (SRM) principle, which defines a trade off between the quality of the approximation of given data set and the complexity of approximating function, motivated by statistical learning theory. In recent years, SVM and its modified algorithms have been widely used in many research fields and achieved satisfactory results [9-11].

In this paper, we propose to use a new class of SVR algorithms [12], called *v*-SVR, for sensor signal reconstruction. This algorithm could automatically compute the width of the so-called *ε*-insensitive tube, which must be specified a priori in standard *ε*-SVM methods, thus adjust the generalization accuracy level to the sample data set. Moreover the parameter *v* is asymptotically related to the noise model, therefore, to get better generalization accuracy the specified asymptotical optimal value could be chosen in accordance with the noise model that is in the data set, which is more suitable for the sensor signal reconstruction under the real world condition that the data set are often contaminated by noise.

The main problem in *v*-SVR or *ε*-SVM methods, however, is that of tuning the parameters, because the generalization abilities of these algorithms depend on the choice of kernel parameter, regularization parameter C and parameter *ε* (or *v*), we present a simple and efficient PSO procedure aimed at determining the optimal hyper-parameters and the sensor reconstructed function. In Section II, we briefly review the *ε*-SVM, *v*-SVR and PSO algorithms, and then describe reconstruction algorithm based on the optimal *v*-SVR procedure. In Section III, we build up a simulation model of multifunctional sensor and analyze the experimental result obtained by the proposed approach.

2. Theory and Algorithm
2.1. *ε*-SVR and *v*-SVR

SVM was originally developed for binary classification problem, and then V. Vapnik generalized the results obtained for the pattern recognition problem to the problem of regression by introducing a novel loss function, *ε*-insensitive loss function, which could be defined as follows:

$$L\left(\left|y - f\left(\mathbf{x}\right)\right|_\varepsilon\right) = \begin{cases} 0 & \text{if } \left|y - f(\mathbf{x})\right| \le \varepsilon \\ \left|y - f(\mathbf{x})\right| - \varepsilon & \text{otherwise} \end{cases} \tag{1}$$

For a given independent and identically distributed (i.i.d.) data set $\left\{\mathbf{x}_i, y_i\right\}_{i=1}^{l}$ with the input data $\mathbf{x}_i \in \Box^n$ and output data $y_i \in \Box$. The ε-SVM is intended to estimate the following function:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad \mathbf{w}, \mathbf{x} \in \Box^n, \quad b \in \Box \tag{2}$$

by minimizing the regularized risk functional:

$$\frac{1}{2}\|\mathbf{w}\|^2 + C\frac{1}{l}\sum_{i=1}^{l}\left|y_i - f(\mathbf{x}_i)\right|_\varepsilon \tag{3}$$

Where C is the regularization parameter that control the trade-off between minimizing the model complexity (the former) and the empirical risk (the latter). To minimizing the Equation (3) is

equivalent to the following constrained optimization problem by introducing two set of nonnegative slack variables $\{\xi_i\}_{i=1}^{l}$ and $\{\xi_i^*\}_{i=1}^{l}$ [13]:

$$\underset{\mathbf{w},\xi,\xi^*}{\text{minimize}} \quad J(\mathbf{w},\xi,\xi^*) = \frac{1}{2}\|\mathbf{w}\|^2 + C\frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) \tag{4}$$

Subject to:

$$\begin{cases} \mathbf{w}^T\mathbf{x}_i + b - y_i \le \varepsilon + \xi_i \\ y_i - \mathbf{w}^T\mathbf{x}_i - b \le \varepsilon + \xi_i^* \\ \varepsilon, \xi_i, \xi_i^* \ge 0 \end{cases} \tag{5}$$

The slack variables measure the deviation of data outside the ε-insensitive tube, and they are penalized in the equation (4). Although the parameter ε affects the desired accuracy of the approximation and the sparseness of the solution, it is difficult to find out the optimal value of ε for the lacking of a priori information about the data set. Therefore, it is advisable to automatically compute the ε from the data set, which is the idea of the v-SVR.

In the v-SVR formulation, the value of ε is also a variable, which is traded off against the model complexity and slack variables via a constant $v \in (0,1]$:

$$\underset{\mathbf{w},\xi,\xi^*,\varepsilon}{\text{minimize}} \quad Q(\mathbf{w},\xi,\xi^*,\varepsilon) = \frac{1}{2}\|\mathbf{w}\|^2 + C(v\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)) \tag{6}$$

Subject to the constraint (5), the constraint optimization problem of Equation (6) results in a convex optimization problem with a global minimum by using Lagrange multipliers techniques and dual theorem, which is similar to the Vapnik's ε-SVM, therefore the dual form for the v-SVR optimization problem could be stated as follows:

$$\underset{\alpha,\alpha^*}{\text{maximize}} \quad W(\alpha,\alpha^*) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)\mathbf{x}_i^T\mathbf{x}_j \tag{7}$$

Subject to:

$$\begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i, \alpha_i^* \le C/l \\ \sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \le Cv \end{cases} \tag{8}$$

And the approximating function can be expressed as:

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)\mathbf{x}_i^T\mathbf{x} + b \tag{9}$$

Usually, only some of the coefficients $(\alpha_i^* - \alpha_i)$ are nonzero, and the corresponding data vectors are called support vectors (SVs). Furthermore, to make the v-SVR algorithm nonlinear, the input data vector $\mathbf{x}_i$ can be mapped into a high-dimensional feature space through some nonlinear mapping $\Phi(\square)$, then solve the optimization problem (7) in the feature space, which means the inner product $\mathbf{x}_i^T\mathbf{x}_j$ in (7) is replacing by the inner product of the input vectors induced in the feature space, $\Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j)$. According to Mercer's theorem, these expensive calculations of inner product in the high-dimensional feature space can be

significantly reduced and the explicit form of the nonlinear mapping is no need by choosing a suitable kernel function such that:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j) \tag{10}$$

Then we can get the nonlinear form of Equation (9):

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) k(\mathbf{x}_i, \mathbf{x}) + b \tag{11}$$

The typical choices of kernel function include polynomial kernels, sigmoid kernels and gaussian kernels.

For the v-SVR Algorithm, the theoretical significance of parameter v is that v is an upper bound on the fraction of errors and a lower bound on the fraction of SVs, where errors refer to the training data vectors lying outside the tube and fraction refer to the relative numbers divided by the total number of training data points, thus v can control the number of SVs and the number of data points lying outside the tube. Moreover it has been theoretically proven that parameter v can be asymptotically optimal chosen for a given class of noise model, such as v can be set to 1 or 0.54 for Laplacian or Gaussian noise model, respectively [14].

2.2. PSO for Hyperparameters Selection

It can be found from above that the generalization performance of v-SVR depends on a good setting of C, v and kernel parameter, however, the principled approach for the selection of hyperparameters is still an open and further complicated research area, which is usually treated as user defined inputs that based on a priori knowledge or expertise [15]. Actually the optimal parameters selection can be regarded as an optimal search process, and the estimation accuracy is computed as a function of hyperparameters, therefore optimal hyperparameters can be automatically found by optimization techniques.

PSO is a novel evolutionary computation technique motivated by the social behaviors of flocking birds or swarming insects, which is a population based stochastic optimization technique that can be used for both discrete and continuous optimization problems, and the cooperation and information sharing of an entire flock implies the intelligence and efficiency of algorithm [16]. Each particle is a moving point in the solution space, and the particle's traversal in the search space is influenced by the best solution that it has found, pbest, and the best solution found by the swarm of particles, gbest, respectively. The common PSO algorithm consists of the velocity and position equation as following:

$$
\begin{aligned}
v_i(k+1) &= wv_i(k) + c_1 rand_1(k)(pbest_i(k) - x_i(k)) \\
&\quad + c_2 rand_2(k)(gbest_i(k) - x_i(k))
\end{aligned} \tag{12}
$$

$$x_i(k+1) = x_i(k) + v_i(k+1) \tag{13}$$

Where x is the position information that reflects the value of hyperparamters, v is the velocity information which is dynamically adjusted according to the flying experience of both particle and swarm, w is inertia weight that control the trade-off between the global exploration and local exploitation abilities of the swam. $c1$, $c2$ are acceleration constants, and $rand_1$, $rand_2$ are random number between (0, 1). In this paper, the PSO is applied to v-SVR algorithm to estimate the optimal value of hyperparameters.

2.3. Algorithm

For a multifunctional sensor, any output signal should represent the unique input signal, which can be called one-to-one, otherwise it is impossible to distinguish a input signal from another. Thus, the inverse mapping of multifunctional transfer function is unique based on inverse mapping theorem, and the system equation, according to Figure 1, can be written as:

$$\begin{cases} x_1(t) = g_1(y_1(t),\cdots,y_n(t)) \\ \qquad\vdots \\ x_n(t) = g_n(y_1(t),\cdots,y_n(t)) \end{cases} \tag{14}$$

Then the multifunctional sensor reconstructed signal can be obtained through the estimation of Equation (14),

$$\begin{cases} \hat{x}_1(t) = f_1^{v-svr}(y_1(t),\cdots,y_n(t),\theta_1) \\ \qquad\vdots \\ \hat{x}_n(t) = f_n^{v-svr}(y_1(t),\cdots,y_n(t),\theta_n) \end{cases} \tag{15}$$

Where $f_i^{v-svr}(y_1(t),\cdots,y_n(t),\theta_i)$ is the v-SVR regression estimation of measured quantity x_i, and θ_i is the optimal hyperparameters calculated by PSO.

3. Result and Analysis

To verify the feasibility of the proposed method, a physical model of the two-input/output multifunctional sensor used in the experiment has been built up and shown in Figure 2.

Where x and y are the input signals, which represent the rate of the slide resistor lower side resistances to the entire resistances, u and v (Voltage) are the output signals. To test the ability of the algorithm to match the noise, two independent and identically distributed (iid) noise, ζ_1 and ζ_2 are added to the input signals. According to KCL, the system transfer function can be described as follows:

$$\begin{cases} u = \dfrac{5[2x^* + y^* + 5x^*y^*(2 - x^* - y^*)]}{3 + 5[y^*(1 - y^*) + x^*(1 - x^*)]} \\[4mm] v = \dfrac{5[2y^* + x^* + 5x^*y^*(2 - x^* - y^*)]}{3 + 5[y^*(1 - y^*) + x^*(1 - x^*)]} \end{cases} \tag{16}$$
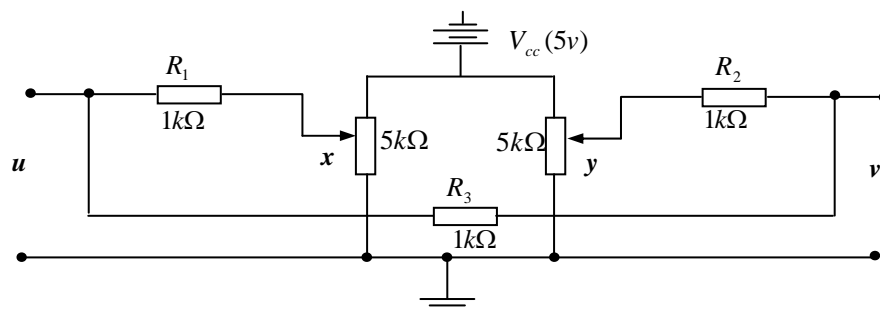

Figure 2. Circuit Model of Two Input/Output Sensor

Where $x^* = x + \zeta_1$ and $y^* = y + \zeta_2$. As a training set, we use 144 samples (x_i^*, y_i^*, u_i, v_i) generated by the above function. Here, the training input set (x_i, y_i) is a cartesian product of two input signal sets, which are both composed of 12 equally spaced data points over the interval (0.1, 0.9), and $\zeta_{1,2}$ are iid additive noise. The risk, generalization error, is computed with respect to the Equation (16) without noise, thus the test set consists of 196 samples (x_i, y_i, u_i, v_i) generated from the noiseless Equation (16), where the test input set is also a cartesian product that are composed of 14 equally spaced data points in the interval (0.1, 0.9).

In this experiment, we add Gaussian noise with zero mean and standard deviation to the data, which is the common assumption, and the aim is to observe whether the proposed

method with theoretically predicted value of *v* can lead to good generalization performance in practical for different noise level, and whether the noise level has any influence on PSO based hyperparameters selection procedure. Therefore, we first compute the optimal hyperparameters for varied noise level (different standard deviation), here the fitness is the root mean squared relative error (RMSRE) between the estimations and the true values of noiseless test input signals, and then we calculate the relative error of estimations with the optimal parameters for each noise level.

Figure 3 and Figure 4 illustrate the performances of PSO approaches on the test set with different noise setting to find the global optimum by plotting the best fitness versus the number of iterations. It is clear from the figures that for all noise levels the value of RMSRE decreases very fast to the high quality solutions at the early iterations (about 100 iterations) and then the curves become very flat, which implies that PSO can converge to the global optimum very quickly. It also can be seen that the best RMSRE of signals increase with the increase of the noise level, however the biggest value is still less than 1%, thus it demonstrates that the proposed PSO procedure can effectively prevent the premature convergence and significantly enhance the convergence rate and accuracy in the evolutionary process, independent of the noise levels.
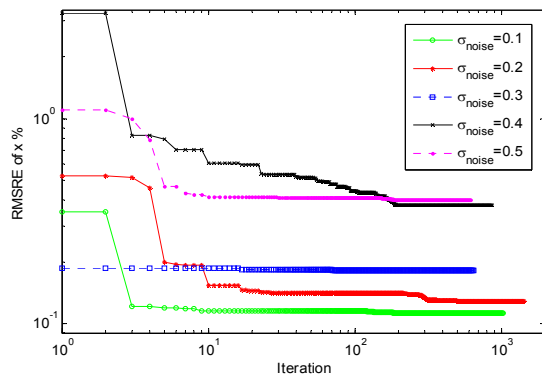

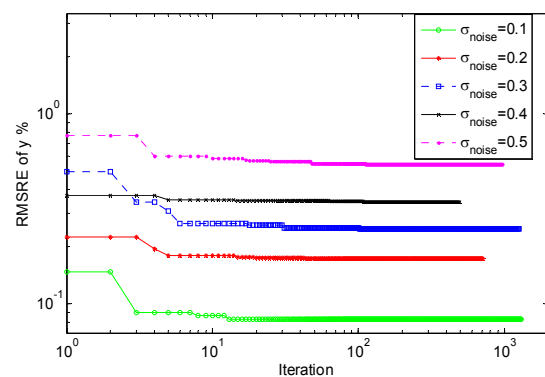Figure 3. RMSRE of x versus Iteration for Different Noise Level

Figure 4. RMSRE of y versus Iteration for Different Noise Level
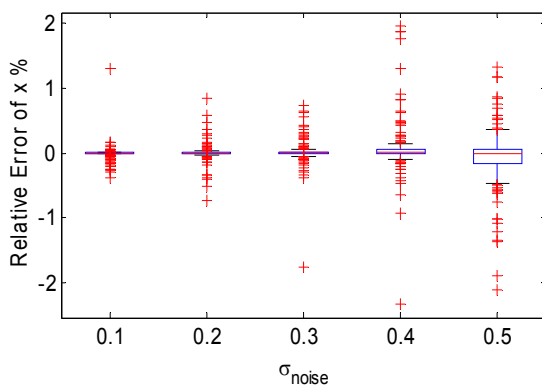

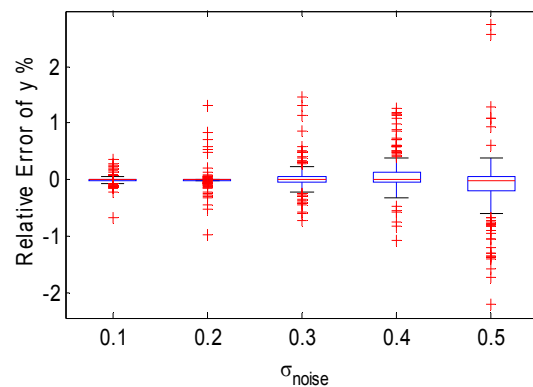Figure 5. Relative Error of x for Different Noise Level

Figure 6. Relative Error of y for Different Noise Level

To evaluate the reconstruction performance of *v*-SVR algorithm with optimal parameters setting, the box plots of the relative error of *x* and *y* for different noise levels are shown in Figure 5 and Figure 6. Each box plot is based on the results of test data set with varying added noise: from left to right, $\sigma_{noise} \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. As shown in figures, the relative error of both reconstructed signals *x* and *y* are approximately zero-mean or very close to zero-mean for all noise levels. This is confirmed by the mean values, given at the bottom of each plot, which

implies that the proposed algorithm is unbiased. Note from the figures that the range and standard deviation of the relative error increase with the growing value of noise level, however, the most of relative error for each box plot is within a reasonable range, such as all of the whisker range are within the range from -0.5% to 0.5%. Therefore, it proves that the proposed method is stable and accurate with respect to the added noise in the data set.

4. Conclusion

This paper presents an optimal SVR algorithm that profit from the combination of v-SVR and PSO for multifunctional sensor signal reconstruction. The v-SVR method is able to cope with a high-dimensional signal processing in small sample size situations. Moreover, the higher generalization accuracy could be achieved, since the parameter v is in accordance with the noise model that is in the data set. And the PSO based parameters optimization procedure is simple, efficient, and easy to implement. The experiment results suggest that the proposed method is suitable for the multivariable condition and enhance the generalization performance and stability of signal reconstruction under different noise levels. Hence, the proposed approach can be immediately used by practitioners interested in applying SVM to various application domains.

References

[1] Yuji J, Shida K. New Multifunctional Tactile Sensing Technique by Selective Data Processing. *IEEE Transactions on Instrumentation and Measurement*. 2000; 49(5): 1091-1094.
[2] Sun JW, Shida K. Multilayer Sensing and Aggregation Approach to Environmental Perception with One Multifunctional Sensor. *IEEE Sensors Journal*. 2002; 2(2): 62-72.
[3] Wei G, Shida K. Estimation of Concentrations of Ternary Solution with NaCl and Sucrose based on Multifunctional Sensing Technique. *IEEE Transactions on Instrumentation and Measurement*. 2006; 55(2): 675-681.
[4] Sun JW, Liu X, Sun SH. TLS Algorithm-based Study on Multifunctional Sensor Data Reconstruction. *Acta Electronica Sinica*. 2004; 32(3): 391-394.
[5] Wang X, Zhang XP. Optimal Look-up Table-based Data Hiding. *IET Signal Processing*. 2011; 5(2): 171-179.
[6] Liu X, Sun JW, Liu D. Nonlinear Multifunctional Sensor Signal Reconstruction based on Total Least Squares. *Journal of Physics: Conference Series*. 2006; 48(1): 281-286.
[7] Bot RI, Lorenz N. Optimization Problems in Statistical Learning: Duality and Optimality Conditions. *European Journal of Operational Research*. 2011; 213(2): 395-404.
[8] Vapnik V. The Nature of Statistical Learning Theory. 2nd edition, New York: Springer. 2000: 147-150.
[9] Yu YL Zhou. Acoustic Emission Signal Classification Based On Support Vector Machine. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2012; 10(5): 1027-1032.
[10] Charrada A. Support Vector Machines Regression for MIMOOFDM Channel Estimation. *IAES International Journal of Artificial Intelligence*. 2012; 1(4): 214-224.
[11] Bo Y, L Liang, W Xue. Color Calibration Model in Imaging Device Control using Support Vector Regression. *TELKOMNIKA Indonesian Journal of Electrical Engineering. 2013*; 11(10): 5530-5538.
[12] Scholkopf B, Smola AJ, Williamson RC, Bartlett PL. New Support Vector Algorithms. *Neural Computation*. 2000; 12(5): 1207-1245.
[13] Smola A, Scholkopf B. A Tutorial on Support Vector Regression. *Statistics and Computing*. 2004; 14(3): 199-222.
[14] Chalimourda A, Scholkopf B, Smola A. Experimentally Optimal *v* in Support Vector Regression for Different Noise Models and Parameter Settings. *Neural Networks*. 2004; 17(1): 127-141.
[15] Cherkassky V, Ma Y. Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*. 2004; 17(1): 113-126.
[16] Parpinelli RS, Teodoro FR, Lopes HS. A Comparison of Swarm Intelligence Algorithms for Structural Engineering Optimization. *International Journal for Numerical Methods in Engineering*. 2012; 91(6): 666-684.