Optimal Signal Multi-Resolution by Genetic Algorithms to Support Artificial Neural Network Models for Financial Forecasting

Taeksoo Shin, Ingoo Han
Graduate School of Management, Korea Advanced Institute of Science and Technology
tsshin@msd.kaist.ac.kr; ighan@kgsm.kaist.ac.kr

Abstract

Detecting the features of significant patterns from their own historical data is crucial in getting an optimal performance especially in time series forecasting. Wavelet analysis which processes information effectively at different scales can be very useful in accomplishing this.

One of the most critical issues to be solved in the application of the wavelet analysis is to choose the correct filter types and the filter parameters. If the threshold is too small or too large, the wavelet shrinkage estimator will tend to overfit or underfit the data. The threshold is often selected arbitrarily or by adopting certain theoretical or statistical criteria. Recently, new and versatile techniques have been introduced to solve that problem.

In this study, we propose an integrated thresholding design of the optimal wavelet transform by genetic algorithms (GAs) to represent a significant signal most suitable in neural network models especially for use in chaotic financial markets. The results show that a hybrid system using GAs has better performance than any other method.

1. Introduction

Detecting the features of significant patterns from their own historical data is crucial in getting optimal performance, especially in time series forecasting. The methods used for time series analysis are conventionally and heavily based on the concepts of stationarity and linearity.

Recently, there has been a renewal of interest in linear expansions of signals, particularly using wavelets and some of their generalization (Daubechies [15], Mallat [30], Rioul and Vetterli [43]). A new data filtering method (or multi-signal decomposition) namely, wavelet analysis is considered more useful for handling the time-series that contain strong quasi-cyclical components than other methods. Wavelet analysis theoretically presents much more clear local information according to different time intervals from the filtered data.

In this study, we suggest a new model architecture of the neural networks supported by a wavelet analysis as multi-signal decomposition to detect the features of significant patterns. We apply the hybrid model architecture to forecasting one day ahead Korean Won / U.S. Dollar currency market as a case study. A strategy is devised using wavelet transform to construct a filter that is significantly matched to the frequency of the time-series within the combined model.

From our experimental results we analyze the effects of several wavelet filtering (i.e. thresholding) criteria to support the neural network learning optimization at present. We also suggest a new general optimal filtering criterion of multi-signal decomposition methods from our experimental learning and validation results of the neural networks. That is, we propose a new extended neural network model which is four-layered neural network architecture having a multi-scale extraction layer before arriving at the input layer. The model learning is supported by genetic algorithms (GAs) and a hill climbing algorithm (HC). Through their hybrid learning we tried to solve efficiently the present threshold problems about the optimal filter design in extracting the significant information from the original data.

The rest of this paper is organized as follows. In the next section, we briefly review fractal structure of financial markets. The third section introduces a multiresolution approach to wavelet and then wavelet transformation methodology about optimal decomposition from the original time series. The fourth section suggests a new methodology for hybrid system using GAs and shows the experimental results in the fifth section. The final section contains concluding remarks.

2. Fractal Structure of Chaotic Financial Markets

Fractional Brownian Motion (fBm) has long been considered a plausible model for financial markets. A fractal structure of the market, indicating the presence of correlations across time, hints at the possibility of some predictability. Recent advances in time-frequency localized transforms by the applied mathematics and electrical engineering communities provide us with new methods for the analysis of this type of process. In fact, it has been proven by Wornell [56], [57] that the wavelet

0-7695-0446-9/99 $10.00 © 1999 IEEE

transform with a Daubechies basis is an optimal transform for fBm processes.

The structure of the market also has the feature of market heterogeneity. Market heterogeneity suggests that the different intentions among market participants result in sensitivity by the market to several different time-scales. Different types of traders view the market with different time resolutions, for example, hourly, daily, weekly, and so on. Short-term traders evaluate the market at high frequency and have a short memory. Small movements in the exchange rate mean a great meal to the short-term trader. The long-term trader evaluates lower frequency data with a much longer memory of past data. He is only interested in large movements in the price. These different types of traders create the multiscale dynamics of the time series.

Throughout our study, we focus on this concept. This reasoning reinforces that a multiresolution embedding using a wavelet analysis is very useful for discovering whether some time-scales are more predictable than others.

3. Multiresolution Approach to Wavelets

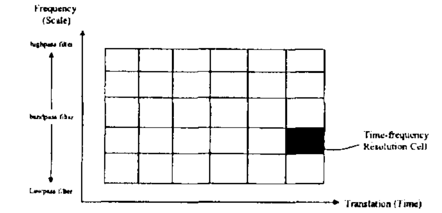
The wavelet analysis is a robust tool that may be used to obtain qualitative information for highly nonstationary time series - specifically, it may be used to detect a small-amplitude harmonic forcing term even when the dynamics are chaotic and even for short total times. (Permann and Hamilton [40])

For present purposes, we focus on the multi-resolution structure of curves or spectra. We intuitively view high frequency noise differently from broad, low frequency components due to e.g. baseline effects.

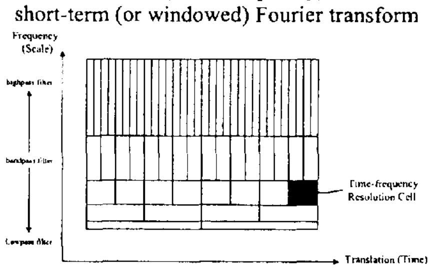
By employing the multi-resolution view, we can build and dismantle curves according to resolution level, so the wavelet functions are constructed to focus on different resolution details in the signal at different positions. This feature is possible because of the special structure of the wavelet basis functions.

The coverage of the time-frequency plane for the wavelet analysis is shown in Figure 1(b). Even though the windowed Fourier transform (WFT) including the discrete Fourier transform (DFT) usually displays the coverage as shown in Figure 1(a), it has their own limitation compared to the wavelet transform. For example, the DFT spreads frequency information over all time and, thus, the loss of frequency characteristics of a time series in the time domain. The transform process is said to be non-local in the time domain. We can partially compensate for this lack of localization by applying either the WFT or the short-time Fourier transform (STFT) to introduce time dependency. But, the WFT filters are evenly spaced in the frequency domain.

Figure 2 shows an example of multi-resolution of a DWT filter, i.e. Daubechies with order 4 using daily Korean won / US dollar returns.



(a) Two dimensional (Time-frequency) resolution of a short-term (or windowed) Fourier transform



(b) Two dimensional (Time-frequency) resolution of a discrete wavelet transform

Figure 1. Two dimensional (Time-frequency) resolution

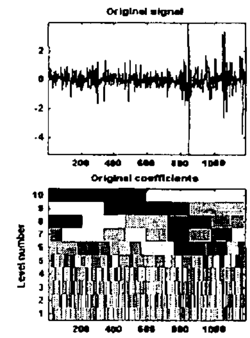


Figure 2. An example of two dimensional (Time-frequency) resolution of a discrete wavelet transform (Daubechies with order 4) using the daily Korean Won/US Dollar returns

3.1. Discrete Wavelet Transform (DWT) and Wavelet Packet Transform (WPT)

A DWT is expressed as a pyramid or tree algorithm (Mallat [30]). In the pyramid algorithm the detail branches are not used for further calculations; only the approximations at each level of resolution are treated to yield approximation and detail obtained at level m+1. But, application of the transform to both the detail and the approximation coefficients results in an expansion of the structure of the wavelet transform tree algorithm to the full binary tree (Coifman and Wickerhauer [11], Coifman et al. [12]). It is called a wavelet packet transform (WPT). This is a more general transform than the DWT. The main difference is that, while in the DWT the detail coefficients are kept and the approximation coefficients are further analyzed at each step, in the WPT both the approximation signal and the detail signal are analyzed at each step. This results in redundant information, as each level of the transform retains n samples.

Therefore, the WPT produces an arbitrary frequency split, which can be adapted to the signal. While wavelet packets create arbitrary binary slicing of frequencies (with associated time resolution), they do not change over time. Often a signal is first arbitrarily segmented, and then, the wavelet packet decomposition is performed on each segment in an independent manner.

There exist simple and efficient algorithms for both wavelet packet decomposition and optimal decomposition selection. We can then produce adaptive filtering algorithms with direct applications in optimal signals.

3.2. Highpass, Lowpass, and Bandpass Filters

The subspaces created by the wavelet transform roughly correspond to the frequency subbands partitioning the frequency bandwidth of the data set. These subspaces then form a disjoint cover of the frequency space of the original data set. In other words, the subspaces have no elements in common and the union of the frequency subbands spans the frequency span of the original data set.

Any set of subspaces which are a disjoint cover of the original data set is understood on an orthonormal basis. The wavelet transform basis is then but one of a family of orthonormal bases with different subband intervals.

According to the frequency-response characteristics, the frequency subbands or subspaces are again categorized into the four basic filter types, i.e. lowpass, highpass, bandpass, and bandstop filter.

The lowpass, highpass, and bandpass filters predominantly used in this study are characterized by two parameters, i.e. frequency and width. These parameters are difficult to specify. There is no simple calculation to provide the correct value.

3.3. The Optimal Multi-Resolution of Time Series

Based on the organization of the wavelet packet library, it is natural to count the decompositions issued from a given orthogonal wavelet. As a result, a signal of length $N = 2^p$ can be expanded out at most 2N different ways, the number of binary subtrees of a complete binary subtree of depth p. As this number may be very large, and since explicit enumeration is generally unmanageable, it is interesting to find an optimal decomposition with respect to a convenient criterion, computable by an efficient algorithm. We search for a minimum of the criterion.

For this purpose, we try to suggest a new criterion of choosing the optimal decomposed sub-series from original series by discrete wavelet transforms in the following research model architecture.

4. Research Model Architecture

In general, for a one-dimensional discrete-time signal, the high frequencies influence the details of the filter levels, while the low frequencies influence the deepest levels and the associated approximations.

The original signal can be expressed as an additive combination of the wavelet coefficients at the different resolution levels.

In this section, we suggest our research framework and a new hybrid time series forecasting model architecture as shown in Figure 3 and 4. Our research framework consists of 4 phases. The first phase decomposes the financial time series into different decomposed series components using a discrete wavelet transform. In the second phase, we extract the refined highpass, lowpass, and bandpass filters from the decomposed time series, which is based on the feedback from the next phase. Figure 4 shows our hybrid neural network model as extended neural network architecture in comparison to prior models. We extract a significant scale component generation automatically from the original data within our model. This function achieves a multi-scale extraction layer of our model. The third phase uses a neuro-genetic approach to learn neural network models with GAs and then the fifth phase includes additional weight learning process by HC algorithms to reinforce a generalization of learning parameters or weights generated in the previous phase. For this purpose, the fourth phase is compared with the third phase in terms of the forecasting accuracy.

The resolution of time series can be adjusted to local parameters to detect its present features including promising features in close time areas with more sensitivity. Using multiple scales of resolution, the time series forecasting can be refined in areas. Feature based segmentation techniques detect local features (such as transitions, lines, curves, in general referred to as edges) based on values of appropriate local operators.

To improve local prediction, a signal parameter such as the refined lowpass filters, highpass filters, and bandpass filters is proposed to control multiple scales of resolution within our research framework. For example, each of the 10 scales was then multiplied by a weighting factor (0-1) and the weighted transform inverted back to a time-series when our time series is decomposed into 10 scales (Figure 5).

Among the various cost measures that one can choose for finding adaptive time-frequency decompositions, we select an evolutionary data driven criteria. The benefits of this are as follows. Since our model uses a final model performance measure within a unified model framework as wavelet thresholding cost measure, the performance measure can solve the combining problems to detect the optimal signal decomposition criteria relevant to a neural network model and also the generalization problems of our hybrid model.

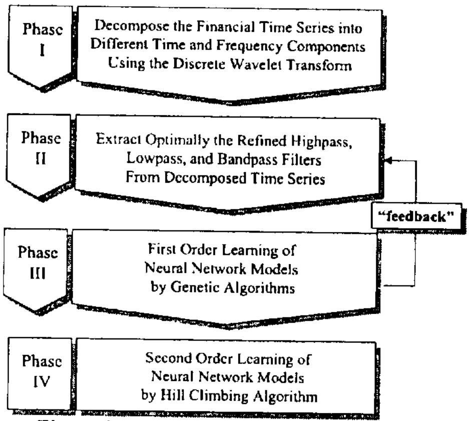


Figure 3. Proposed research framework

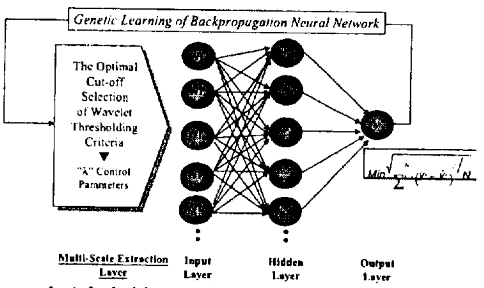


Figure 4. A hybrid neural network model architecture

$$Min\left(\sqrt{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2/N}\right) \quad (3)$$
$$\hat{y}_i = f(IWT_i)$$
$$IWT_i = \sum_{j=\lambda_1}^{\lambda_2} DS_i(j)$$
$$s.t.\ X_i = \sum_{j=1}^{N} DS_i(j)$$
$$1 \le \lambda_1 \le \lambda_2 \le k$$

where

y_i = Actual output of *ith* case of population at (t+1)*th* day,

$\hat{y}_i$ = Neural network output of *ith* case of population at (t+1)*th* day,

f(·) = Neural network model,

IWT_i = The *ith* case of refined wavelet filtered inputs of population at (t)*th* day,

X_i = The *ith* input case of population at (t)*th* day,

$DS_i(j)$ = The *jth* automatically decomposed time series by Daubechies wavelet transform of *ith* case of population at (t)*th* day,

N = The population size,

k = The maximum length of decomposition levels of pyramid or tree algorithms,

λ_1 = The cut-off level of lowpass filter on condition of (λ_2 = k),

λ_2 = The cut-off level of highpass filter on condition of (λ_1 = 1) except that both λ_1 and λ_2 are the cut-off levels of bandpass filter on condition of ($1 \le \lambda_1 \le \lambda_2 \le k$).

4.1. Phase I: Decomposing The Financial Time Series Using DWT

A scalogram (Rioul and Flandrin [42]) is defined as wavelet periodogram referring to the absolute value of the wavelet coefficients at each scale. The scalogram usually is plotted logarithmically as a function of both the scale and location indices. Inspection of the scalogram (or of the wavelet coefficients themselves) is useful when one needs to view frequency/scale and location information at the same time. In the same manner as the periodogram produces an ANOVA decomposition of energy of a signal to different Fourier frequencies, the scalogram decomposes the energy to level components (See Table 1).

Table 1. Scalogram of daily Korean Won/US Dollar returns ($\ln X_t - \ln X_{t-1}$)

Decomp. Series (DS)	*Frequency*	*Energy (power)*
DS1	1-4	762.3629045
DS2	5-8	272.2633141
DS3	9-16	65.28630562
DS4	17-32	37.19868193
DS5	33-64	16.77289267
DS6	65-128	6.330477868
DS7	129-256	2.911816767
DS8	257-512	2.363012764
DS9	513-1024	1.593836382
DS10	1025-2048	0.91605887

The scalogram of the discrete wavelet transform of a time series is the key tool used to decompose the series into cycles of different frequencies. However, it is mostly difficult for model experts to extract the multi-cyclic structure from the original data by analyzing the distribution shape of scalogram.

From Figure 5 we can extract lowpass filter, highpass filter, and bandpass filter from a decomposed band series (i.e. Band 1-10). For example, Band 1 corresponds to the highest frequency component in the data. These components indicate signals with very short periods. In addition, we can effectively extract a bandpass filter to the data by eliminating the highest and lowest bands from the total bands.

Therefore, we combine decomposed band series (Band1-10) by the DWT into generating 3 separate bands, i.e.

refined lowpass, highpass, and bandpass filters to extract the optimally or near optimally refined wavelet filters in the following phase II.

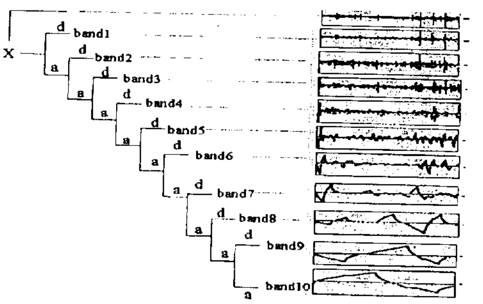


Figure 5. Pyramid or tree algorithm using daily Korean Won / US Dollar returns [X: original series, a: approximation components, d: detailed components, band 1 (the highest highpass filter) → band10 (the lowest lowpass filter)].

4.2. Phase II: Extracting Optimally The Refined Wavelet Filters (i.e. Highpass, Lowpass, and Bandpass Filters)

Since the scales of the WT may be viewed as a filter bank, the degree to which any single scale reflects the probability distribution of the frequency characteristics of the time series can be calculated (i.e. a measure of relevance). The relevance of each scale to all the time series can then be translated into a weighting factor.

A measure of the cumulative relevance of each scale over all the examples was calculated in the following way. The WT for each example used to train the neural network is calculated. Some scales are weighted exactly 1 (index scale), while all other scales of the transform are set to 0, and the transform is inverted back into a time series.

Various techniques for optimal bandwidth selection have been studied in the following studies (Brillinger [5], Jenkins and Watts [26], Moulin et al. [36], Priestley [41], Wahba [50], Wahba and Wold [51]). They produce estimates that have a good overall bias variance tradeoff. However, the bias versus variance tradeoff is generally not optimal locally.

4.3. Phase III: The First Order Learning of Neural Network Model by GAs

In this study, the basic model we experiment with is Backpropagation neural network (BPN) model (Rumelhart et al.[44]) which has a parsimonious 4 input nodes, 4 hidden nodes and 1 output node with one type of wavelet filters, i.e. highpass, lowpass or bandpass filters within the network structure. The other model we experiment with is BPN model which has 8 input nodes, 8 hidden nodes and 1 output node with two types of filters, i.e. both highpass and lowpass filters or has 12 input nodes, 12 hidden nodes and 1 output node with three types of filters, i.e. highpass, lowpass and bandpass filters. Each filter as an input node also consists of its own 4 daily delayed inputs.

GAs are also used to search the weight space without use of any gradient information (Whitley and Hanson [55], Montana and Davis [34]). For example the fitness could be given by the value of the cost function for that set of weights. Starting with a random population of such strings, successive generations are constructed using genetic operators such as mutation and crossover to construct new strings out of old ones, with some form of survival of the fittest; fitter strings are more likely to survive and to participate in mating (crossover) operations.

In this study, GAs are basically used to automatically determine both the wavelet thresholding cut-off parameters and the learning parameters of neural networks. The wavelet thresholding parameters are adjusted to optimize the performance of the financial forecasting over the entire samples (i.e. training samples).

For this use with the GA, every weight in the neural networks has been coded by real number with values in a limited range, [-4, 4]. In addition, our suggested multi-scale extraction layer's weights, i.e. the cut-off levels (λ_1, λ_2) of the wavelet filters with the range from 1 to λ_1, λ_1 to λ_2 (s.t. $1 \leq \lambda_1 \leq \lambda_2 \leq k$; k = the maximum length of resolution levels by wavelet transformation) and λ_2 to k are added to the present neural network model by being connected with the input layer.

In the current study we use a population size of 50 and the same GA adjusted parameters are maintained over the entire study in order to estimate the average performance of the NN models for different learning methods. The crossover rate ranges 0.5 - 0.8 and the mutation rate ranges 0.01- 0.06 for our experiment. As a stopping condition, we use 5,000 trials.

4.4. Phase IV: The Second Order Learning of Neural Network Model by Hill Climbing Algorithm

In the previous phase we used GAs to learn neural networks, but one problem with GAs is their inefficiency in fine-tuned local search, thus the scalability of these methods are in question (Yao [58]).

To solve this learning problem, Kitano [27] presents a method that combines GAs with HC. He does this by using the GA to determine the starting weights for a network, which is then refined by HC.

According to this learning strategy, we tried to solve the limitation of learning by GAs in Phase III with the second order learning, i.e. HC learning of neural networks in this phase.

5. Experimental Results

In this section, we evaluate our framework using a case of the daily Korean Won / U.S. Dollar exchange rates which are transformed to the returns using the logarithm through standardization from January 10, 1990 to June 25, 1997. That is, the returns are defined as the logarithm of today's exchange rate divided by the logarithm of yesterday's exchange rate. The learning phase involved observations from January 10, 1990 to August 4, 1995, while the testing phase ran from August 7, 1995 to June 25, 1997. We transform the daily returns into the decomposed series such as an approximation part and a detail part by Daubechies wavelet transform with 4 coefficients (DAUB4) for neural network forecasting models.

The experimental results of our hybrid neural network architecture are showed in Table 2. First, our hybrid system has better performance than random walks and NN by HC or GA with original signals as inputs.

Secondly, we compare two learning methods within the hybrid system. Our hybrid model is trained using two learning methods, i.e. only a GA method and a combined learning method by GA and HC (GA-HC). After that, they are compared with other models in terms of the performance of wavelet thresholding algorithms. As shown in Table 2, a combined GA-HC method gets better performance than one GA method.

Thirdly, the performance is also different according to filter type. The model using highpass, lowpass, and bandpass filters at once demonstrates better performance than models using partial filters. However, lowpass and bandpass filters have the same cut-off range result. This result indicates that bandpass filters rarely impact on our model performance applied to daily Korean won / US dollar market.

Finally, we compare our model performance with the performance of benchmark models, i.e. the models using prior representative thresholding methods to evaluate our hybrid forecasting system. The results show that our hybrid model is better than any other model in terms of forecasting performance (Table 3). In this experiment, we use a few benchmark models to compare our model's performance as follows. Namely, we use well known three wavelet thresholding algorithms, i.e. best basis selection (Daubechies [14], Coifman and Wickerhauser [11], Mallat and Zhang [31], Chen [6], Chen and Donoho [7], Chen et al. [8], Donoho [17]), cross-validation (Nason [38], [39], Jensen and Bultheel [25]), and best level tree (Coifman et al. [9]) techniques in the literature. Table 3 shows that GA method has significantly better performance than any other wavelet thresholding Algorithm.

Table 2. The comparison of the BPN model performance using test samples

Filter Types	*Cut-off Range* (λ_1, λ_2)	*Learning Methods*	*BPN*[g] *Structure (I-H-O)*[h]	*Performance (RMSE)*[i]
-	-	RW[c]		2.939007
-	-	GA[d]	(4-4-1)	1.780642
-	-	HC[e]	(4-4-1)	1.754525
Highpass	(1-2)	GA	(4-4-1)	1.629141
Lowpass	(3-10)	GA	(4-4-1)	1.726126
Bandpass	(2-4)	GA	(4-4-1)	1.750383
Combined[b]	(1-5, 2-10, 1-5)[a]	GA	(12-12-1)	**1.343301**
Highpass	(1-2)	GA-HC[f]	(4-4-1)	1.516126
Lowpass	(3-10)	GA-HC	(4-4-1)	1.721580
Bandpass	(2-4)	GA-HC	(4-4-1)	1.713277
Combined[b]	(1-5, 2-10, 1-5)[a]	GA-HC	(12-12-1)	**1.119327**

a: Highpass (1-5), Lowpass (2-10), Bandpass (1-5), b: Highpass+Lowpass+Bandpass filters, c: Random walks, d: Genetic algorithms, e: Hill climbing algorithms, f: Genetic algorithms + Hill climbing algorithms, g: Backpropagation neural network, h: (I: Input nodes , H: Hidden nodes, O: Output nodes), i: Root mean squared error.

Table 3. The BPN performance comparison between different wavelet filtering criteria using test samples

Wavelet Threshold Algorithms	*Filter Types*	*Learning Methods*	*BPN Structure (I-H-O)*[i]	*Performance (RMSE)*[j]
Best Basis[a]	LP[e]&HP[f]	HC[h]	(8-8-1)	1.74329
Cross Validation[b]	LP&HP	HC	(8-8-1)	1.676247
Best Level[c]	LP&HP	HC	(8-8-1)	1.746597
GA[d]	LP&HP&BP[g]	GA-HC	(12-12-1)	**1.119327**

a: Coifman and Wickerhauser (1992), b: Nason (1994), c: (Coifman et al., 1994), d: Genetic algorithms, e: Lowpass filter, f: Highpass filter, g: Bandpass filter, h: Hill climbing algorithms, i: (I: Input nodes , H: Hidden nodes, O: Output nodes), j: Root mean squared error.

6. Concluding Remarks

We have described a new framework for modeling and analyzing signals at multiple scales in financial forecasting.

In conclusion, we illustrated a decision support to build a hybrid forecasting system by using GAs throughout our study. The decision support process is summarized as follows. First, a multiple scale resolution of financial time series is implemented easily by discrete wavelet transform techniques. Once the financial time series has been segmented into areas with relative homogeneous value levels, i.e. filtered band series, the transformed information is evaluated through a refining process of the band series. That is, the desired multi-scale input structure in neural network models is optimally or near optimally extracted by genetic algorithms, so the final input structure consists of the refined highpass and lowpass filtered inputs in our experiment. The experimental results showed the enhanced filtering or signal multi-resolution power of wavelet analysis to the performance of the neural network models. It also means that our wavelet thresholding algorithm by GAs is better than other thresholding algorithms in increasing forecasting performance.

7. References

[1] B.K. Alsberg, A.M. Woodward, and D.B. Kell, "An Introduction to Wavelet Transforms for Chemometricians: A Time-Frequency Approach," *Chemometrics and Intelligent Laboratory Systems* 37, 1997, pp.215-239.

[2] M.A. Ariño and B. Vidakovic, "On Wavelet Scalograms and Their Applications in Economic Time Series," *Discussion paper 95-21*, ISDS, Duke University, 1995.

[3] A. Aussem, J. Compbell and F. Murtagh, "Wavelet-Based Feature Extraction and Decomposition Strategies for Financial Forecasting," *Journal of Computational Intelligence in Finance*, March/April, 1998, pp.5-12.

[4] Breiman, L., J. Friedman, R. Olshen, and C.J. Stone, *Classification and Regression Trees*, Belmont, CA: Wadsworth, 1983.

[5] Brillinger, D.R., *Time Series: Data Analysis and Theory*. Holt, New York, 1975.

[6] Chen, S.S. *Basis Pursuit*, Ph.D. Thesis, Department of Statistics, Stanford University, Stanford, CA, Nov. 1995.

[7] S.S. Chen and D.L. Donoho, "Basis Pursuit," *Technical Report*, Stanford University, Statistics Department, 1995.

[8] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Journal on Scientific Computing*, Vol. 20, No. 1, 1998, pp. 33-61.

[9] R.R. Coifman, Y. Meyer, S. Quake, and M.V. Wickerhauser, "Signal Processing and Compression with Wavelet Packets," *in Wavelets and Their Applications*, J.S. Byrnes, J.L. Byrnes, K.A. Hargreaves, and K. Berry (eds.), 1994, pp.363-379

[10] Coifman, R.R., Y. Meyer, and M.V. Wickerhauser, *Progress in Wavelet Analysis and Applications*, in Y. Meyer and S. Roques (eds.), Editions Frontieres, France, 1993, pp.77-93.

[11] R.R. Coifman and M.V. Wickerhauser, "Entropy Based Methods for Best Basis Selection," *IEEE Transactions on Information Theory* 38, No. 2, 1992, pp.719-746.

[12] R.R. Coifman and M.V. Wickerhauser, "Wavelets and Adapted Waveform Analysis. A Toolkit for Signal Processing and Numerical Analysis," *In Processing of Symposia in Applied Mathematics* 47, 1993, pp.119-145.

[13] G. Davis, S. Mallat, and Z. Zhang, "Adaptive Time-Frequency Decompositions with Matching Pursuit," *Wavelet Applications, Proc. SPIE*, 1994, pp.402-409.

[14] I. Daubechies, "Time-Frequency Localization Operators: A Geometric Phase Space Approach," *IEEETIT* 34, 1988, pp.605-612.

[15] Daubechies, I., Ten Lectures on Wavelets, *SIAM*, Philadelphia, PA, 1992.

[16] K. Dohan and P.H. Whifield, "Identification and Characterization of Water Quality Transients Using Wavelet Analysis. I. Wavelet Analysis Methodology," *Wat. Sci. Tech.* Vol. 36, No. 5, 1997, pp.325-335.

[17] D.L. Donoho, "Denoising by Soft-Thresholding," *IEEE Transactions on Information Theory*, vol.41, 1995, pp.6130-6270.

[18] D.L. Donoho and I.M. Johnstone, "Ideal Spatial Adaption via Wavelet Shrinkage," *Bimetrika*, vol. 81, 1994, pp.425-455.

[19] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet Shrinkage: Asymptopia?," *Journal of the Royal Statistical Society*, Series B, vol. 57, 1995, pp.301-369.

[20] J.L. Elman, "Finding Structure in Time," *Cognitive Science* 14, 1990, pp.179-211.

[21] D.J. Farmer and J.J. Sidorowich, "Predicting Chaotic Time Series," *Physical Review Letters*, 59, 1987, pp.845-848.

[22] A.B. Geva, "ScaleNet-Muliscale Neural-Network Architecture for Time Series Prediction," *IEEE Transaction on Neural Networks*, Vol. 9, No.5, Sep. 1998.

[23] Goldberg D.E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Mass. Addison-Wesley, 1989.

[24] S.A. Greenblatt, "Wavelets in Econometrics: An Application to Outlier Testing," *Computational Economic Systems: Models, Methods, and Econometrics* (Dordrecht), Manfred Gilli (ed.), *Advances in Computational Economics*, Vol. 5, Kluwer Academic Publishers, 1995, pp.139-160.

[25] M. Jensen and A. Bultheel, "Wavelet Thresholding Using Generalized Cross Validation," *The International Workshop on Wavelets and Statistics*, Duke University, NC, U.S.A., 12-13 October 1997.

[26] Jenkins, G.M. and D.G. Watts, *Spectral Analysis and its Applications*. Holden-Day, San Francisco, 1968.

[27] H. Kitano, "Empirical Studies on The Speed of Convergence of Neural Network Training Using Genetic Algorithms," *In Proceedings of The Eighth National Conference on Artificial Intelligence*, boston. MA. AAAI/MIT Press, 1990, pp.789-795.

[28] C.M. Kuan and K. Hornik, "Convergence of Learning Algorithms with Constant Learning Rates," *IEEE Transactions on Neural Networks*, vol. 2, No.5, 1991, pp. 484-489.

[29] S. Lin, W. Punch III, and E.D. Goodman, "A Hybrid Model

Utilizing Genetic Algorithms and Hopfield Neural Networks for Function Optimization," *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Francisco, 1995.

[30] A. Mallat, "Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1989, pp.674-693.

[31] S.G. Mallat and Z. Zhang, "Matching Pursuit with Time-Frequency Dictionaries," *Technical Report* 619, Courant Institute of Mathematical Sciences, (1993)

[32] Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programming*, Second Extended Edition, Spring-Verlag, 1994.

[33] C.R. Mittermayr, S.G. Nikolov, H. Hutter, and M. Grasserbauer, "Wavelet Denoising of Gaussian Peaks: A Comparative Study," *Chemometrics and Intelligent Laboratory Systems* 34, 1996, pp. 187-202.

[34] D. Montana and C. Davis, "Training Feedforward Neural Networks Using Genetic Algorithms," *Technical Report*, BBN Systems and Technologies Inc., Cambridge (MA), 1989.

[35] P.A. Morettin, "Wavelets in Statistics," *Third International Conference on Statistical Data Analysis Based on The L_1-Norm and Related Methods*, Neuchâtel, Switzerland, Aug. 11-15, 1997.

[36] P. Moulin, J. A. O'Sullivan, and D. L. Snyder, "A Method of Sieves for Multiresolution Spectrum Estimation and Radar Imaging," *IEEE Transactions on Information Theory*, pp. 801-813, March 1992.

[37] U.A. Müller, M. M. Dacorogna, R. Davé, O.V. Pictet, R.B. Olsen, and J.R. Ward, "Volatilities of Different Time Resolutions - Analyzing The Dynamics of Market Components," *The XXXIXth International Conference of the Applied Econometrics Association (AEA), Real Time Econometrics —Submonthly Time Series*, 14-15 Oct 1993.

[38] G.P. Nason, "Wavelet Regression by Cross-validation," *Technical Report* 447, Department of Statistics, Stanford University, April 1994.

[39] G.P. Nason, "Wavelet Shrinkage Using Cross-Validation," *Journal of the Royal Statistical Society*, Series B, vol. 58, 1996, pp.463-479.

[40] D. Permann and I. Hamiltion, "Self-similar and Erratic Transient Dynamics for the Linearly Damped Simple Pendulum," *American Journal of Physics*, Vol.60, No.5, 1992, pp.442-450.

[41] Priestley, M.B., *Spectral Analysis and Time Series*, Academic Press, London and San Diego, 1981.

[42] O. Rioul and P. Flandrin, "Time-Scale Energy Distributions: A General Class Extending Wavelet Transforms," *IEEE Transactions on Signal Processing*, 40, 1992, pp.1746-1757.

[43] O. Rioul and M. Vetterli, "Wavelet and Signal Processing," *IEEE Signal Processing Magazine*, October, 1991, pp.14-35.

[44] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning Internal Representations by Error Propagation," *In Parallel Distributed Processing: Exploration in the Microstructure of Cognition*, (eds.) D.E. Rumelhart and J.L. McClelland, MIT Press, 1986, pp.318-362.

[45] T. Shin and I. Han, "Integration of Joint Time-Frequency Filtering Methods and Neural Network Techniques for Foreign Exchange Return Forecasting," *98 Fall Conference. Proceedings of Korean Expert Systems Society*, 1998, pp.205-217.

[46] K. Shin, T. Shin, and I. Han, "Neuro-genetic Approach for Bankruptcy Prediction: A Comparison to Back-propagation Algorithms," *International Conference of Korea Management Information Systems Society*, Seoul, Korea, 1998.

[47] C. Taswell, "Near-Best Basis Selection Algorithms with Non-additive Information Cost Functions," In Moeness G. Amin, (ed.) *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Philadelphia, PA, Oct. 1994, pp.13-16.

[48] C. Taswell, "The What, How, and Why of Wavelet Shrinkage Denoising," *Technical Report* CT-1998-09, Computational Toolsmiths, Stanford, 1998.

[49] B. Vidakovic, "Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors," *Journal of the American Statistical Association*, Vol. 93, No. 441, March 1998, pp. 173-179.

[50] G. Wahba, "Automatic Smoothing of the Log Periodogram," *J. Am. Stat. Soc.* Vol. 75, No. 369, 1980, pp.122-132.

[51] G. Wahba and S. Wold, "Periodic Splines for Spectral Density Estimation: The Use of Cross-Validation for Determining the Degree of Smoothing," *Comman. Statist.* Vol. 4, No. 2, 1975, pp.125-141.

[52] B. Walczak and D.L. Massart, "Noise Suppression and Signal Compression Using the Wavelet Packet Transform," *Chemometrics and Intelligent Laboratory Systems* 36, 1997, pp.81-94.

[53] A.S. Weigend, B.A. Huberman, and D.E. Rumelhart, "Predicting the Future: A Connectionist Approach," *International Journal of Neural Systems*, Vol. 1, 1990, pp.193-209.

[54] Weigend, A. and N.A. Gershenfeld (eds.), *Time Series Prediction: Forecasting The Future and Understanding The Past*, Addison-Wesley Publishing Company, 1994.

[55] D. Whitely and T. Hanson, "Optimizing Neural Networks Using Faster, More Accurate Genetic Search," *Proceedings of The Third International Conference on Genetic Algorithms*, San Mateo, 1989, pp. 391-396.

[56] G.W. Wornell, "A Karhunen-Loeve-like Expansion for 1/f Processes via Wavelet," *IEEE Transaction on Information Theory*, vol. IT-36, pp. 859-861, July 1990.

[57] G.W. Wornell, "Wavelet-based Representations for The 1/f Family of Fractal Processes," *Proceedings of the IEEE*, 81(10), (1993), pp.1428-1450.

[58] X. Yao, "Evolutionary Artificial Neural Networks," *International Journal of Neural Systems*, 4(3), 1993, pp.203-222.

[59] Young, R.K., *Wavelet Theory and Its Applications*, Kluwer Academic Publishers, 1993, pp.18-122.