http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (18): 4666-4672, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4666.4672

Study on Nonlinear Combination Forecasting Model for Grain Yield

[1]Liu Ran and [2]Bu Hui
[1]School of Software, North China University of Water Resources and Electric Power, 450011, Zhengzhou, China
[2]Department of Information Engineering, North China University of Water Resources and Electric Power, 450011, Zhengzhou, China

Abstract: In view of the existing problems in grain yield, this study introduces a nonlinear combination forecasting model based on adaptive neural network, support vector machine and relevance vector machine. In this combination model, we take the prediction result of the previous single model as the input data of the next single model. In addition, a new method of determining the weight coefficient is proposed based on rough set theory. During the forecast operation, the core modules of the three models are organized together organically which improves the precision and stability of the combination model. Test results show that this method overcomes random and mutations of traditional methods and the mean absolute error of the prediction result is lower than the traditional model.

Key words: Self-adaptive neural network, combination forecasting model, SVM, RVM, rough set

INTRODUCTION

In today's society agriculture has become the foundation of the national economy. With the accurate forecasts of grain output, we can understand the supply and demand gap of future grain and can adopt appropriate measures to resolve the problems timely which will play an important role on social stability. Current prediction methods of grain yield include regression prediction, rough set theory, gray prediction, BP neural network and so on. Due to the restrictions of normality, stability, independence of time series, the traditional modeling methods are not suitable for the forecast of grain output time series. With the fast development of artificial intelligence technology and information technology, many new forecast models such as BP neural network, Wavelet analysis, SVM, RVM and so on are proposed. However, the traditional neural network is easy to fall into local minimum, as well as the topology structure is difficult to define theoretically. The prediction modes based on wavelet transform can get good results only when the input data is stationary time series. SVM method can not obtain the uncertainty in forecasting because of its lack of necessary probability information. In grey forecasting model, it is difficult to find appropriate membership functions which cause the lower reliability of predictions. RVM can get good predictions but to the large sample, it is not the best method. Furthermore, RVM method relies on excessive artificial parameters which limit its Successful application. Forecasting method based on rough set need lots of rules but prediction accuracy rate is not high (Bu and Liu, 2011).

Therefore, different forecasting methods provide different useful information and prediction accuracy and the focus are often different. In order to utilize the advantages of single method fully and make up for the shortcomings of single prediction method and achieve better prediction results, two or more prediction models can be integrated by some nonlinear mechanisms to construct combination forecasting models. In this study, we introduce a nonlinear combination forecasting model based on adaptive neural network, support vector machine and relevance vector machine (Tipping, 2001). In addition, a new method of determining the weight coefficient is proposed based on rough set theory. Test results show that this method overcomes random and mutations of traditional methods and the mean absolute error of the prediction result is lower than the traditional model.

WEIGHT COEFFICIENT OF THE COMBINATION FORECASTING MODEL

Suppose there are n prediction models we can use to predict the same data sample. Then, the combination forecasting model can be viewed as follows:

Corresponding Author: Liu Ran, School of Software, North China University of Water Resources and Electric Power, 450011, Zhengzhou, China

$$\hat{y}_t = \sum_{i=1}^{n} k_i \cdot \hat{y}_t^{(i)}$$

$\hat{y}_t$ is the prediction value of the combination forecasting model at the moment t.

$\hat{y}_t^{(i)}$ is the prediction value of the prediction model i at the moment t (i = 1, 2, .., n).

k_i is the weight coefficient of prediction model i at the moment t (i = 1,2,..., n):

$$\sum_{i=1}^{n} k_i = 1, k_i \geq 0$$

In the following, let fitted values of each single forecasting model in the combination model denote conditional property C and C = ($\hat{y}^{(1)}$, $\hat{y}^{(2)}$,..., $\hat{y}^{(n)}$). Let observed value of the forecast object denote decision attribute D and D = {y}. Discourse domain U = {u_1, u_2, …, u_n} and u_t = ($\hat{y}^{(1)}$, $\hat{y}^{(2)}$,..., $\hat{y}^{(n)}$, y_t), t = 1, 2, ..., n, in which $\hat{y}^{(1)}$, $\hat{y}^{(2)}$,..., $\hat{y}^{(n)}$ and y_t denote fitted value and historical data of forecast object of each single forecasting model respectively at moment t (Yuan and An, 2013).

For analyzing the importance degree of each single based on rough set theory, the foundation is to discretize the conditional property and set up knowledge representation system (Zhao *et al*., 2004). In this study, the calculation method of weight coefficient for single model is as follows:

- Compute the dependence degree of decision attribute D upon conditional property C:

$$k = \gamma_c(D) = \sum_{i=1}^{n} |POS_c(y_i)| \Big/ |U|$$

In which: C = {$\hat{y}^{(1)}$, $\hat{y}^{(2)}$,..., $\hat{y}^{(n)}$}, D = {y}, $|POS_c(y_i)|$ denotes positive region of decision attribute D about condition attributes *C*, |U| is the cardinal number of set U

- Remove the prediction model i and compute the dependence degree of decision attribute D upon conditional property $C - \{c_i\}$:

$$\gamma_{C-\{c_i\}}(D) = \sum_{i=1}^{n} |POS_{C-\{c_i\}}(y_i)| \Big/ |U|, i = 1,2,...,n$$

- Compute the importance degree of the prediction model i among all the prediction models:

$$\sigma_{CD}(c_i) = \gamma_c(D) - \gamma_{C-\{c_i\}}(D), i = 1,2,...,n$$

- Compute the weight coefficient of the prediction model i:

$$k_i = \frac{\sigma_{CD}(c_i)}{\sum_{i=1}^{n} \sigma_{CD}(c_i)}, i = 1, 2, ..., n$$

STRUCTURAL ANALYSIS OF NONLINEAR COMBINATION FORECASTING MODEL

Based on the analysis above, we adopt the single prediction models of adaptive neural network, support vector machine and relevance vector machine to construct the combination forecasting model. In the following sections, we firstly give the basic working principle of each single model and then analyse the frame construction of the combination forecasting model.

Adaptive neural network prediction model: BP network is a neural network with three layers or more neurons, including input layer, middle layer (hidden layer) and out layer. In this network model, the adjacent two layers are connected completely but no connections exist among the neurons in the same layer (as shown in Fig. 1). Each connection has weight parameters. BP neural network model realizes the forecast function by training sample data. In the process of learning and training, the BP model generates firstly a set of connected weights between the neurons randomly and gets the output by forward propagation. Then, comparing the output with the expected value, if the error is larger than the given value, we need update the connected weights to reduce the errors by back-propagation process. The output computing of forward propagation and the weights updating of back-propagation are executed alternately until the error between the network's actual output and expectations meets the requirements. In this way we can get the satisfactory connected weights and the thresholds (network structure). Then, the prediction process can be carried out by inputting the tester sample into the network structure (Mukherjee *et al*., 1997).

In practical applications, it is found that local minimization could lead to failure of the learning process and that the speed of convergence is very slow in the BP

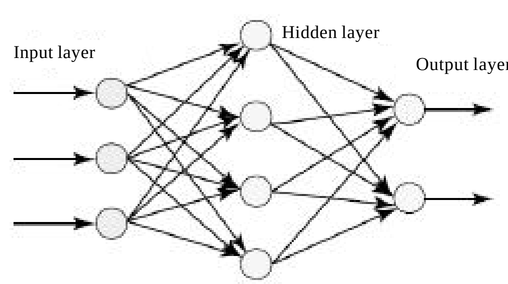

Fig. 1: Basic artificial neural network structure

algorithm. So, we adopt additional momemtum and self-adaptive learning rate to overcome these problems (Wu *et al*., 2008).

In the additional momemtum method, a momemtum term which is proportional to the previous variation of weight (or bias) is added to the previous variation of weight and then a new variation of weight is produced on the basis of error back propagation. The adjust equation of weight and bias are as follows:

$$\Delta\omega_{ij}(k+1) = (1-mc)\eta\delta_i P_j + mc\Delta\omega_{ij}(k)$$
$$\Delta b_j(k+1) = (1-mc)\eta\delta_i + mc\Delta b_j(k)$$

in which, $\Delta\omega_{ij}$ is the variation of weight, Δb_j is the variation of bias, δ is the error term, P_j is input variable, η is learning rate, k is train times, mc is the momemtum term (about 0.95). The additional momemtum method is a method in which the influence of the variations of weight and bias are propagated by a momemtum term. When the momemtum term is zero, the variations weight and bias are generated on the basis of gradient descent method, when the momemtum term equals to one, the new variations of weight and bias exactly equals to the last variation and the part of variation generated on the basis of gradient method can be omitted. The additional momemtum term can promote the adjustation of weight and bias varying toward the direction of the bottom of error surface. When the values of weights of net are located in the flatness of the bottom of error surface, δ_i will be small, so $\Delta\omega_{ij}$ (k+1)≈ $\Delta\omega_{ij}$ (k). Then we can prevent from the occurrence of $\Delta\omega_{ij}$ (k) = 0, which will make net jumping out from the local minimum of error surface.

RVM forecasting model: Relevance vector machine is a sparse probability model based on support vector machine proposed by Michael E Tipping in 2001. Its training is carried on under Bayesian framework, so we can get the distribution of predicted values by regression estimatewith RVM (Wu and Wang, 2010).

The output of RVM model is as follows:

$$y(x) = \sum_{j=1}^{m} \omega_j \phi_j(x) + \omega_0$$

where, ϕ_j (x) is non-linear kernel function, ω_j is model weights. After defining the model basis functions, we can train the model weights ω_j with maximum likelihood method under Bayesian framework, which may avoid learning problems and improve model generalization ability. Therefore, RVM defines priori probability distribution for each model weight:

$$p(\omega_j \mid \alpha_j) = \left[\frac{\alpha_j}{2\pi}\right]^{1/2} \exp\left[-\frac{1}{2}\alpha_j\omega_j^2\right]$$

where, ω_j is hyper-parameter of the priori distribution of model weight α_j (Eyheramendy *et al*., 2003). For a given set of training samples $\{x_i, t_i\}_{i=1}^{N}$, Assume that the target value t_i is independent and the noise of input data obey Gaussian distribution of which the variance is σ^2. In this way, likelihood functionof the given training samples set is as follows:

$$p(t \mid \omega, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2}\|t - \Phi\omega\|^2\right]$$

where, $t = (t_1, t_2, \ldots, t_N,)^T$, $\omega = (\omega_1, \omega_2, \ldots, \omega_N)^T$, Φ is matrixof which the rows include the response of all kernel functions to input x_i:

$$(\Phi)_i = [1, \phi_1(x_i), \phi_2(x_i) \ldots, \phi_n(x_i)]$$

Based on priori probability distribution and likelihood distribution, calculate the posterior probability distribution of model weights with Bayesian method. The formula can be written as:

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

The posterior distribution of model weight is multivariate Gaussian distribution, that is:

$$p(\omega \mid t, \alpha, \sigma^2) = N(\mu, \Sigma)$$

where: $\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}$ is covariance, A is diagonal matrix of $(\alpha_0, \alpha_1, \ldots, \alpha_n)$ and $\mu = \sigma^{-2} \Sigma \Phi^T t$ is mean value. The likelihood distribution of training target value can realize marginalizationby integration:

$$p(\omega \mid t, \alpha, \sigma^2) = p\int (t \mid \omega, \sigma^2) p(\omega \mid \alpha) d\omega$$

In this way, we can get marginal likelihood distribution of the hyper-parameters:

$$p(t \mid \alpha, \sigma^2) = N(0, C)$$

Here, covariance $C = \sigma^2 I + \Phi A^{-1} \Phi^T$.

Finally, the estimated value of model weights in RVM method are given by the mean value of posterior distribution, as well as it is Maximum A Posteriori (MAP) estimation. The MAP estimation of model weight depends on hyper-parameters α and noise variance σ^2 and its

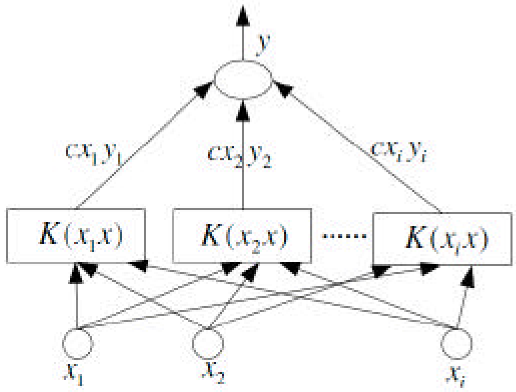

Fig. 2: Network model of SVM

estimated value $\bar{\alpha}$ and $\bar{\sigma}^2$ can be obtained by maximizing the marginal likelihood distribution. The uncertainty of model weight optimal value reflected by posterior distribution may shows the uncertainty of model predictions. For the given input value x* the corresponding probability distribution of the output is as follows:

$$\begin{aligned} &p(t^* \mid x^*, \bar{\alpha}, \bar{\sigma}^2) = \\ &p\int (t^* \mid x^*, \omega, \bar{\sigma}^2) p(\omega \mid t, \bar{\alpha}, \bar{\sigma}^2) d\omega \end{aligned}$$

The above formula obeys the form of Gaussian distribution, that is:

$$p(t^* \mid x^*, \bar{\alpha}, \bar{\sigma}^2) = N(y^*, \sigma^{*2})c$$

Where the predicted mean value $y^* = \mu^T \Phi(x^*)$ and the variance (uncertainty):

$$\sigma^{*2} = \bar{\sigma}^2 + \Phi^T(x^*) \sum \Phi(x^*)$$

RVM solves the problem of parameters selection with significance under Bayesian framework which has wide applicability. Using RVM for regression prediction, we can obtain better predicted value and its variance range.

SVM forecasting model: Support vector machine (also known as support vector network) is proposed by Vapnik which is a learning method based on statistical theory. The main idea of Support Vector Machine is to construct a nonlinear kernel function to map the data from the input space into a possibly high dimensional feature space and then generalize the optimal hyper-plane with maximum margin between the two classes. Similar to a neural network, the output of support vector machine is linear combination of intermediate nodes and each intermediate node corresponds to a vector, the support vector machine network model is shown in Fig. 2.

SVM possesses complete theory for it is based on statistical learning theory (Gartner and Flach, 2001). However, there are still some problems in actual application. A typical problem is the choice of model parameters. The parameters which have important influence on prediction accuracy are penalty factor C and kernel functions. Penalty factor C is used to control the compromise between model complexity and approximation error. At the same time different kinds of kernel functions will generate different number of support vectors. In order to improve the prediction accuracy we can use some optimization algorithms, for example, PSO to optimize the parameters selection of support vectors.

Structure of combination model: As mentioned above, neural network is easy to fall into local minimum, as well as the topology structure is difficult to define theoretically. SVM method cannot obtain the uncertainty in forecasting because of its lack of necessary probability information. At the same time, SVM kernel function must satisfy the mercer conditions which limit the range of choices. RVM can get good predictions but to the large sample, it is not the best method. Furthermore, RVM method relies on excessive artificial parameters which limit its successful application. Therefore, in order to overcome the problems above, we proposed a new combination prediction model which takes use of the advantages of SVM, neural network model and RVM. In this combination model, firstly, we use neural network to classify the initial sample data. The data with similar attributes are divided into a collection. In this procession the abnormal results are removed. Secondly, the processed data is transmitted to kernel function formed by the kernel functions of SVM and RVM. During the forecasting, the three models are organized together and control information can be passed to each model to adjust model parameter. That is to say, the three model work together dynamically and synchronously which enhances the stability and accuracy of forecast results. The algorithm flow chart is shown in Fig. 3.

APPLICATION EXAMPLE AND ANALYSIS

In order to test and verify the effectiveness of the combination model proposed in this study, we predict the grain output with the combination forecast model. As we all know, the importance of grain output analysis is not only reflected in the fitting of historical data but also on short term prediction. Accurate forecasts can help to regulate the market balance of drain supply and demand

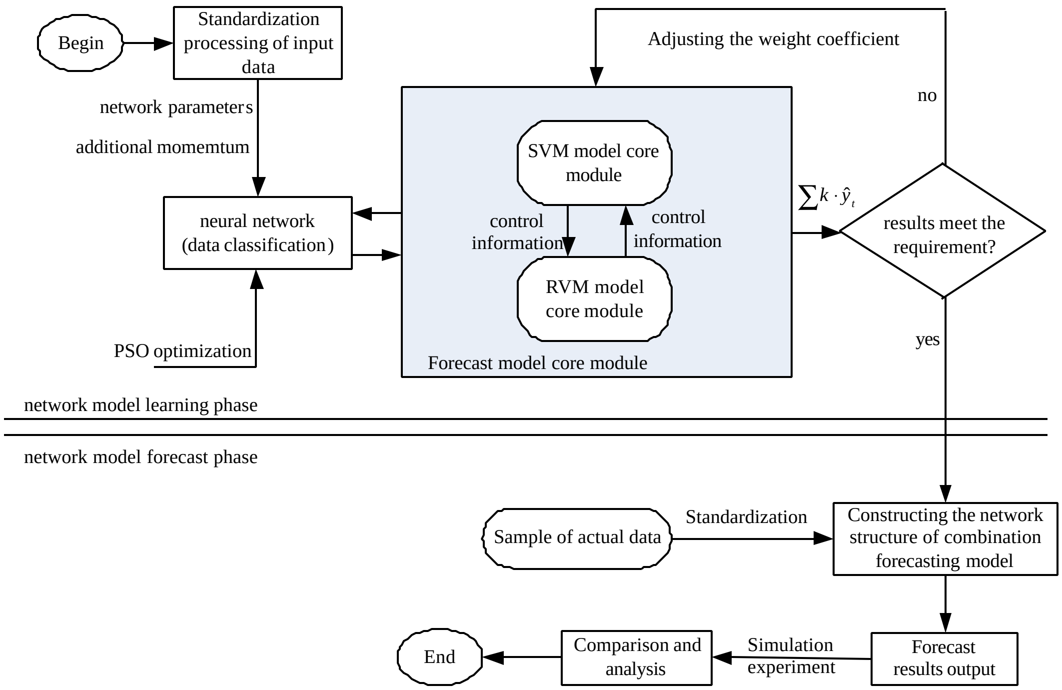

Fig. 3: Algorithm flow chart of combination model

Table 1: Grain yield from 1978-2009

Year	Grain yield	Year	Grain yield
1978	30476.5	1979	33121.2
1980	32055.5	1981	32150.2
1982	35145.0	1983	38172.8
1984	40173.1	1985	37910.8
1986	39115.1	1987	40129.8
1988	39140.8	1989	40175.5
1990	44624.3	1991	43529.3
1992	44265.8	1993	45648.8
1994	44510.1	1995	46661.8
1996	50453.5	1997	49417.1
1998	51229.5	1999	50838.5
2000	46217.5	2001	45263.6
2002	45705.7	2003	43069.5
2004	46946.9	2005	48402.1
2006	49804.2	2007	50160.2
2008	51000.3	2009	52650.8

and prevent insufficient or excessive grain supply because of adverse market effects. In this study, the experimental data we use is from grain output of some province from 1978 to 2009. The data is shown in Table 1.

We can divide the data into two parts: the data from 1978 to 2000 is used as training samples to construct network structure. The data from 2001-2009 is used as test samples to test generalization ability of the model.

In the selection of model parameters, RBF kernel function is chosen as kernel function of SVM model:

$$k(x, y) = \exp\left(-\frac{\|x-y\|^2}{\sigma^2}\right)$$

Penalty factor C and kernel parameter s are set as follows: (C = 69, s = 51). Neural network function module is set four layers and there are 16 neural nodes in every layer. In order to evaluate the performance of forecast models, we chose the RMSE and MAPE as the evaluation indicators.

The RMSE and MAPE are defined as follows:

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

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \bar{y}_i}{y_i}\right|$$

In which, y_i is the true value, $\bar{y}_i$ is forecast value and n is the number of the forecast sample. Table 2 shows the generalization value of various models from the 2001 to 2009.

Figure 4 shows the comparison curves of different forecasting models where combination model (1) denotes the model proposed by (Bu and Liu, 2011) and combination model (2) denotes the combination model proposed in this study.

Table 3 shows the comparison of RMSE and MAPE of various models.

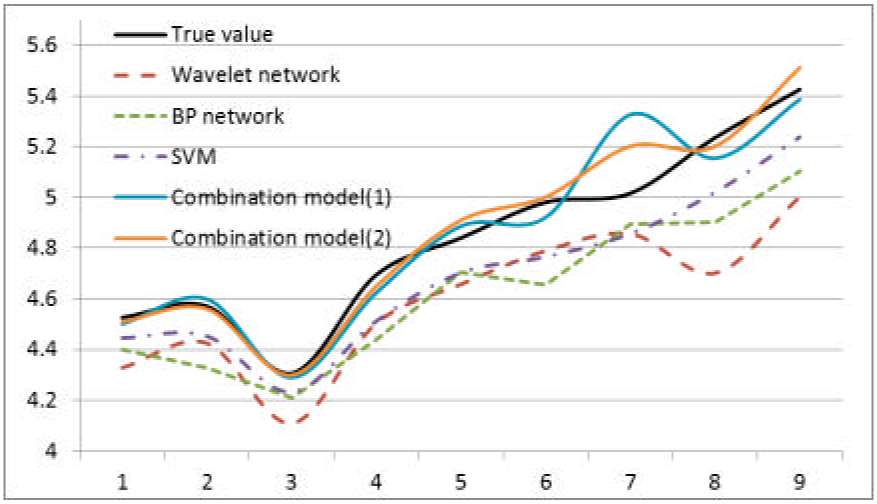

Fig. 4: Comparison curves of different forecasting models

Table 2: Grain yield predicted value of different model from 2001 to 2009

Year	True value	Wavelet network	BP network	SVM	Combination model (1)	Combination model (2)
2001	45263.6	43285.6	44001.3	44456.7	45012.1	45112.6
2002	45705.7	44249.7	43259.4	44528.9	45989.7	45601.9
2003	43069.5	41095.1	42108.6	42329.1	42894.6	42999.9
2004	46946.9	45035.7	44365.8	45123.6	46237.7	46500.6
2005	48402.1	46574.8	47025.3	47025.2	48865.2	49101.5
2006	49804.2	47895.5	46562.7	47654.4	49210.9	50012.8
2007	50160.2	48564.2	48936.1	48569.3	53261.3	52013.7
2008	52356.8	47012.6	49002.5	50200.6	51543.7	52013.7
2009	54254.3	50023.7	51021.9	52379.8	53856.7	55111.1

Table 3: Comparison of RMSE and MAPE of various models

Forecast models	RMSE	MAPE (%)
Wavelet network	3245.23	7.85
BP network	2511.2	6.23
SVM model	1256.4	3.29
Combination model (1)	851.8	1.16
Combination model (2)	650.3	0.85

CONCLUSION

From the Table 2, 3 and Fig. 4, we can see that the prediction accuracy of the combination model is improved greatly. Moreover, the organization topological structure improves the robustness of forecast model and reduces the probability of abnormal results. Test results show that the proposed method in this study has better nonlinear forecasting ability, higher prediction accuracy and broad application prospects.

REFERENCES

Bu, H. and R. Liu, 2011. A Combination Prediction Model Based on SVM and Its Application in Grain Output. In: Applied Informatics and Communication, Jun, Z., (Ed.). Springer, Berlin Heidelberg, pp: 74-81.

Eyheramendy, S., D. Lewis and D. Madigan, 2003. On the naive bayes model for text categorization. Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics, January 3-6, 2003, Hyatt Hotel, Key West, Florida, pp: 456-470.

Gartner, T. and P.A. Flach, 2001. WBC_{Svm}: Weighted bayesian classification based on support vector machine. Proceedings of the 18th International Conference on Machine Learning, June 2001, Willianstown, USA., pp: 154-161.

Mukherjee, S., E. Osuna and F. Girosi, 1997. Nonlinear prediction of chaotic time series using support vector machines. Proceedings of the 7th IEEE Workshop on Neural Networks for Signal Processing, September 24-26, 1997, Amelia Island, FL., pp: 511-520.

Tipping, M.E., 2001. Sparse Bayesian learning and the relevance vector machine. J. Machine Learn. Res., 1: 211-244.

Wu, H. and F. Wang, 2010. RVM-BASED ore grade forecasting model and its application. Proceedings of the 3rd IEEE International Conference on Computer Science and Information Technology, Volume 3, July 9-11, 2010, Chengdu, pp: 449-451.

Wu, H.X., H.X. Dong and J.Q. Su, 2008. 3D reconstruction from section plane views based on Self-adaptive neural network. Proceedings of the 2nd Intelligent Information Technology Application, Volume 3, December 20-22, 2008, Shanghai, pp: 84-88.

Yuan, Y. and Z. An, 2013. Study on SVM nonlinear combination forecasting method for grain yield based on rough set theory. Food Production SVM Nonlinear Combination Forecasting Model Based on Rough Set. http://www.study.edu.cn/releasestudy/content/201303-183

Zhao, X.Q., X.Y. Cao, Z.Q. Lan, S. Rao and Z.Y. Huang *et al.*, 2004. Study of the method for determining weighting coefficient of coal ash slagging fuzzy combination forecast based on rough set theory. J. China Coal Soc., 2: 222-224.