A Reliability-Based RBF Network Ensemble Model for Foreign Exchange Rates Predication

Lean Yu[1], Wei Huang[3], Kin Keung Lai[2,4], and Shouyang Wang[1,3]

[1] Institute of Systems Science, Academy of Mathematics and Systems Science,
Chinese Academy of Sciences, Beijing 100080, China
{yulean, sywang, whuang}@amss.ac.cn
[2] Department of Management Sciences, City University of Hong Kong,
Tat Chee Avenue, Kowloon, Hong Kong
{msyulean, mskklai}@cityu.edu.hk
[3] School of Management, Huazhong University of Science and Technology,
1037 Luoyu Road, Wuhan 430074, China
[4] College of Business Administration, Hunan University, Changsha 410082, China

Abstract. In this study, a reliability-based RBF neural network ensemble forecasting model is proposed to overcome the shortcomings of the existing neural ensemble methods and ameliorate forecasting performance. In this model, the ensemble weights are determined by the reliability measure of RBF network output. For testing purposes, we compare the new ensemble model's performance with some existing network ensemble approaches in terms of three exchange rates series. Experimental results reveal that the prediction using the proposed approach is consistently better than those obtained using the other methods presented in this study in terms of the same measurements.

1 Introduction

Combining the outputs of several neural networks into an aggregate output often gives improved accuracy over any individual output [1]. Usually, the output of an ensemble network is a weighted average of the outputs of each network, with the ensemble weights determined as a function of the relative error of each network determined in training [1]. Accordingly, the resulting ensemble network often outperforms the individual networks. Therefore there is a growing research stream [1-5] into ensemble methods. For example, performance improvement can result from training the individual networks to be decorrelated with each other [2] with respect to their errors. Usually, the weights of the existing ensemble methods are determined by error or error variance [1, 9, 11] without considering the reliability of neural network output. However, error-dependent neural network ensemble methods do not often obtain consistent good performances in forecasting. In some experiments, such as [1, 6], error-dependent ensemble network sometimes performs worse than the individual neural networks.

Under such backgrounds, we propose a novel neural network ensemble forecasting approach that differs in that the ensemble weights are determined the reliability of neural network output. In this study, the reliability is used as a confidence measure of network output. That is, the ensemble weights are proportional to the reliability

I. King et al. (Eds.): ICONIP 2006, Part III, LNCS 4234, pp. 380 – 389, 2006.

measure of the respective outputs. In addition, the neural network type used in this study is radial basis function (RBF) neural network because it can generate a confidence measure due to its specificity.

The motivation of this study is to formulate a reliability-based RBF network ensemble forecasting model for exchange rates prediction and compare its performance with other existing network ensemble forecasting approaches. The rest of the study is organized as follows. The next section presents some previous work done in the ensemble methods in terms of forecasting. Section 3 describes the reliability-based RBF neural network ensemble method in detail. To verify the effectiveness of the proposed method, several experiments are performed in Section 4. Section 5 concludes.

2 Previous Studies

This section presents some earlier work done in ensemble methods in terms of prediction. Suppose there are n individual neural networks trained on a set $\boldsymbol{D} = \{x_i, y_i\}$ (i = 1, 2, … , n).

2.1 The Brief Description of Neural Ensemble Predictor

According to the previous assumption, there are n individual neural network outputs, i.e., $\hat{f}_1(x), \hat{f}_2(x), \cdots, \hat{f}_n(x)$. The main question of neural ensemble predictor is how to combine (ensemble) these different outputs into an aggregate output $\hat{f}(x)$, which is assumed to be a more accurate output. The general form of the model for such an ensemble predictor can be defined as

$$\hat{f}(x) = \sum_{i=1}^{n} w_i \hat{f}_i(x) \tag{1}$$

where w_i denotes the assigned weight of $\hat{f}_i(x)$, and in general the sum of the weight is equal to one, i.e., $\sum w_i = 1$. In neural network ensemble forecasting, how to determine ensemble weights is a focus issue. As earlier mentioned, there are a variety of methods for determining ensemble weights in the past work, which is presented in the following. Typically, there are four ensemble methods, which are described below.

2.2 The Simple Averaging Ensemble Method (SAE)

A simple method for combining network outputs is to simply average individual network predictors together, so this approach is called as the "simple averaging ensemble (SAE) method". The SAE is defined as

$$\hat{f}_{SAE}(x) = \sum_{i=1}^{n} w_i \hat{f}_i(x) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}_i(x) \tag{2}$$

where the weight of each individual network output $w_i = 1/n$.

Simple averaging ensemble method is one of the most frequently used combination approaches is easy to understand and implement [4]. Some experiments [8-9] have shown that this approach by itself can lead to improved performance [1] and it is an effective approach to improve neural network performance. Specially, it is more useful when the local minima of ensemble members are different, i.e., when the local minima of ensemble networks are different. Different local minima mean that ensemble members are diverse. Thus averaging can reduce the ensemble variance.

However, this approach treats each member equally, i.e., it does not stress ensemble members that can make more contribution to the final generalization. That is, it does not take into account the fact that some networks may be more accuracy than others. If the variances of ensemble networks are very different, we do not expect to obtain a better result using simple averaging [10]. In addition, since the weights in the combination are so unstable, a simple average may not the best choice in practice [6].

2.3 The Simple MSE Ensemble Method (SMSE)

The simple MSE method estimates the linear weight parameter wi in Equation (1) by minimizing the mean squared error (MSE) [11], that is, for i = 1, 2, …, n,

$$w_{opt,i} = \arg\min_{w_i}\left\{\sum_{j=1}^{m}\left(w_i^T \hat{f}_i(x_j) - d_{ji}(x_j)\right)^2\right\} = \left(\sum_{j=1}^{m}\hat{f}_i(x_j) f_i^T(x_j)\right)^{-1}\sum_{j=1}^{m} d_{ji}(x)\hat{f}_i(x_j) \qquad (3)$$

where d (x) is the expected value.

The simple MSE solution seems to be reasonable, but, as Breiman [9] has pointed out, this approach has two serious problems in practice: a) the data are used both in the training of each predictor and in the estimation of w_i, and b) individual predictors are often strongly correlated since they try to predict the same task. Due to these problems, this approach's generalization ability will be poor.

2.4 The Stacked Regression Ensemble Method (SRE)

The stacked regression method was proposed by Breiman [9] in order to solve the problems associated with the previous MSE method. Thus, the stacked regression method is also called the modified MSE method. This approach utilizes cross-validation data to modify the simple MSE solution, i.e.,

$$w_{opt,i} = \arg\min_{w_i}\left\{\sum_{j=1}^{m}\left(w_i^T g_i(x_j) - d_{ji}(x_j)\right)^2\right\}, \quad i = 1,2,\ldots,n \qquad (4)$$

where $g_i(x_j) = \left(\hat{f}_i^{(1)}(x_j; D_{cv}), \cdots, \hat{f}_i^{(m)}(x_j; D_{cv})\right)^T \in \Re^M$ is a cross-validated version $\hat{f}_i(x_j) \in \Re^M$ and D_{cv} is the cross-validated data.

Although this approach overcomes the limitations of the simple MSE method, the solution is based on the assumption that the error distribution of each validated set is normal [10]. In practice, however, this normal assumption does not always hold and thus this approach does not lead to the optimal solution in the Bayes sense [10].

2.5 The Variance-Based Weighting Ensemble Method (VWE)

The variance-based weighting ensemble approach estimates the weight parameter w_i by minimizing error variance σ_i^2[1]; all predictors are error-independent networks, i.e.,

$$w_{opt,i} = \arg\min_{w_i} \left\{ \sum_{i=1}^{n} \left(w_i \sigma_i^2) \right)^2 \right\}, \quad (i = 1, 2, \cdots, n) \tag{5}$$

under the constraints $\sum_{i=1}^{n} w_i = 1$ and $w_i \geq 0$. Using the Lagrange multiplier, the optimal weights are:

$$w_{opt,i} = \frac{\left(\sigma_i^2\right)^{-1}}{\sum_{j=1}^{n} \left(\sigma_j^2\right)^{-1}}, \quad (i = 1, 2, \cdots, n) \tag{6}$$

The variance-based weighting method is based on the assumption of error independence. Moreover, as earlier mentioned, individual predictors are often strongly correlated for the same task. This indicates that this approach has serious drawbacks for minimizing error-variance when neural predictors with strong correlation are included within the combinatorial members.

According to the previous descriptions and literature review, the above four neural network ensemble methods have widely been used, but we can also find that the weights of these ensemble methods are determined by the error or error variance. In practice, error-dependent network ensemble model may not often obtain good forecasting performance when error is correlated each other [2]. Furthermore, the accuracy of each individual network predictor is different in different conditions because neural network learning by itself is "the state of the art". Thus, it is necessary to measure the confidence of the forecasting results of each output before they are combined. In such situations, a novel reliability-based neural ensemble method is proposed to overcome the above problems, which is presented in the following. It is worth noting that the reliability is used to measure the confidence of neural network output and the RBF network is used to constitute a neural network ensemble.

3 The Reliability-Based RBF Network Ensemble Model

In this section, we first introduce the radial basis function (RBF) neural network briefly. Then the confidence measure — reliability is induced from the RBF neural network. Finally, based on the reliability measure, a reliability-based RBF neural network ensemble method is formulated.

3.1 Overview of RBF Neural Networks

The RBF neural network [12, 13] is generally composed of three layers: input layer, hidden layer and output layer. The input layer feeds the input data to each of the nodes of the hidden layer. The hidden layer of nodes differs greatly from other neural networks in that each node represents a data cluster which is centered at a particular point with a given radius. Each node in the hidden layer calculates the distance from the input vector to its own center. The calculated distance is transformed via some basis function and the result is output from a node. The output from the node is multiplied by a constant or weighting value and fed into the output layer. The output layer consists of only one node which acts to sum the outputs of the previous layer and to yield a final output value. A generic architecture of an RBF network with *k* input and *m* hidden nodes is illustrated in Fig. 1.

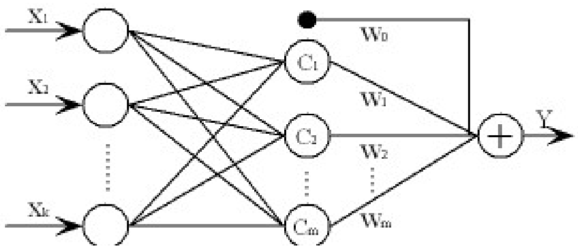

Fig. 1. The generic architecture of the RBF neural network

The computation process of the RBF neural network follows the following procedures. When the network receives a k dimensional input vector $\boldsymbol{X}$, the network computes a scalar value using the following formula:

$$Y = f(X) = w_0 + \sum_{i=1}^{m} w_i \varphi(D_i) \quad (7)$$

where w_0 is the bias, w_i is the weight parameter, m is the number of nodes in the hidden layers of the RBF neural network, and $\varphi(D_i)$ is the radial basis function. In this study, the Gaussian function is used as radial basis function, as shown below.

$$\varphi(D_i) = \exp(-D_i^2 / \sigma^2) \quad (8)$$

where σ is the radius of the cluster represented by the center node, the D_i representes the distance between the input vector $\boldsymbol{X}$ and all the data centers. It is clear that $\varphi(D_i)$ will return values between 0 and 1. Usually, the Euclidean norm is used to calculate distance, but other metrics can also be used. The Euclidean norm is calculated by

$$D_i = \sqrt{\sum_{j=1}^{k} (x_j - c_{ji})^2} \quad (9)$$

where c is a cluster center for any of the given nodes in the hidden layer.

Complex nonlinear systems such as time series data are generally difficult to model using standard linear regression [13]. Dissimilar to the regression, neural networks are nonlinear and their parameters are determined by some learning techniques and search algorithms such as error back propagation and steep gradient algorithm. The main drawback of the standard neural networks is that their parameters learning algorithm is time-consuming and have a tendency to get stuck at local minima [14]. But RBF neural networks overcome the above problems to obtain good performance since their parameters that need to be trained are the ones in the hidden layer of the network. Finding their values is the solution of a linear problem and can be obtained through interpolation [12]. Therefore, their parameters are found much faster than in other neural networks. Furthermore, the RBF network can usually reach near perfect accuracy on the training data set without trapping into local minima [13].

3.2 Confidence Measure and Reliability

Once the parameters are determined by training the data, the RBF network can be applied to perform prediction for any unknown input vectors. But the confidence of the output results is an important issue for any prediction problems. In this study, other than the basic output, the RBF network is also used to generate confidence measure to verify the reliability of network output.

Actually, using RBF neural network to generate a confidence measure is not a new idea. Lee [15] used this principle in a handwritten digit recognition program and achieved good results. Lee used as a measure of confidence the numeric difference between the two output nodes with the highest values. Leonard et al. [16] explored two methods of measuring the reliability of prediction: the maximum activation value function and Parzen estimator and concluded the Parzen method superior as it took into account the distribution of the training data. Wedding and Cios [13] also used the output of the Gaussian basis function to calculate a certainty factor value. The certainty factor was calculated from the input vector's proximity to the hidden layer's node centers rather than the data distribution of the training vectors. The reliability proposed in this study is a variation of [13].

In this study, a new confidence measure called reliability for RBF neural networks is proposed. The reliability can be generated by RBF through using the $\varphi(D_i)$ value from the basis functions in the hidden layer, as shown in Equation (8). When the output $\varphi(D_i)$ of a hidden layer node is high (near one), then this indicates that a data point lies near the center of a cluster and the data is familiar to that particular node and therefore the network is confident in the output. Conversely, if the $\varphi(D_i)$ is very small (near zero), this implies that the data point lies outside of a cluster and thus the network is not reliable in the result. Because the reliability is used to measure the overall network, then all the values from different centers should be combined into a single one. According to the positive correlation between the reliability and $\varphi(D_i)$, the reliability measure of overall network can be computed by

$$R_j = \frac{1}{m}\sum_{i=1}^{m}\varphi(D_i) \qquad \textbf{(10)}$$

where R_j is the reliability of the individual network j and $\varphi(D_i)$ is the corresponding output value at hidden neuron i. As Equation (10) revealed, the larger the R, the larger the reliability. Then the weight of this individual network should be larger than that of others in the ensemble. That is, the large reliability should correspond to large weight in the ensemble network so as to improve the prediction accuracy. In terms of reliability measure, a reliability-based RBF network ensemble forecasting model is formulated in the following.

3.3 Reliability-Based Weighted Ensemble (RWE) Method

Instead of choosing error-dependent weights, we allow the weights to adjust to be proportional to the reliability measure of the respective network outputs. We might achieve better performance. Define the reliability-based weighted ensemble (RWE) network as:

$$\hat{f}_{RWE}(x) = \sum\nolimits_{i=1}^{n} w_i \hat{f}_i(x) \tag{11}$$

where the w_i s are as:

$$w_i = \frac{R_i}{\sum\nolimits_{j=1}^{n} R_j} \tag{12}$$

The w_i s sum to one, so $\hat{f}_{RWE}$ is a weighted average of the individual network outputs. The difference is that the weight vector is determined by confidence measure — reliability to try to give the best prediction under consideration, instead of choosing error-dependent weights with respect to a cross validation set. Each RBF network's contribution to the sum is proportionate to its confidence measure, i.e., reliability. For testing, several experimental examples are presented below.

4 Simulations

4.1 Data

In this study three foreign exchange series are selected for comparison purposes. The foreign exchange data used in this paper are monthly and are obtained from Pacific Exchange Rates Services (http://fx.sauder.ubc.ca/), provided by Professor Werner Antweiler, University of British Columbia, Vancouver, Canada. They consist of the US dollar against each of the three currencies — German marks (DEM), British pound (GBP) and Japanese yen (JPY) studied in this paper. We take monthly data from January 1971 to December 2000 as in-sample (training periods) data sets (360 observations including 60 samples for validations). We also take the data from January 2001 to December 2004 as out-of-sample (testing periods) data sets (48 observations), which is used to evaluate the good or bad performance of prediction based on some evaluation measurement. In order to save space, the original data are not listed here, and detailed data can be obtained from the website or from the authors. For comparison, two typical indicators, mean squared error (*MSE*) and directional statistics (D_{stat}) were used in this study. Usually, *MSE* is an ordinary level predication measurement and while D_{stat} is a directional prediction measurement, as defined in [7].

4.2 Empirical Results

In this study each of the three ensemble methods was implemented and tried on several data sets for comparison. Ten feed-forward neural networks with sigmoidal activation functions and four hidden nodes were trained for each training set, then tested as an ensemble for each method for the testing set. Each network was trained with standard back-propagation for 100 iterations with a learning rate of 0.2 using the Matlab software package, which is producted by Mathworks Laboratory Corporation. In addition, the best individual network using cross-validation (CV) (NCV for short) [3] method (i.e., select the individual network by minimizing the mean squared error on CV) is chosen as benchmark model for comparison. Accordingly, the results obtained are reported in Tables 1 and 2.

Table 1. A comparison of MSE between different methods for the three exchange rates

	DEM		GBP		JPY	
	MSE	Rank	*MSE*	Rank	*MSE*	Rank
NCV	0.0035	5	0.0019	5	0.0048	6
SAE	0.0038	6	0.0018	4	0.0043	5
SMSE	0.0031	4	0.0021	6	0.0040	4
SRE	0.0029	3	0.0015	2	0.0037	3
VWE	0.0025	2	0.0016	3	0.0031	2
RWE	0.0022	1	0.0009	1	0.0028	1

Table 2. A comparison of D_{stat} between different methods for the three exchange rates

	DEM		GBP		JPY	
	D_{stat} (%)	Rank	D_{stat} (%)	Rank	D_{stat} (%)	Rank
NCV	64.58	6	70.83	5	62.50	6
SAE	72.91	4	68.75	6	66.67	5
SMSE	66.67	5	70.83	4	75.00	4
SRE	79.17	3	83.33	3	77.08	3
VWE	81.25	2	85.41	2	81.25	2
RWE	87.50	1	89.58	1	83.33	1

Tables 1 and 2 give clear comparisons of various methods for the three currencies via *MSE* and D_{stat}. Generally speaking, these tables provide comparisons of level and direction among these different methods. Experimental results reveal that the prediction performance of the dynamically weighted ensemble forecasting model is better than those of other ensembles models.

Focusing on the *MSE* indicator, our proposed ensemble method performs the best in all the cases, followed by the VWE, SRE. Interestingly, the *MSE* of the SAE are not better than those of the best individual network model for the DEM testing case, and the *MSE* of the NCV are worse than those of the SMSE for the GBP case, implying that the SAE and SMSE does not consider the fact that some networks may be more accurate than the others.

However, the low *MSE* does not necessarily mean that there is a high hit ratio for foreign exchange movement direction prediction. Thus the D_{stat} comparison is necessary for practitioners. Focusing on D_{stat} of Table 2, we are not hard to find that the proposed dynamically weighted ensemble forecasting model outperforms the other ensemble models and the benchmark model according to the rank; furthermore, from the business practitioners' point of view, D_{stat} is more important than MSE because the former is an important decision criterion in foreign exchange trading decision. With reference to Table 2, the differences between the different models are very significant. For instance, for the JPY testing case, the D_{stat} for the best individual ANN model via NCV is only 62.50%, for the SAE method it is 66.67%, for the SMSE, it is 75.00%, and the D_{stat} for SRE is 77.08%, and for the VWE, D_{stat} is 81.25%; while for the RWE method, D_{stat} reaches 83.33%. Furthermore, like *MSE* indicator, the proposed ensemble method performs the best in all the cases, followed by VWE, SRE, and the poorest is the individual network model via NCV. The main reason is that our proposed approach can adjust its weights dynamically, giving it an advantage over other ensemble methods whose weights are fixed as a part of training.

5 Conclusions

This study proposes a reliability-based RBF neural network ensemble forecasting model to obtain accurate prediction results and improve prediction quality further. In terms of the empirical results, we can find that across different ensemble models for the test cases of three main currencies — German marks (DEM), British pound (GBP) and Japanese yen (JPY) — on the basis of different criteria, our proposed ensemble method performs the best. In the proposed reliability-based RBF network ensemble model testing cases, the *MSE* is the lowest and the D_{stat} is the highest, indicating that the proposed reliability-based RBF network ensemble model can be used as a viable alternative ensemble solution to exchange rates prediction.

Acknowledgements

This work is partially supported by National Natural Science Foundation of China, Chinese Academy of Sciences, Key Research Institute of Humanities and Social Sciences in Hubei Province-Research Center of Modern Information Management and Strategic Research Grant of City University of Hong Kong (SRG No. 7001677).

References

1. Perrone, M.P., Cooper, L.N.: When Networks Disagree: Ensemble Methods for Hybrid Neural Networks. In Mammone, R.J. (ed.): Neural Networks for Speech and Image Processing, Chapman-Hall (1993) 126-142
2. Rosen, B.E.: Ensemble Learning Using Decorrelated Neural Networks. Connection Science 8 (1996) 373-384
3. Krogh, A., Vedelsby, J.: Neural Network Ensembles, Cross Validation, and Active Learning. In Tesauro, G., Touretzky, D., Leen, T. (eds.): Advances in Neural Information Processing Systems, Cambridge, MA, MIT Press 7 (1995) 231-238
4. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
5. Tumer, K., Ghosh, J.: Error Correlation and Error Reduction in Ensemble Classifiers. Connection Science 8 (1996) 385-404
6. Kang, B.H.: Unstable Weights in the Combination of Forecasts. Management Science 32 (1986) 683-695
7. Yu, L., Wang, S.Y., Lai, K.K.: A Novel Nonlinear Ensemble Forecasting Model Incorporating GLAR and ANN for Foreign Exchange Rates. Computers and Operations Research 32 (2005) 2523-2541
8. Hansen L.K., Salamon, P.: Neural Network Ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 993-1001
9. Breiman, L.: Bias, Variance, and Arcing Classifiers. Technical Report TR460, Department of Statistics, University of California, 1994

10. Ueda, N.: Optimal Linear Combination of Neural Networks for Improving Classification Performance. IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 207-215
11. Benediktsson, J.A., Sveinsson, J.R., Ersoy, O.K., Swain, P.H.: Parallel Consensual Neural Networks. IEEE Transactions on Neural Networks 8 (1997) 54-64
12. Bishop, C.M.: Improving the Generalization Properties of Radial Basis Function Neural Networks. Neural Computation 3 (1991) 579-588
13. Wedding II, D.K., Cios, K.J.: Time Series Forecasting by Combining RBF Networks, Certainty Factors and the Box-Jenkins Model. Neurocomputing 10 (1996) 149-168
14. Chen, S., Billings, S.A., Cowan, C.F.N., Grant, P.M.: Nonlinear Systems Identification Using Radial Basis Functions. International Journal of Systems Science 21 (1990) 2513-2539
15. Lee, Y.C.: Handwritten Digit Recognition Using K Nearest Neighbor, Radial Basis Function and Backpropagation Neural Networks. Neural Computation 3 (1991) 440-449
16. Leonard, J.A., Kramer, M.A., Unger, L.H.: A Neural Network Architecture that Computes its Own Reliability. Computers and Chemical Engineering 16 (1992) 819-835