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 (24): 8066-8073, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.8066.8073

A New Nonlinear Combination Forecasting Algorithm and its Application in Digital Mine

Wu Huixin[1], Mo Duo[1] and Liu Ran[2]
[1]Department of Information Engineering,
North China University of Water Resources and Electric Power, 450046, Zhengzhou, China
[2]School of software, North China University of Water Resources and Electric Power,
450011, Zhengzhou, China

Abstract: Mineral resource prediction is the basis for many spatial operations in digital mine. Thus, there is a growing interest in constructing effective prediction algorithm, both in applications and in science. In this study, a new nonlinear combination forecasting algorithm based on adaptive neural network, support vector machine and relevance vector machine is presented to overcome the limitation in linear combination forecasting. Furthermore, a new method of selecting weight coefficient is proposed based on rough set theory. Theoretical analysis and forecasting examples all show that the new techniques has reinforcement learning properties and universalized capabilities. Finally, future trends for research and development in this area are highlighted.

Key words: Combination forecasting, SVM, RVM, BP neural network, digital mine

INTRODUCTION

With the rapid development of the three-dimensional (3D) modeling theory, computer graphics and the technology of computer software and hardware, the 3D modeling technique in the digital mine has been studied thoroughly. How to get complete geological data and manage the data efficiently become the key to the successful application of 3D modeling. To get complete geological data, traditionally method is using original geological drill hole data to infer attribute data information unknown areas based on the experience of technicians. This approach has the obvious disadvantage of subjective arbitrariness, that is, for the same geological drill hole information, different geological sections can be obtained by different technicians. For example, for the area between the two drill holes ZK_1 and ZK_2 shown in Fig. 1, there will be two different presumable results (The thick line represents sampling of ore and the thin line represents sampling of rock) (Wu and Wang, 2010).

In recent years, many new methods for getting complete geological data are proposed such as neural network, rough set theory, support vector machine (SVM), relevance vector machine(RVM) and so on. However, 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 turns subjective division into objective division under probability, which makes classification function reach likelihood function maximum for the training set.

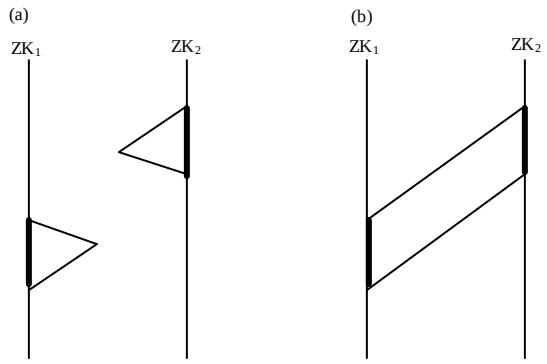

Fig. 1: Limitation of the traditional method

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, a new nonlinear combination forecasting algorithm based

Corresponding Author: Wu Huixin, North China University of Water Resources and Electric Power, 450011, Zhengzhou, China

on adaptive neural network, support vector machine and relevance vector machine (Tipping, 2001) is presented to overcome the limitation in linear combination forecasting. Furthermore, a new method of selecting weight coefficient is proposed based on rough set theory. Theoretical analysis and forecasting examples all show that the new techniques has reinforcement learning properties and universalized capabilities.

ALGORITHM FOR SELECTING WEIGHT COEFFICIENT OF NET WORKS

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:

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

$\hat{y}_t$ 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_1 …, 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:

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

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}

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

- 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

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

COMPOSITION FRAMEWORK FOR NONLINEAR COMBINATION ALGORITHM

In order to achieve the purpose of combination forecasting and get effectively forecast data. In this study, the single prediction models of adaptive neural network, support vector machine and relevance vector machine are selected to construct the combination forecasting model. Firstly, we discuss the basic working principle of each single model. Then we design and carry out a combined forecasting algorithm.

The optimized bp neural network: Taking into account the accuracy and reliability of engineering application, we select the widely used Back-propagation (BP) neural network. The BP neural network is a kind of multi-layer feed-forward network proposed by Rumelhart and Hinton (1986). The BP networks always consist of at least three hierarchical layers of neurons, an input layer, a hidden layer and an output layer. Each layer is fully connected to the next layer. Neural networks learn by adjustments of the weights of the various connections. The goal of earning is to minimize the error between the desired output and the actual output of the network. But 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 algorithm. So, in the study, a BP algorithm modified with additional momemtum and self-adaptive learning rate is adopted (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.

The changes of weight and bias of each training cycle are determined by the learning rate. The big learning rate can cause instability of net system but the small one can result in convergence speed slower and training time longer. The learning rate usually is determined through experiment or by experience. Self-adaptive learning rate refers to that the learning rate can be adjusted in training cycle based on the change of weights. If a modified value of weight decreases the error function, we should increase the learning rate. If not, we should decrease the learning rate. The adjusting equation for learning rate is as follows (in which, SSE is the sum squared errors):

$$\eta(k+1) = \begin{cases} 1.05\eta(k) & SSE(k+1) < SSE(k) \\ 0.7\eta(k) & SSE(k+1) > 1.04 \times SSE(k) \\ \eta(k) & others \end{cases}$$

SVM forecasting algorithm: 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.

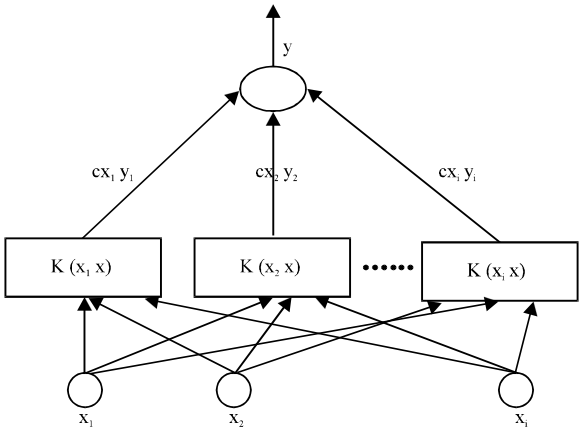

Fig. 2: Network model of SVM

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 (Bu and Liu, 2011).

RVM forecasting algorithm: Relevance vector machine is a sparse probability model based on support vector machine proposed by Tipping (2001). Its training is carried on under Bayesian framework, so we can get the distribution of predicted values by regression estimate with 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, $\phi_j(x)$ 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 function of 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 matrix of which the rows include the response of all kernel functions to input x_i:

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

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:

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

is covariance, A is diagonal matrix of (α_0, $\alpha_1 \ldots$ α_n) and:

$$\mu = \sigma^{-2}\sum \Phi^T t$$

is mean value. The likelihood distribution of training target value can realize marginalization by 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 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:

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

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

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

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.

In this study, the prediction algorithm of RVM can be summarized as the following steps (Wu and Wang, 2010):

- Initialize hyper-parameter $\{\alpha_i\}$ and variance σ^2
- Calculate posteriori statistic of weight μ and Σ
- Calculate all the y_I and re-estimate the $\{\alpha_i\}$ and σ^2
- If being convergent, go to step 5, otherwise go back step 2

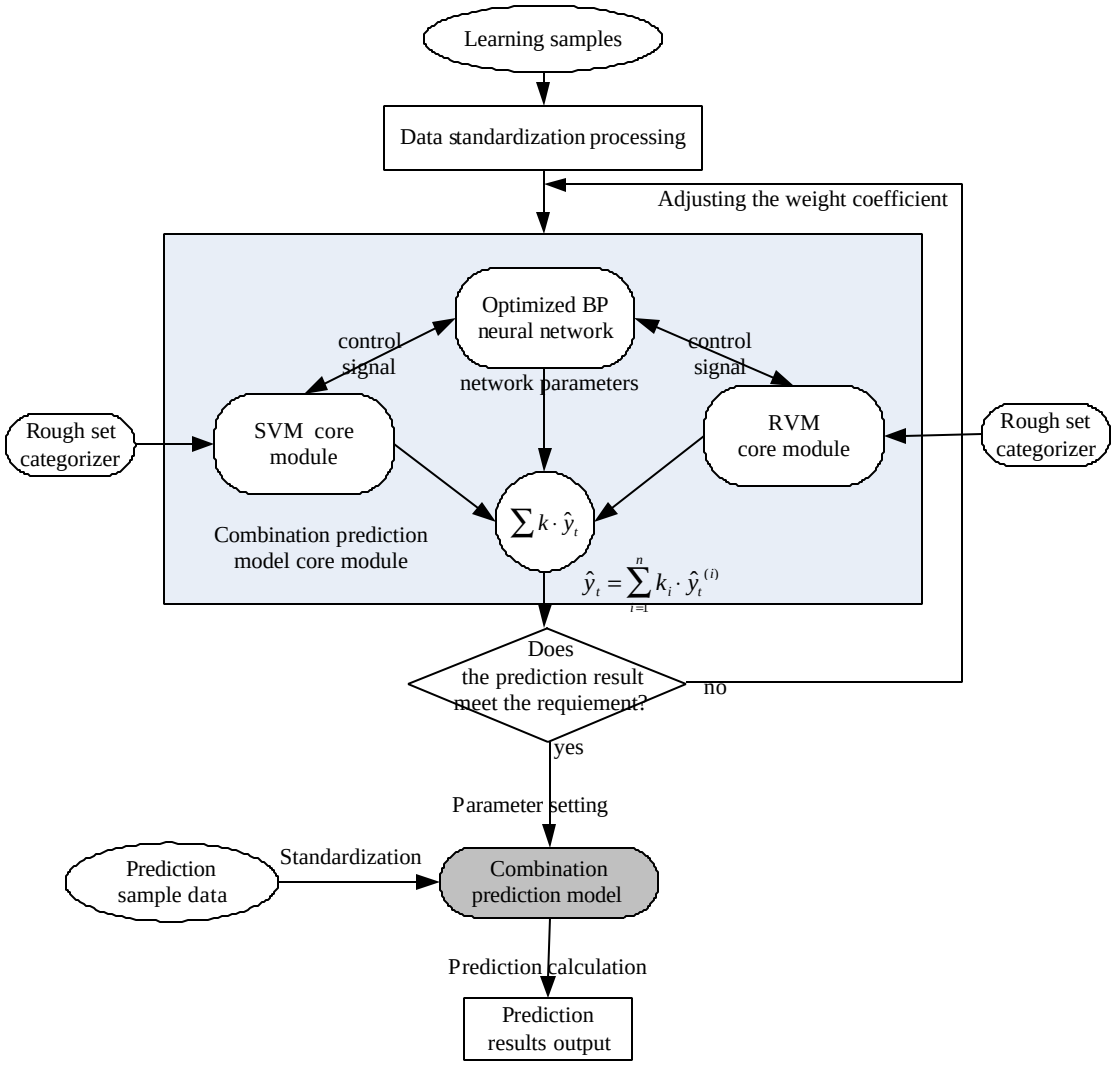

Fig. 3: Algorithm flow for combination prediction model

- Delete the weights and basis function of which $\alpha_i = \infty$
- Forecast new data and the mean value is y (x, μ)

Algorithm design and implementation for composition framework: The traditional combination forms are often linear. As the network structure is linear, prediction error may be passed from one forecasting model to another forecasting model, which can lead to error accumulation. Furthermore, the linear forecasting models have a lower robustness and have poor network tolerance. Therefore, in order to overcome the problems above, we proposed a new nonlinear combination forecasting model based on RVM, BP neural network and SVM. In the combination model, the prediction engine is formed by three kernel functions (BP neural network core, SVM core nodule and RVM core nodule). During the forecasting, the three kernel functions are organized together and control signal can be passed to each core nodule to adjust model parameter. That is to say, the three kernel functions work together dynamically and synchronously which enhances the stability and accuracy of forecast results. The algorithm framework of the combination model proposed in this study is shown in Fig. 3.

APPLICATION EXAMPLE AND COMPARATIVE EXPERIMENTAL ANALYSIS

In order to test and verify the effectiveness of the combination model proposed in this study, we prepare

Table 1: Training samples from zk729

Sample No.	Sample grade (%)	Sample No.	Sample grade (%)
1	11.26	13	50.67
2	15.14	14	50.07
3	46.79	15	53.59
4	12.14	16	60.68
5	11.57	17	44.05
6	12.54	18	56.27
7	10.35	19	50.86
8	9.45	20	59.34
9	12.25	21	53.65
10	18.62	22	7.69
11	54.48	23	7.35
12	53.47	24	6.01

Table 2: Training samples from zk729

Sample No.	Sample grade (%)	Sample No.	Sample grade (%)
1	18.55	13	67.64
2	53.96	14	58.10
3	50.62	15	42.49
4	55.65	16	18.23
5	53.72	17	56.44
6	60.87	18	56.44
7	54.11	19	7.45
8	55.44	20	52.01
9	54.55	21	7.23
10	51.48	22	6.17
11	52.27	23	5.69
12	46.97	24	5.01

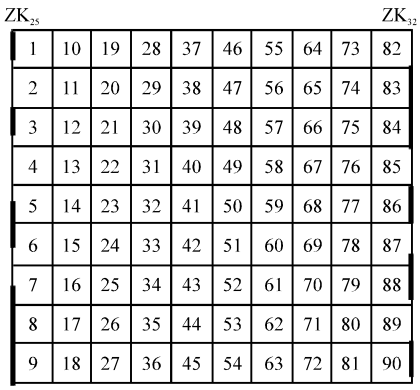

Fig. 4: Region segmentation for forecasting

complete ore grade attribute data of each geological cross-section. Generally speaking, we have only ore grade attribute data of each drill hole, however the attribute data between two drill holes we do not know. Therefore, the key question is to estimate ore grade of information unknown areas with the existing drill hole data. In this study, the area information unknown is divided into many small squares averagely and each small square is a basic unit to be estimated (as shown in Fig. 4).

Select geological drill hole data of ZK_{729} and ZK_{750} as Training samples. Standardized input data is show in Table 1 and 2. The output data is ore grade of each small ore block between ZK_{729} and ZK_{750}.

Table 3: Ore grade predicting results with SVM Wu and Wang (2010)

Ore no.	Predicted value	Measured value	Ore no.	Predicted value	Measured value
1	25.27	29.510	…	…	…
2	27.89	29.890	…	…	…
3	11.01	12.010	…	…	…
4	12.13	9.540	94	48.71	50.07
5	37.14	38.830	95	52.73	46.43
6	44.51	42.770	96	41.47	44.21
7	43.27	40.250	97	41.07	47.64
8	46.10	49.780	98	38.91	42.27
9	38.62	41.290	99	21.30	24.05
10	20.24	23.860	100	14.23	12.47

Table 4: Ore grade predicting results with RVM Wu and Wang (2010)

Ore no.	Predicted value	Measured value	Ore no.	Predicted value	Measured value
1	26.27	29.51	…	…	…
2	28.12	29.89	…	…	…
3	11.21	12.01	…	…	…
4	11.03	9.54	94	49.02	50.07
5	39.10	38.83	95	51.07	46.43
6	43.51	42.77	96	42.31	44.21
7	42.12	40.25	97	45.1	47.64
8	47.18	49.78	98	39.48	42.27
9	39.47	41.29	99	23.11	24.05
10	21.23	23.86	100	13	12.47

Table 5: Ore grade predicting results with combination algorithm

Ore no.	Predicted value	Measured value	Ore no.	Predicted value	Measured value
1	28.32	29.51	…	…	…
2	28.72	29.89	…	…	…
3	12.75	12.01	…	…	…
4	10.13	9.54	94	49.42	50.07
5	37.10	38.83	95	48.37	46.43
6	41.51	42.77	96	42.86	44.21
7	41.23	40.25	97	46.21	47.64
8	48.20	49.78	98	40.25	42.27
9	39.89	41.29	99	25.01	24.05
10	22.09	23.86	100	11.83	12.47

In order to test and verify the effectiveness of the combination model proposed in this study, comparative experimental analysis are made with single SVM and single RVM. The experimental results are shown in Table 3, 4 and 5.

Figure 5 shows the comparison curves of SVM predicted values and the measured values of ore grade.

Figure 6 shows the comparison curves of RVM predicted values and the measured values of ore grade.

The predicting result with the combination algorithm proposed in this study is shown in Table 5.

Figure 7 shows the comparison curves of combination model predicted values and the measured values of ore grade.

In order to evaluate the performance of forecast models, we chose the RMSE and MAPE as the evaluation indicators.

Table 6: Comparison of RMSE and MAPE of various models

Forecast models	RMSE	MAPE (%)
SVM model	5.35	7.23
RVM model	4.21	5.32
Combination model	3.49	3.56

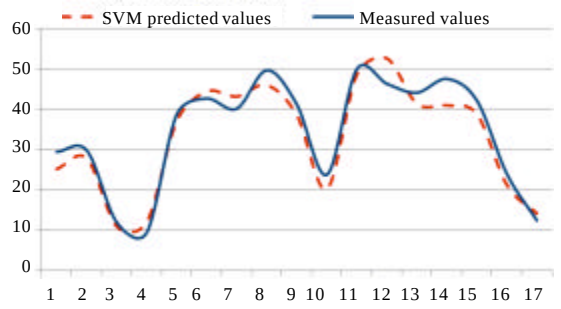

Fig. 5: Prediction result of single SVM

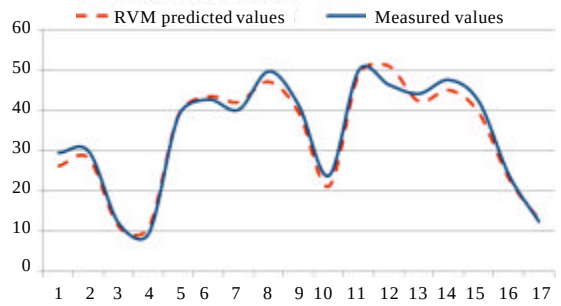

Fig. 6: Prediction result of single RVM

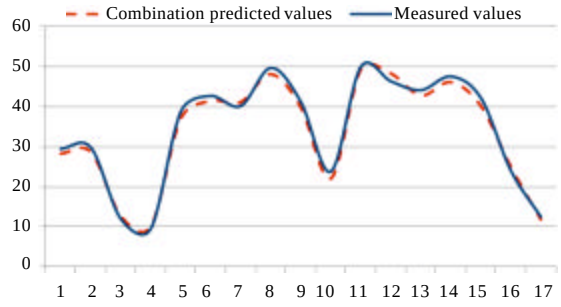

Fig. 7: Prediction result of single the combination model proposed in this study

The RMSE and MAPE are defined as follows:

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

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

In which, y_i is the true value, $\bar{y}_i$ is forecast value and n is the number of the forecast sample. Table 6 shows the comparison of RMSE and MAPE of various models.

CONCLUSIONS

From the comparative experimental analysis above, we can see that the prediction accuracy of the combination model is improved greatly. Moreover, the new technique has reinforcement learning properties and universalized capabilities. The organization topological structure improves the robustness of forecast model and reduces the probability of abnormal results. In the following work, the convergence speed of the network, network stability and so on will be the focus of our research work.

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.

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.paper.edu.cn/releasepaper/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.