Chinese Grain Production Forecasting Method Based on Particle Swarm Optimization-based Support Vector Machine

Sheng-Wei Fei*, Yu-Bin Miao and Cheng-Liang Liu

School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, P.R. China

Received: September 8, 2008; Accepted: October 27, 2008; Revised: October 27, 2008

Abstract: Forecasting of grain production is an important resource for establishing agriculture policy. Particle swarm optimization-based support vector machine (PSO-SVM) is applied to forecast grain production in this paper. In PSO-SVM model, particle swarm optimization (PSO) is used to determine free parameters of support vector machine. PSO is a new optimization method, which is motivated by social behavior of bird flocking or fish schooling. The optimization method not only has strong global search capability, but also is very easy to implement. The Chinese grain production is used to illustrate the performance of proposed PSO-SVM model. The experimental results indicate that the PSO-SVM method can achieve greater forecasting accuracy than grey model, artificial neural network in Chinese grain production forecasting. Consequently, PSO-SVM is a proper method in Chinese grain production forecasting.

Keywords: Grain production forecasting, support vector machine, particle swarm optimization, time series forecasting, parameters optimization.

1. INTRODUCTION

Forecasting of grain production is an important resource for establishing agriculture policy [1]. The accurate forecasting of grain production avails to the reasonable adjustment of agricultural structure and healthy development of agricultural economy. At present, the international popular grain production forecasting methods are meteorologic analysis method [2,3], remote sensing technology [4,5] and statistical dynamics growth simulation method [6]. The forecasting lead time of this three methods is only 1-2 months and their forecasting error of grain production is usually 5-10%, which can effect on the establishment of agriculture policy and reasonable adjustment of agricultural structure. Thus, time series prediction method is introduced to forecasting of grain production by some experts [7]. In the method, grain production is forecasted according to the production in former years and future change of grain production is forecasted in the years ahead.

In recent years, various time series prediction techniques have been proposed, including artificial neural network [8], grey model [9], support vector machine [10], etc. Artificial neural network (ANN) is one of the commonest methods used in non-linear forecasting, which has strong parallel processing and fault tolerant ability. However, the practicability of ANN is limited due to several weaknesses, such as requirement for a large amount of training data, ‘over-fitting’, slow convergence velocity and relapsing into local extremum easily [11]. Superior forecasting accuracy can be gained with a small quantity of training data by using grey model. However, grey model only depicts a monotonously increasing or decreasing process with time as exponential law, and the change of gain production usually takes on the fluctuation state. So a certain error is always generated in forecasting gain production based on GM. Support vector machine (SVM) is a new machine learning method based on the statistical learning theory, which solves the problem of ‘over-fitting’, local optimal solution and low convergence rate existed in ANN and has excellent generalization ability in the situation of small sample [12]. However, the practicability of SVM is effected due to the difficulty of selecting appropriate SVM parameters [13]. Particle swarm optimization (PSO) motivated by social behavior of bird flocking or fish schooling is a new optimization technology [14]. The optimization method is very easy to implement and there are few parameters to adjust. It has been successfully applied to solve multidimensional optimization problem in artificial neural network training, function optimization, etc [15, 16]. Thus, in the study, the proposed PSO-SVM model is applied to forecast Chinese grain production, among which PSO is used to determine free parameters of support vector machine.

This paper is organized as follows: Section 2 introduces particle swarm optimization-based support vector machine. Section 3 testifies the performance of the proposed PSO-SVM model with Chinese grain production data. Finally, current & future developments is provided in Section 4.

2. PARTICLE SWARM OPTIMIZATION-BASED SUPPORT VECTOR MACHINE

2.1. Support Vector Regression

The basic concept of SVM regression is to map nonlinearly the original data x into a high- dimensional feature space, and to solve a linear regression problem in this feature space. The regression approximation estimates a function according to a given data set $T=\{(x_i,y_i)\}_i^l$, where x_i denotes the input vector, y_i denotes the corresponding output value and l denotes the total number of data patterns, the SVM regression function is :

*Address correspondence to this author at the School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, P.R. China; Tel: +862134206075; E-mail: feishengwei@sohu.com

1872-2121/09 $100.00+.00

$$f(x) = w \cdot \varphi(x) + b \tag{1}$$

Where $\varphi(x)$ denotes the high-dimensional feature space, w denotes the weight vector and b denotes the bias term.

The coefficients w and b are estimated by minimizing the following regularized risk function:

$$R(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(y_i, f(x_i)) \tag{2}$$

Where C denotes a cost function measuring the empirical risk. $\frac{1}{2}\|w\|^2$ is the regularization term. $L_\varepsilon(y_i, f(x_i))$ is called the ε-insensitive loss function, which is defined as:

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon & |y_i - f(x_i)| \geq \varepsilon \\ 0 & |y_i - f(x_i)| < \varepsilon \end{cases} \tag{3}$$

In Eq. (3), the loss equals zero if the error of forecasting value is less than ε, otherwise the loss equals value beyond ε.

Two positive slack variables ξ and ξ^* are introduced to represent the distance from actual values to the corresponding boundary values of the ε-tube. Then, Eq. (2) is transformed into the following constrained form:

Minimize

$$\phi(w, \xi, \xi^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}(\xi + \xi^*) \tag{4}$$

Subject to

$$\begin{cases} y_i - [w \cdot \varphi(x)] - b \leq \varepsilon + \xi_i & \xi_i \geq 0 \\ [w \cdot \varphi(x)] + b - y_i \leq \varepsilon + \xi_i^* & \xi_i^* \geq 0 \end{cases}$$

This constrained optimization problem is solved using the following Lagrangian form:

Maximize

$$H(\partial, \partial^*) = -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\partial_i - \partial_i^*)(\partial_j - \partial_j^*)K(x_i, x_j) \tag{5}$$

$$+\sum_{i=1}^{l} y_i(\partial_i - \partial_i^*) - \varepsilon\sum_{i=1}^{l}(\partial_i + \partial_i^*)$$

Subject to

$$\sum_{i=1}^{l}(\partial_i - \partial_i^*) = 0 \quad \partial_i, \partial_i^* \in [0, C]$$

Where ∂_i and ∂_i^* are the so-called Lagrangian multipliers. $K(x_i, x_j)$ is called the kernel function. The value of the kernel function equals the inner product of $\varphi(x_i)$ and $\varphi(x_j)$ which are produced by mapping two vectors x_i and x_j into the higher dimensional feature space, that is $K(x_i, x_j) = \varphi(x_i)\varphi(x_j)$.

By the Lagrange multipliers ∂_i and ∂_i^* calculated, an optimal desired weight vector of the regression hyperplane is obtained, that is:

$$w^* = \sum_{i=1}^{l}(\partial_i - \partial_i^*)K(x_i, x) \tag{6}$$

Hence, the regression function is:

$$f(x) = \sum_{i=1}^{l}(\partial_i - \partial_i^*)K(x_i, x) + b \tag{7}$$

In SVM, radial basis function(RBF) $K(x_i, x) = \exp(-\|x_i - x\| / 2\sigma^2)$, polynomial basis function $K(x_i, x) = ((x_i \cdot x) + b)^d$ ($b \geq 0$, d is the natural number), and sigmoid function $K(x_i, x) = \tanh(k(x_i \cdot x) + v)$ ($k > 0$, $v < 0$) are commonly used. Only one variable needs determination in radial basis function, thus fewer free parameters in SVM need determination using radial basis function, which avails to optimize parameters. In addition, SVM constructed by radial basis function has excellent nonlinear forecasting performance. Thus, in this work, RBF is used in the SVM. Here, C, σ and ε are user-determined parameters.

2.2. The Principle of PSO

PSO is a populated search method, which derives from the research for the movement of organisms in a bird flocking or fish schooling. Similar to genetic algorithms, PSO performs searches using a population (called swarm) of individuals (called particles) that are updated from iteration to iteration [17]. To discover the optimal solution, each particle moves in the direction of its previously best position (pbest) and its best global position (gbest). The velocity and position of particles can be updated by the following equations:

$$\begin{aligned} v_{ij}(t+1) = {} & w \cdot v_{ij}(t) + c_1 \cdot rand_1 \cdot (pbest_{ij}(t) - p_{ij}(t)) \\ & + c_2 \cdot rand_2 \cdot (gbest_{ij}(t) - p_{ij}(t)) \end{aligned} \tag{8}$$

$$p_{ij}(t+1) = p_{ij}(t) + \beta \cdot v_{ij}(t+1) \tag{9}$$

Where t denotes the iteration counter, v_{ij} is the velocity of particle i on the j th dimension , whose value is limited to the range[$-v_{\max}, v_{\max}$], p_{ij} is the position of particle i on the j th dimension, whose value is limited to the range[$-p_{\max}, p_{\max}$], $pbest_{ij}$ is the pbest position of particle i on the j th dimension , and $gbest_{ij}$ is the gbest position of the swarm on the j th dimension. The inertia weight w is used to balance the global exploration and local exploitation. The $rand_1$ and $rand_2$ are random function in the range [0, 1], β is constraint factor used to control the velocity weight, whose value is usually set to 1. Positive constant c_1 and c_2 are personal and social learning factors, whose values are usually set to 2.

2.3. Parameters Optization of SVM Based on PSO

In SVM, the election of the parameters, C, σ and ε has a great influence on the performance of SVM, so the particle is composed of three parts, C, σ and ε. Fig. (**1**) presents the process of optimizing the SVM parameters with PSO, which is described below:

(1) Initialization. PSO is initialized with a population of random particles and velocities.

(2) Training SVM model and fitness evaluation. SVM model is trained with the parameters, C, σ and ε included in current particle. The k-fold cross validation is used to evaluate fitness. In k-fold cross validation, the training data set is randomly divided into k mutually exclusive subsets of approximately equal size. Among which $k-1$ subsets are used as the training set, the last subset is used as validation. The above procedure is repeated k times, so that each subset is used once for validation. The fitness function is defined as the $\text{MAPE}_{\text{cross validation}}$, which is shown in Eq. (10) and Eq. (11).The solution with a smaller $\text{MAPE}_{\text{cross validation}}$ has a smaller fitness value.

$$\text{Fitness}=\text{MAPE}_{\text{cross validation}}=\frac{1}{k}\sum_{i=1}^{k}\left|\bar{e}_i\right|\times 100\% \tag{10}$$

$$\bar{e}_i=\frac{1}{m}\sum_{j=1}^{m}\left|\frac{y_j-\hat{y}_j}{y_j}\right| \tag{11}$$

Where y_j is the actual value and $\hat{y}_j$ is the validation value; m is the number of samples in validation subset; $\bar{e}_i$ is mean relative error of the validation subset.

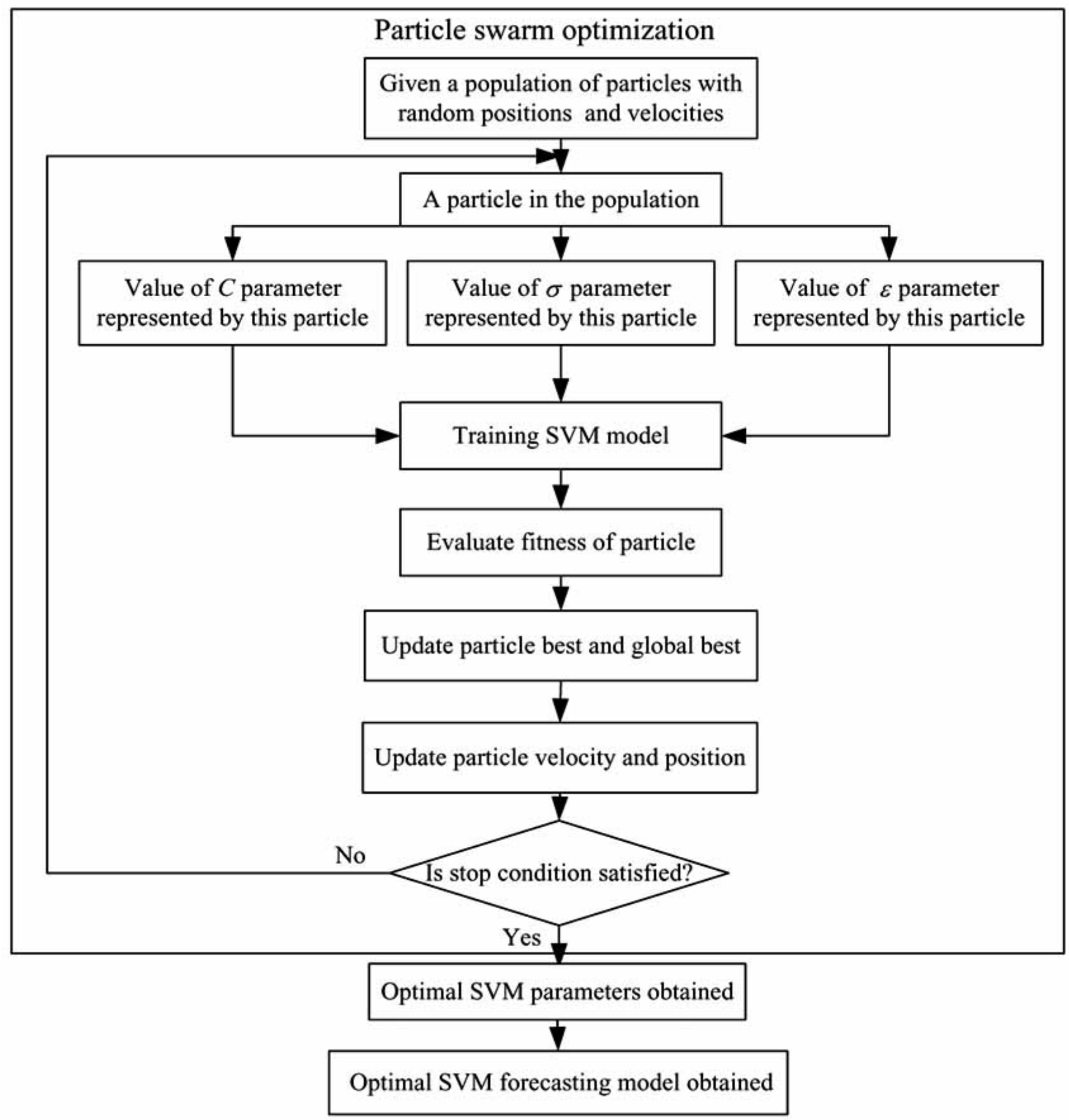

Fig. (1). The process of optimizing the SVM parameters with particle swarm optimization.

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

(4) Update the velocity and position value of each particle. The velocity of each particle is calculated by Eq. (8). Each particle moves to its next position according to Eq. (9).

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

3. FORECASTING FOR CHINESE GRAIN PRODUCTION BASED ON PSO-SVM

Forecasting of grain production is the time series forecasting problem. And the goal is to search a forecasting model with excellent generalization ability by utilizing the training sample obtained by historical data. The SVM regression mapping function can be described as below:

$$a_t = f(a_{t-1}, a_{t-2}, \ldots, a_{t-m}) \tag{12}$$

Where a_t is the output value of time t, $(a_{t-1}, a_{t-2}, \ldots, a_{t-m})$ is the input vector, m is the dimension of the input vector.

Chinese gain production data is shown in Fig. (**2**), among which the data from 1985 to 2000 is used as training data and the data from 2001 to 2005 is used as testing data. Before training sample sets are constructed, the primary data should be normalizated, which can improve operation speed and generalization capability of forecasting model. The training data is used to construct training sample sets according to the dimension of the input vector. Here, the dimension of the input vector is set to 4. In the training stage, firstly the free parameters C, σ and ε of SVM are optimized by PSO, the validation error is measured by Eq. (10), the adjusted parameters with minimum validation error are selected as the most appropriate parameters. Then, the optimal parameters are utilized to train SVM model. The testing data sets are used to examine the accuracy of the forecasting model. A mean absolute percentage error (MAPE) is used to evaluate the forecasting accuracy, which is as follows:

$$\text{MAPE} = \frac{1}{N}\sum_{h=1}^{N}\left|\frac{y_h - \hat{y}_h}{y_h}\right| \times 100\% \tag{13}$$

Where y_h and $\hat{y}_h$ represent the actual and forecasting values respectively, N is the number of forecasting points.

The suitable parameters for the PSO-SVM model of grain production forecasting are illustrated in Table **1**. The ANN compared with PSO-SVM is one hidden-layer BP neural network with 4 input nodes, 10 hidden nodes and 1 output nodes. Comparison of the forecasting results among PSO-SVM, BP and GM are shown in Fig. (**3**) and Table **2**.

In Fig. (**3**), the forecasting values of GM take on the monotonously increasing state, when actual values are descending, the forecasting values of GM can't be adjusted and go on presenting ascending trend. Then, biggish error is generated. Thus, it can be seen that GM isn't suitable for using under the circumstances that the change of data takes on the convex or concave state. However, the change of gain production usually takes on the fluctuation state, which is shown as Fig. (**3**). PSO-SVM implements the principle of

Table 1. The Suitable Parameters for the PSO-SVM Model of Grain Production Forecasting

PSO-SVM parameters			Training MAPE/%
C	σ	ε	
15.365	0.85162	0.011252	2.9928

Table 2. The Comparison of the Forecasting Results Among GM, BP, PSO-SVM

Year	Actual Value/10^4 t	GM		BP		PSO-SVM	
		Forecasting Value/10^4 t	Error/%	Forecasting Value/10^4 t	Error/%	Forecasting Value/10^4 t	Error/%
2001	45264	51983	14.852	47092	4.0436	45695	0.9573
2002	45706	52936	15.814	46971	2.7644	44848	-1.8797
2003	43070	53904	25.160	45836	6.4240	43407	0.7842
2004	46947	54893	16.924	51117	8.8830	44860	-4.447
2005	48401	55897	15.486	51176	5.7310	45773	-5.431
MAPE/%		17.647		5.5692		2.6998	

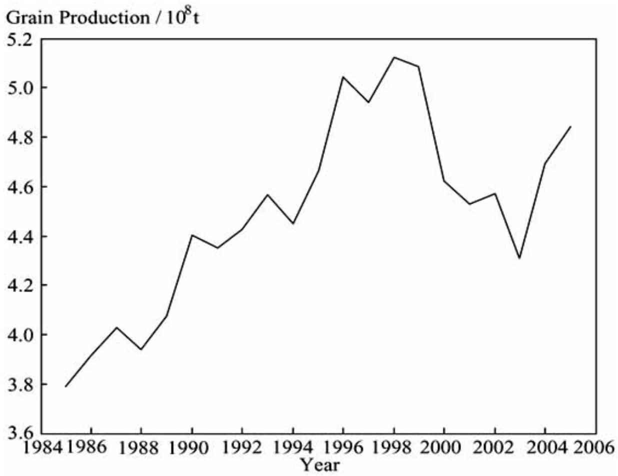

Fig. (2). Chinese gain production data.

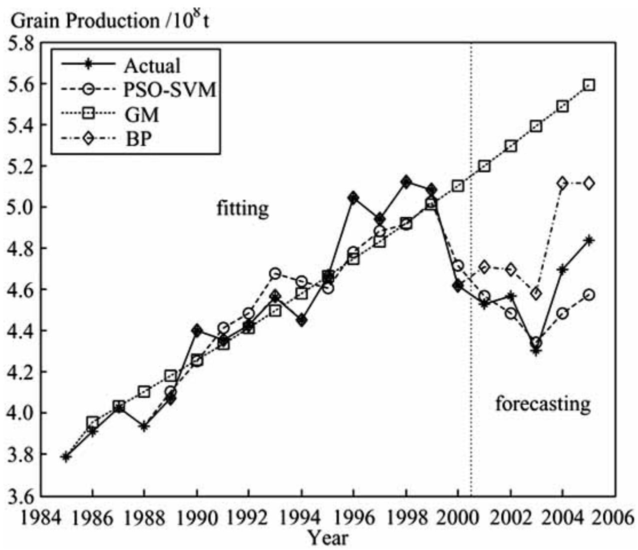

Fig. (3). Forecasting values of GM, BP, PSO-SVM.

structural risk minimization in place of experiential risk minimization, which makes it have excellent generalization ability in the situation of small sample, it has no requirement in data series distribution, and can be in the wake of changing disciplinarian of data. Thus, the forecasting error of PSO-SVM is also small even if under the circumstances that the change of data takes on great fluctuation. ANN which implements the principle of experiential risk minimization not only needs a large amount of training data, but also the stabilization of its forecasting results is very bad. It indicates that PSO-SVM has more excellent performance than ANN, GM in forecasting Chinese gain production by comparison of their forecasting results.

4. CURRENT & FUTURE DEVELOPMENTS

In this paper, PSO-SVM is applied to forecast Chinese grain production. In the PSO-SVM approach, PSO is used to select suitable parameters of SVM, which avoids over-fitting or under-fitting of the SVM model occurring because of the improper determining of these parameters. PSO is a new optimization method, which not only has strong global search capability, but also is very easy to implement. So it is very suitable for parameters selection of SVM. The Chinese grain production is used to illustrate the performance of proposed PSO-SVM model. The experimental results indicate that the PSO-SVM method can achieve greater forecasting accuracy than grey model, artificial neural network in Chinese grain production forecasting. This study demonstrates that PSO-SVM is a proper method in Chinese grain production forecasting. Future research focuses on the development of grain production forecasting system.

CONFLICT OF INTEREST

The author has no conflict of interest to declare.

REFERENCES

[1] Traill WB, Arnoult MHP, Chambers SA, *et al*. The potential for competitive and healthy food chains of benefit to the countryside. Trends Food Sci Technol 2008; 19(5), 248-254.
[2] Kobayashi, T., Hasegawa, T., Hua, D., Uchida, K.: US20077227625 **(2007)**.
[3] Wakayama, T.: US20006097329 **(2000)**.
[4] Payne, F.L., Molloy, J.J.: US4169243 **(1979)**.
[5] Whelan, T.W., Kashani, A.H.M.: US20006067044 **(2000)**.
[6] Ding, Y., Newman, K.W.: US20067076397 **(2006)**.
[7] Huifei J, Deyong W, Shuhua L, Shuyan L, Wenliang W. Prediction of grain yield using chaos theory. J China Agric Univ 2006; 11 (1), 47-52(in Chinese).
[8] Tseng CM, Jan CD, Wang JS, Wang CM.Application of artificial neural networks in typhoon surge forecasting. Ocean Eng 2007; 34(11-12), 1757-1768.
[9] Wu AX, Xi Y, Yang BH, Chen XS, Jiang HC. Study on grey forecasting model of copper extraction rate with bioleaching of primary sulfide ore. Acta Metallurgica Sin (English Letters) 2007; 20(2), 117-128.
[10] Yongsheng D, Xinping S, Yueming Z. Forecasting financial condition of Chinese listed companies based on support vector machine. Expert Syst Applic 2008; 34(4): 3081-3089.
[11] Rosario C, Renato De L, Emilia M. Integrating support vector machines and neural networks. Neural Networks 2007; 20(5): 590-597.
[12] Vapnik V. The Nature of Statistical Learning Theory. Springer-Verlag, New York 1995.
[13] Kristof Coussement, Dirk Van den Poel.Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. Expert Syst Applic 2008; 34(1), 313-327.
[14] Liu B, Wang L, Jin Y-H. An effective hybrid PSO-based algorithm for flow shop scheduling with limited buffers. Comput Oper Res 2008; 35(9): 2791-2806.
[15] Shen Q, Shi WM, Yang XP, Ye BX. Particle swarm algorithm trained neural network for QSAR studies of inhibitors of platelet-derived growth factor receptor phosphorylation. Eur J Pharm Sci 2006; 28(5), 369-376.
[16] Brits R, Engelbrecht AP, van den Bergh F.Locating multiple optima using particle swarm optimization. Appl Math Comput 2007; 189(2): 1859-1883.
[17] Sha DY, Hsu C-Y. A new particle swarm optimization for the open shop scheduling problem. Comput Oper Res 2008; 35(10), 3243-3261.