Stock Index Modeling Using Hierarchical Radial Basis Function Networks

Yuehui Chen[1], Lizhi Peng[1], and Ajith Abraham[1,2]

[1]School of Information Science and Engineering
Jinan University, Jinan 250022, P.R.China
Email: yhchen@ujn.edu.cn
[2]IITA Professorship Program, School of Computer Science and Engg.
Chung-Ang University, Seoul, Republic of Korea
Email: ajith.abraham@ieee.org

Abstract. Forecasting exchange rates is an important financial problem that is receiving increasing attention especially because of its difficulty and practical applications. This paper proposes a Hierarchical Radial Basis Function Network (HiRBF) model for forecasting three major international currency exchange rates. Based on the pre-defined instruction sets, HRBF model can be created and evolved. The HRBF structure is developed using the Extended Compact Genetic Programming (ECGP) and the free parameters embedded in the tree are optimized by the Degraded Ceiling Algorithm (DCA). Empirical results indicate that the proposed method is better than the conventional neural network and RBF networks forecasting models.

1 Introduction

Exchange rates are affected by many highly correlated economic, political and even psychological factors. These factors interact in a very complex fashion. Exchange rate series exhibit high volatility, complexity and noise that result from an elusive market mechanism generating daily observations [1]. Much research effort has been devoted to exploring the nonlinearity of exchange rate data and to develop specific nonlinear models to improve exchange rate forecasting, i.e., the Autoregressive Random Variance (ARV) model [2], Autoregressive Conditional Heteroscedasticity (ARCH) [3], self-exciting threshold autoregressive models [4]. There has been growing interest in the adoption of neural networks, fuzzy inference systems and statistical approaches for exchange rate forecasting problem [5][13][14].

For a recent review of neural networks based exchange rate forecasting, please consult [7]. The input dimension (i.e. the number of delayed values for prediction) and the time delay (i.e. the time interval between two time series data) are two critical factors that affect the performance of neural networks. The selection of dimension and the number of time delays has great significance in time series prediction.

Hierarchical neural networks consist of multiple neural networks assembled in different level or cascade architecture. Mat Isa et al. used Hierarchical Radial Basis Function (HiRBF) to increase RBF performance in diagnosing cervical cancer [17]. HiRBF cascading together two RBF networks, where both networks have different structure but using the same learning algorithms. The first network classifies all data and performs a filtering process to ensure that only certain attributes to be fed to the second network. The study shows that HiRBF performs better then the single RBF model. HiRBF has been proved effective in the reconstruction of smooth surfaces from sparse noisy data points [18]. In order to improve the model generalization performance, a selective combination of multiple neural networks by using Bayesian method was proposed in [19].

In this paper, an automatic method for constructing HiRBF network is proposed. Based on the pre-defined instruction sets, a HiRBF network can be created and evolved. The HiRBF network also allows input variables selection. In our previous studies, in order to optimize Flexible Neural Tree (FNT) the hierarchical structure of FNT was evolved using Probabilistic Incremental Program Evolution algorithm (PIPE) [8][9] and Ant Programming with specific instructions. In this research, the hierarchical structure is evolved using the Extended Compact Genetic Programming (ECGP), a tree-structure based evolutionary algorithm. The fine tuning of the parameters encoded in the structure is accomplished using the degraded ceiling algorithm [16]. The proposed method interleaves both optimizations. The novelty of this paper is in the usage of HiRBF model for selecting the important inputs and/or time delays and for forecasting foreign exchange rates.

2 The Hierarchical RBF Model

A function set F and a terminal instruction set T used for generating a hierarchical RBF model are described as $S = F \bigcup T = \{+_2, +_3, \ldots, +_N\} \bigcup \{x_1, \ldots, x_n\}$, where $+_i (i = 2, 3, \ldots, N)$ denote non-leaf nodes' instructions and taking i arguments. x_1,x_2,...,x_n are leaf nodes' instructions and taking no other arguments. The output of a non-leaf node is calculated as a RBF neural network model (see Fig.1). From this point of view, the instruction $+_i$ is also called a basis function operator with i inputs.

The basis function operator is shown in Fig.1(left). In general, the basis function networks can be represented as

$$y = \sum_{i=1}^{m} \omega_i \psi_i(x; \theta) \tag{1}$$

where $x \in R^n$ is input vector, $\psi_i(x; \theta)$ is ith basis function, and ω_i is the corresponding weights of ith basis function and θ is the parameter vector used in the basis functions. In this research, Gaussian radial basis functions are used,

$$\psi_i(x; \theta) = \prod_{j=1}^{n} exp(-\frac{\| x_j - b_j \|^2}{{a_j}^2}) \tag{2}$$

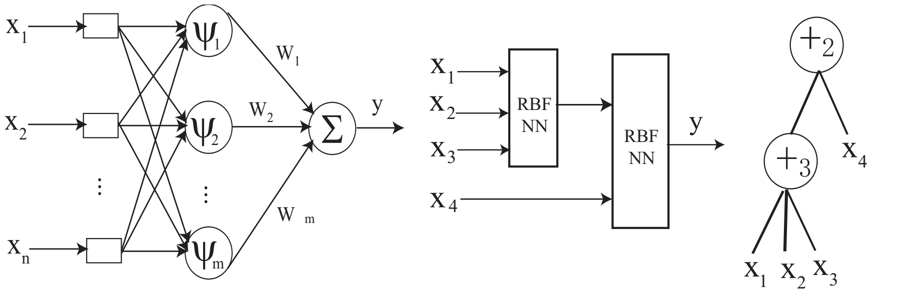

Fig. 1. A RBF network (left), a hierarchical RBF network (middle), and a tree-structural representation of the HRBF (right)

and the number of basis functions used in hidden layer is same with the number of inputs, that is, $m = n$.

In the creation process of HiRBF tree, if a nonterminal instruction, i.e., $+_i(i = 2, 3, 4, \ldots, N)$ is selected, i real values are randomly generated and used for representing the connection strength between the node $+_i$ and its children. In addition, $2 \times n^2$ adjustable parameters a_i and b_i are randomly created as Gaussian radial basis function parameters. The output of the node $+_i$ can be calculated by using (1) and (2). The overall output of HiRBF tree can be computed from left to right by depth-first method, recursively.

Tree Structure Optimization. Finding an optimal or near-optimal HiRBF is formulated as a product of evolution. In our previously studies, the Genetic Programming (GP), Probabilistic Incremental Program Evolution (PIPE) have been explored for structure optimization of the FNT [8][9]. In this paper, the Extended Compact Genetic Programming (ECGP) [11] is employed to find an optimal or near-optimal HiRBF structure.

ECGP is a direct extension of ECGA to the tree representation which is based on the PIPE prototype tree. In ECGA, Marginal Product Models (MPMs) are used to model the interaction among genes, represented as random variables, given a population of Genetic Algorithm individuals. MPMs are represented as measures of marginal distributions on partitions of random variables. ECGP is based on the PIPE prototype tree, and thus each node in the prototype tree is a random variable. ECGP decomposes or partitions the prototype tree into sub-trees, and the MPM factorises the joint probability of all nodes of the prototype tree, to a product of marginal distributions on a partition of its sub-trees. A greedy search heuristic is used to find an optimal MPM mode under the framework of minimum encoding inference. ECGP can represent the probability distribution for more than one node at a time. Thus, it extends PIPE in that the interactions among multiple nodes are considered.

Parameter Optimization with Degraded Ceiling Algorithm. Simulated annealing is one of the most widely studied local search meta-heuristics. It was proposed as a general stochastic optimization technique in 1983 [15] and has been applied to solve a wide range of problems including the weights optimization of

```
Set the initial solution S
Calculate initial fitness function f(s)
Initial ceiling B=f(s)
Specify input parameter dB
While not some stopping condition do
   define neighbourhood N(s)
   Randomly select the candidate solution s* in N(s)
   If ( f(s*) < f(s) )   or  ( f(s*) <= B )
   Then accept s*
```

Fig. 2. The Degraded ceiling algorithm

a neural network. The basic ideas of the simulated annealing search are that it accepts worse solutions with a probability $p = e^{-\frac{\delta}{T}}$, where $\delta = f(s^*) - f(s)$, the s and s^* are the old and new solution vectors, $f(s)$ denotes the cost function, the parameter T denotes the temperature in the process of annealing. Originally it was suggested to start the search from a high temperature and reduce it to the end of the process by an equation: $T_{i+1} = T_i - T_i * \beta$. However, the cooling rate β and initial value of T should be carefully selected since it is problem dependent.

The degraded ceiling algorithm also keeps the acceptance of worse solutions but with a different manner [16]. It accepts every solution whose objective function is less than or equal to the upper limit B, which is monotonically decreased during the search. The procedure for the degraded ceiling algorithm is given in Fig.2.

Procedure of the General Learning Algorithm. The general learning procedure for constructing the HiRBF network can be described as follows.

1) Create an initial population randomly (HiRBF trees and its corresponding parameters);
2) Structure optimization is achieved by using ECGP algorithm;
3) If a better structure is found, then go to step 4), otherwise go to step 2);
4) Parameter optimization is achieved by the DCA algorithm as described in subsection 2. In this stage, the architecture of HiRBF model is fixed, and it is the best tree developed during the end of run of the structure search. The parameters (weights and flexible activation function parameters) encoded in the best tree formulate a particle.
5) If the maximum number of local search is reached, or no better parameter vector is found for a significantly long time then go to step 6); otherwise go to step 4);
6) If satisfactory solution is found, then the algorithm is stopped; otherwise go to step 2).

Variable Selection using Hierarchical RBF Paradigms. It is often a difficult task to select important variables for a classification or regression problem, especially when the feature space is large. Conventional RBF neural network usually cannot do this. In the perspective of hierarchical RBF framework, the nature of model construction procedure allows the HiRBF to identify important input features in building an HiRBF model that is computationally efficient and effective. The mechanisms of input selection in the HiRBF constructing procedure are as follows. (1) Initially the input variables are selected to formulate the HiRBF model with same probabilities; (2) The variables which have more contribution to the objective function will be enhanced and have high opportunity to survive in the next generation by a evolutionary procedure; (3) The evolutionary operators i.e., crossover and mutation, provide a input selection method by which the HiRBF should select appropriate variables automatically.

3 Exchange Rates Forecasting Using HiRBF Paradigms

3.1 The Data Set

We used three different datasets in our forecast performance analysis. The data used are daily forex exchange rates obtained from the Pacific Exchange Rate Service [12], provided by Professor Werner Antweiler, University of British Columbia, Vancouver, Canada. The data comprises of the US dollar exchange rate against Euros, Great Britain Pound (GBP) and Japanese Yen (JPY). We used the daily data from 1 January 2000 to 31 October 2002 as training data set, and the data from 1 November 2002 to 31 December 2002 as evaluation test set or out-of-sample datasets (partial data sets excluding holidays), which are used to evaluate the good or bad performance of the predictions, based on evaluation measurements.

The forecasting evaluation criteria used is the normalized mean squared error (NMSE),

$$NMSE = \frac{\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y}_t)^2} = \frac{1}{\sigma^2}\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2, \tag{3}$$

where y_t and $\hat{y}_t$ are the actual and predicted values, σ^2 is the estimated variance of the data and $\bar{y}_t$ the mean.

3.2 Feature/Input Selection with HiRBF

It is often a difficult task to select important variables for a forecasting or classification problem, especially when the feature space is large. A fully connected NN classifier usually cannot do this. In the perspective of HiRBF framework, the nature of model construction procedure allows the HiRBF to identify important input features in building a forecasting model that is computationally efficient and effective. The mechanisms of input selection in the HiRBF constructing procedure are as follows. (1) Initially the input variables are selected to formulate the HiRBF model with same probabilities; (2) The variables which have more contribution to the objective function will be enhanced and have high opportunity to survive in the next generation by a evolutionary procedure; (3) The evolutionary operators provide a input selection method by which the HiRBF should select appropriate variables automatically.

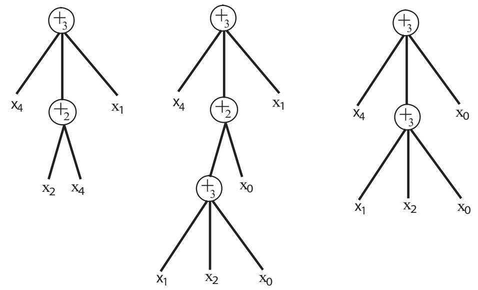

Fig. 3. The evolved HRBF trees for forecasting euros (left), British pounds (middle) and Japanese yen (right).

Table 1. Forecast performance evaluation for the three exchange rates (NMSE for testing)

Exchange rate	Euros	British Pounds	Japanese Yen
MLFN [13]	0.5534	0.2137	0.2737
ASNN [13]	0.1254	0.0896	0.1328
RBF-NN	0.1130	0.0852	0.1182
HRBF-NN (This paper)	0.0240	0.0212	0.0095

3.3 Experimental Results

For simulation, the five-day-ahead data sets are prepared for constructing HiRBF models. A HiRBF model was constructed using the training data and then the model was used on the test data set. The instruction sets used to create an optimal HiRBF forecaster is $S = F \bigcup T = \{+_2, +_3\} \bigcup \{x_1, x_2, x_3, x_4, x_5\}$. Where $x_i (i = 1, 2, 3, 4, 5)$ denotes the 5 input variables of the forecasting model.

The optimal HiRBF models evolved for three major internationally traded currencies: British pounds, euros and Japanese yen are shown in Figure 3. It should be noted that the important features for constructing the HiRBF models were formulated in accordance with the procedure mentioned in the previous section.

For comparison purpose, three single-stage RBF networks are also employed with structure of {5-10-1} for forecasting three major internationally traded currencies. The forecast performances of a traditional multi-layer feed-forward network (MLFN) model and an adaptive smoothing neural network (ASNN) model are also shown in Table 1. The actual daily exchange rates and the predicted ones for three major internationally traded currencies are shown in Figure 4. From Tables 1, it is observed that the proposed HiRBF forecast models are better than the considered neural networks models for three major internationally traded currencies.

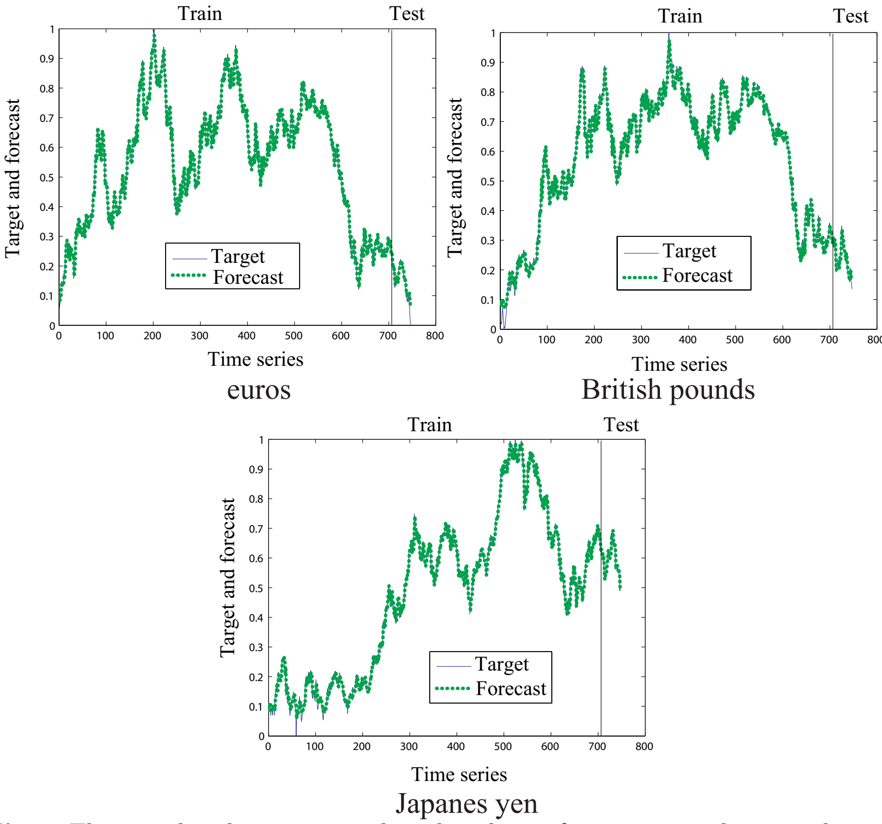

Fig. 4. The actual exchange rate and predicted ones for training and testing data set.

4 Conclusions

In this paper, we presented a HiRBF model for forecasting three major international currency exchange rates. We have demonstrated that the evolved HiRBF forecasting model may provide better forecasts than the traditional MLFN forecasting model, the ASNN forecasting model and a traditional single RBF network. The comparative evaluation is based on a statistical measures (NMSE). Our experimental analyses reveal that the NMSE for three currencies using the HiRBF model are significantly better than those using the MLFN model, the ASNN model and the RBF model. This implies that the proposed HiRBF model can be used as a feasible solution for exchange rate forecasting.

Acknowledgment

This research was partially supported the Natural Science Foundation of China under contract number 60573065, and The Provincial Science and Technology Development Program of Shandong under contract number SDSP2004-0720-03.

References

1. P. Theodossiou, "The stochastic properties of major Canadian exchange rates". The Financial Review 29(2):193-221, 1994.
2. M. K. P. So, K. Lam and W. K. Li., "Forecasting exchange rate volatility using autoregressive random variance model", Applied Financial Economics, 9:583-591, 1999.
3. D. A. Hsieh. "Modeling heteroscedasticity in daily foreign-exchange rates", Journal of Business and Economic Statistics, 7:307C317, 1989.
4. D. Chappel, J. Padmore, P. Mistry and C. Ellis. "A threshold model for French franc/Deutsch mark exchange rate", Journal of Forecasting, 15:155-164, 1996.
5. A. N. Refenes,M. Azema-Barac, L. Chen and S. A. Karoussos. "Currency exchange rate prediction and neural network design strategies", Neural Computing and Application, 1:46-58, 1993.
6. L.Yu, S. Wang, and K. K. Lai. "Adaptive Smoothing Neural Networks in Foreign Exchange Rate Forecasting", V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3516, pp. 523-530, 2005.
7. W. Wang, K.K.Lai, Y.Nakamori, S.Wang. "Forecasting Foreign Exchange Rates with Artificial Neural Networks: A Review". International Journal of Information Technology & Decision Making, 3(1):145-165, 2004.
8. Y. Chen, B. Yang and J. Dong, "Nonlinear System Modeling via Optimal Design of Neural Trees", International Journal of Neural Systems, 14(2):125-137, 2004.
9. Y. Chen, B. Yang, J. Dong, and A. Abraham, "Time-series Forecasting using Flexible Neural Tree Model", Information Science, 174(3-4):219-235, 2005.
10. J.T. Yao, C.L. Tan. "A case study on using neural networks to perform technical forecasting of forex", Neurocomputing, 34:79-98, 2000.
11. K. Sastry and D. E. Goldberg. "Probabilistic model building and competent genetic programming", In R. L. Riolo and B. Worzel, editors, Genetic Programming Theory and Practise, chapter 13, pp. 205-220. Kluwer, 2003.
12. http://fx.sauder.ubc.ca/
13. L. Yu, S. Wang, and K. K. Lai, "Adaptive Smoothing Neural Networks in Foreign Exchange Rate Forecasting", V.S. Sunderam et al. (Eds.): ICCS 2005, LNCS 3516, 523-530, 2005.
14. A. Abraham, "Analysis of Hybrid Soft and Hard Computing Techniques for Forex Monitoring Systems", IEEE International Conference on Fuzzy Systems (IEEE FUZZ'02), IEEE Press pp. 1616 -1622, 2002.
15. L. Kirkpatrick et al., "Optimization by simulated annealing", Science, 220:671-680, 1983.
16. E.K. Burke et al., "A new local search approach with execution time as an input parameter", Technical Report No. NOTTCS-TR-2002-3, School of Computer Science and Information Technology, University of Nottingham, (2002).
17. N. A. Mat Isa, Mashor, M. Y., and Othman, N. H., "Diagnosis of Cervical Cancer using Hierarchical Radial Basis Function (HiRBF) Network", In Sazali Yaacob, R. Nagarajan, Ali Chekima (Eds.), Proc. of the Int. Conf. on Artificial Intelligence in Engineering and Technology, pp. 458-463, 2002.
18. S. Ferrari, I. Frosio, V. Piuri, and N. Alberto Borghese, "Automatic Multiscale Meshing Through HRBF Networks", IEEE Trans. on Instrumentation and Measurment, 54(4):1463-1470, 2005.
19. Z. Ahmad, J. Zhang, "Bayesian selective combination of multiple neural networks for improving long-range predictions in nonlinear process modelling", Neural Comput & Applic., 14:78C87, 2005.