유전자알고리즘을 이용한 웨이블릿분석 및 인공신경망기법의 통합모형구축
A Hybrid System of Wavelet Transformations and Neural Networks Using Genetic Algorithms: Applying to Chaotic Financial Markets

Taeksoo Shin, Ingoo Han**

Abstract

인공신경망을 시계열예측에 적용하는 경우에 고려되어야 할 문제중, 특히 모형에 적합한 입력변수의 생성이 중요시되고 있는데, 이러한 분야는 인공신경망의 모형생성과정에서 입력변수에 대한 전처리기법으로써 다양하게 제시되어 왔다.

가장 최근의 입력변수 전처리기법으로써 제시되고 있는 신호처리기법은 전통적 주기분할처리방법인 푸리에변환기법(Fourier transforms)을 비롯하여 이를 확장시킨 개념인 웨이블릿변환기법(wavelet transforms) 등으로 대별될 수 있다. 이는 기본적으로 시계열이 다수의 주기(cycle)들로 구성된 상이한 시계열들의 집합이라는 가정에서 출발하고 있다. 전통적으로 이러한 시계열은 전기 또는 전자공학에서 주파수영역분할, 즉 고주파 및 저주파수를 분할하기 위한 기법에 적용되어 왔다. 그러나, 최근에는 이러한 연구가 다양한 분야에 활발하게 응용되기 시작하였으며, 그 중의 대표적인 예가 바로 경영분야의 재무시계열에 대한 분석이다. 전통적으로 재무시계열은 장,단기의사결정을 가진 시장참여자들간의 거래특성이 시계열에 각기 달리 가격으로 반영되기 때문에 이러한 상이한 집단들의 고유한 거래움직임으로 말미암아 예를 들어, 주식시장이 프랙탈구조를 가지고 있다고 보기도 한다. 이처럼 재무시계열은 다양한 사회현상의 집합체라고 볼 수 있으며, 그만큼 예측모형을 구축하는데 어려움이 따른다.

본 연구는 이러한 시계열의 주기적 특성에 기반을 둔 신호처리분석으로서 기존의 시계열로부터 노이즈를 줄여 주면서 보다 의미있는 정보로 변환시켜 줄 수 있는 웨이블릿분석 방법론을 새로운 필터링기법으로 사용하여 현재 많은 연구가 진행되고 있는 인공신경망과의 모형결합을 통해 기존연구과는 다른 새로운 통합예측방법론을 제시하고자 한다.

본 연구에서 제시하는 통합방법론은 크게 2 단계 과정을 거쳐 예측모형으로 완성이 된다. 즉, 1 차 모형단계에서 원시 재무시계열은 먼저 웨이블릿분석을 통해서 노이즈가 필터링 되는 동시에, 과거 재무시계열의 프랙탈 구조, 즉 비선형적인 움직임을 보다 잘 반영시켜 주는 다차원 주기요소를 가지는 시계열로 분해,생성되며, 이렇게 주기에 따라 장단기로 분할된 시계열들은 2 차 모형단계에서 신경망의 새로운 입력변수로서 사용되어 최종적인 인공 신경망모델을 구축하는 데 반영된다.

기존의 주기분할방법론은 모형개발자입장에서 여러가지 통계기준치중에서 최적의 기준치를 합리적으로 선택해야 하는 문제가 추가적으로 발생하며, 본 연구에서는 이상의 제반 문제들을 개선시키기 위해 통합방법론으로서 기존의 인공신경망모형을 구조적으로 확장시켰다. 이 모형에서는 기존의 입력층 이전단계에 새로운 층이 제시된다. 이 층은 장단기 주기를 가지는 다수의 의미 있는 재무시계열을 생성시키는 웨이블릿 주기분할층으로 정의된다. 이렇게 해서 생성된 새로운 통합모형은 기존모형에서 생성되는 기본적인 학습파라미터와 더불어, 본 연구에서 새롭게 제시된 주기분할층의 파라미터들이 모형의 학습성과를 높이기 위해 함께 고려된다.

한편, 이러한 학습과정에서 추가적으로 고려해야 할 파라미터의 갯수가 증가함에 따라서, 본 모델의 학습성과가 local minimum 에 빠지는 문제점이 발생될 수 있다. 즉, 웨이블릿분석과 인공신경망모형을 모두 전역적으로 최적화시켜야 하는 문제가 발생한다.

본연구에서는 이 문제를 해결하기 위해서, 최근 local minimum 의 가능성을 최소화하여 전역적인 학습성과를 높여 주는 인공지능기법으로서 유전자알고리즘기법을 본연구의 통합모델에 반영하였다.

이에 대한 실증사례 분석결과는 일일 환율예측문제를 적용하였을 경우, 기존의 방법론보다 더 나은 예측성과를 나타내었다.

Key words: Fractal structure, Wavelet Transform, Wavelet Packet Transform, Optimal Decomposition, Neural Networks, Genetic Algorithms, Embedding Dimension

* Graduate School of Management, Korea Advanced Institute of Science and Technology

Introduction

Detecting the features of significant patterns from their own historical data is so much crucial to good performance specially in time-series forecasting.

The methods used for time series analysis are conventionally and heavily based on the concepts of stationarity and linearity. Linear models as autoregressive (AR), moving average (MA) and mixed (ARMA) models are often used. These are special cases of the general discrete linear model, which is a linear combination of present and past value of a white noise process.

But, there are fields where non-stationary and non-linear models are necessary, as in economics, oceanography, engineering, medicine, etc. A wide variety of non-linear models have been considered in the literature, as the bilinear models, threshold models, ARCH models, etc.

Recently, there has been a renewal of interest in linear expansions of signals, particularly using wavelets and some of their generalization (Daubechies, 1992; Mallat, 1989; Rioul and Vetterli, 1991).

That is, a new data filtering method (or multi-signal decomposition) such as specially, wavelet analysis is considered more useful for handling the time-series that contain strong quasi-cyclical components than other methods. The reason is that wavelet analysis theoretically makes much better local information according to different time intervals from the filtered data.

Generally multi-signal decomposition method such as Fourier analysis and wavelet analysis is a good method for the extraction of cyclical information bearing signals from corrupted observations. Specially, wavelet method having that advantage is considered as a concept that will need to be developed further for use in chaotic time series such as short-term economic or financial markets. To date, the present studies related to wavelet analysis are increasingly being applied to many different fields.

In this study, we suggest a hybrid learning architecture using multi-signal decomposition methods (i.e. wavelet analysis) and apply the combined learning 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.

We also use an artificial neural network model as a basic time series forecasting model. But, current artificial neural network model building needs lengthy experimentation and tinkering which is a major roadblock for the extensive use of the method. So recently a new combined model architecture using several algorithms has been suggested to overcome the limitation of single neural network model.

Through experimental results with wavelet filtering techniques we show the present different filtering criteria of wavelet analysis to support the neural network learning optimization and analyze the critical issues related to the optimal filter design problems in wavelet analysis. That is, those issues include that the human expert as a model developer should be confronted with the judgmental problem of finding the optimal filter parameter to generate significant input variables for the forecasting model.

Finally, from the second purpose we suggest new general optimal filtering criteria 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 signal decomposition layer before arriving at the input layer. Through that hybrid learning we tried to solve the present threshold problems about the optimal filter design in extracting the significant information from the original data.

In summary, the principal objective of this paper is to develop a new framework for multiscale signal representation in financial forecasting using artificial neural network models.

The rest of this paper is organized as follows. In the following section, we briefly review fractal structure of financial market, and in the third section also describe the chaos analysis, specially embedding dimension. The fourth section introduces multiresolution approach to wavelet and then wavelet transformation methodology about optimal decomposition from the original time series. The fifth section suggests a new methodology of hybrid system using genetic algorithms and then shows the experimental results and the conclusion contains final comments.

Fractal Structure of Chaotic Financial Markets

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.

A scheme called Multiresolution Embedding is constructed to discover whether some time-scales are more predictable than others. To achieve this goal, an Artificial Neural Network technique and a wavelet analysis are adopted in this study.

The investigation of multiscale representations of signals and the development of multiscale algorithms has been and remains a topic of much interest in many contexts. In some cases, such as in the use of fractal models for signals, the motivation has directly been the fact that the phenomenon of interest exhibits patterns of importance at multiscales. A second motivation has been the possibility of developing highly parallel and iterative algorithms based on such representations.

One of the DWT applications is to use it as an analysis to determine if the fractal dimension of a market indicator maintains consistency through different levels of scale. The reason is that the selection of the orthonormal wavelet used in the transform may influence the result, since these wavelets are themselves recursively defined and fractal in nature.

Embedding Dimension

In a time series $x_i(i=1,...,t,...,\infty)$, when the values of points previous to t were observed and x_t is to be predicted, the group of d data which are immediately previous to x_t, is to be predicted, the group of d data which are immediately previous to x_t, is used to predict x_t:

$$(x_{t-d+1}, \ldots, x_{t-2}, x_{t-1}) \rightarrow X_t \qquad (1)$$

where X_t is the predicted value of X_t. The d data are applied to the input nodes and X_t is produced in the output node as the predicted value. At the training stage, using the goup of d data which are immediately previous to $x_{t'}$ as a learning sample, $X_{t'}$ is produced by the network:

$$(x_{t'-d+1}, \ldots, x_{t'-2}, x_{t'-1}) \rightarrow X_{t'} \qquad (2)$$

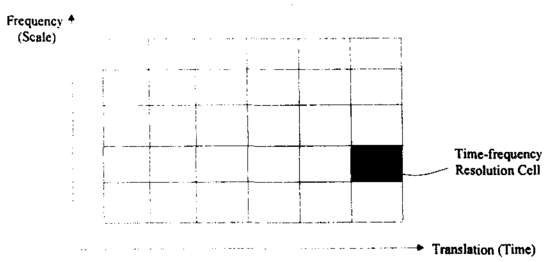
The weight values of the network are estimated using the error backpropagation method (Rumelhart et al., 1986) in which the sum of the errors between $x_{t'}$ and $X_{t'}$ for n learning samples is minimized. At the prediction stage, using these estimated weight values, x_t is produced. In the chaos theory, d is called the embedding dimension (Farmer and Sidorowich, 1987). With a chaotic time series, it is proved that the original characteristics of the chaos can be reconstructed from a single time series by using a proper embedding dimension.

In this study, we apply the dimension information to a model building, specially to selecting the number of time lagged input variables of neural network models.

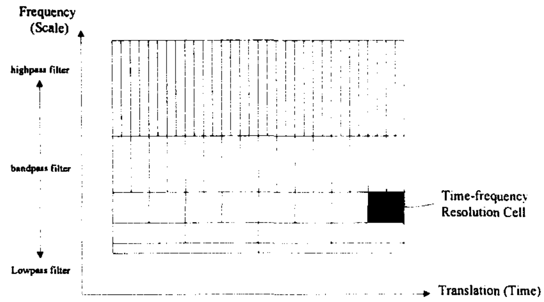
Multiresolution Approach to Wavelets

The wavelet analysis is a robust tool that may be used to obtain qualitative information for highly nonstationary time series - specifically, that it may be used to detect a small-amplitude harmonic forcing term even when the dynamics is chaotic and even for short total times. (Permann and Hamilton, 1992)

Each stage of resolution can be considered as a space which could be imagined to be represented as a linear combination of some suitable basis. Each resolution space is a subset of the resolution space which has a higher resolution. So we could say that 'mountain space' is a subset of 'house-car space' which is a subspace of 'human-animal-chair space' which is a subspace of 'key-spoon-coin space'... which is a subspace of 'atomic space'.



(a) Two Dimensional (Time-Frequency) Resolution of a Short-term Fourier Transform



(b) Two Dimensional (Time-Frequency) Resolution of a Discrete Wavelet Transform

Figure 1. Two Dimensional (Time-Frequency) Resolution

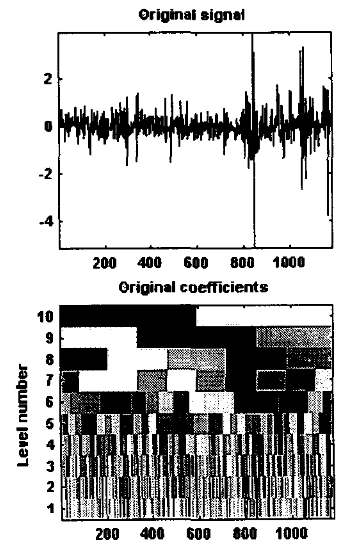


Figure 2. A case of Two Dimensional (Time-Frequency) Resolution of a Discrete Wavelet Transform (Daubechies with order 4) using the daily Korean Won/US Dollar returns

For present purposes we will be interested in the multiresolution structure of curves or spectra. Intuitively, we view high-frequency noise differently from broad, low-frequency components due to e.g. baseline effects.

By employing the multiresolution 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 aim of a wavelet transform is to decompose any signal f into a summation of all the possible wavelet bases at the different scales.

The coverage of the time-frequency plane for the wavelet analysis is shown in Figure 1(b).

Even though and the windowed Fourier transform (WFT) including the discrete Fourier transform (DFT) usually show the coverage as shown in Figure 1(a), they have 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 and the DWT filters shown in Figure 1 are related by a scaling function equal to 2. The scaling function could be any number, even a fraction. In addition, In the WFT, the filter bandwidths are constant, while the DWT filters are, again, related by a factor of two. Thus, for DWT filters, the width of the filter is proportional to its center frequency. One way of looking at this is that the wavelet approach partitions the data into blocks of equal information content - representing a potentially very useful characteristic of wavelet filters (Figure 1(b)).

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

In the pyramid algorithm the detail branches are not used for further calculations, i.e. only the approximations at each level of resolution are treated to yield approximation and detail obtained at level m+1. 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, 1993; Coifman *et al.*, 1993).

Coifman and Wickerhauser (1993) developed a wavelet packet transform; this is a more general transform than the discrete wavelet transform. The main difference is that while in the discrete wavelet transform the detail coefficients are kept, and the approximation coefficients are further analyzed at each step, in the wavelet packet transform 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.

The main characteristic of the wavelet packet transform is that it produces an arbitrary frequency split, which can be adapted to the signal. While wavelet packet 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

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 forms 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 span the frequency span of the original data set.

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

Lowpass filters pass all frequencies below the specified frequency, and they are usually employed for smoothing. Highpass filters pass all frequencies above the specified frequency. They are usually used to extract information on local variation while suppressing overall signal levels. Bandpass filters pass only those periodic components in the vicinity of the specified frequency.

The most basic type of filter, the one from which practically all other filters are derived, is the bandpass filter. As the name implies, this filter passes a single band of periodic components, stopping all components having higher or lower frequencies. We initially focus on the simple and eminently useful style that has a single frequency of maximum response and that tapers smoothly to zero on both sides of this frequency. It resembles the filter shown in Figure 3.

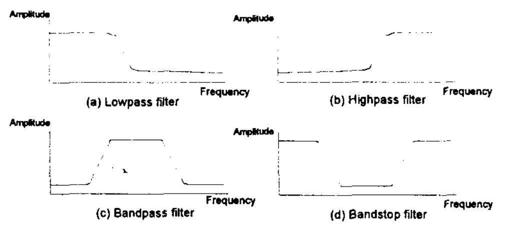


Figure 3. Frequency-response characteristics of the four basic filter types

The bandpass filter predominantly used in this study is characterized by two parameters. The center frequency, called f_o, is the frequency maximally favored by the filter. It may be any value from 0.0 to 0.5 (the Nyquist frequency) cycles per sample. The width, called s, defines the width of the passband. Note that the literature contains many definitions of the width of a passband. We define the width in terms of the filter equation, which will be seen shortly. The width is specified in the same units as the frequency, and it typically ranges from 0.01 to 0.2 or so.

The frequency response function of the bandpass filter used here is a Gaussian function. This is the function that multiplies the DFT of the time series before transforming back to the time domain.

These filters including lowpass, highpass, or bandpass filters are characterized by two parameters (frequency and width).

For lowpass filters, the frequency is the cutoff below which periodic components are passed and above which periodic components are obstructed. The width is the transition range over which the response of the filter goes from one extreme (unimpeded passage) to the other extreme (total cutoff). The frequency parameter always has the range 0.0 to 0.5.

The important point is that the reciprocal of the frequency parameter is the period of the periodic component. The width parameter is trickier to specify.

There is no simple calculation to provide the correct value. It is arbitrary choice. Unfortunately, a real tradeoff is involved.

Choosing the Optimal Decomposition

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 $n = 2^N$ can be expanded in at most 2^n different ways, the number of binary subtrees of a complete binary subtree of depth N. 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 are looking for a minimum of the criterion.

Functionals verifying an additivity-type property are well suited for efficient searching of binary-tree structures and the fundamentals splitting. Classical entropy-based criteria match these conditions and describe information-related properties for an accurate representation of a given signal. Entropy is a common concept in many fields, mainly in signal processing. But these criteria have a few limitations on choosing the optimal decomposition sub-series from original time series. That is, they have an inefficient learning problem and a misspecification problem of an object function for global model optimization.

Therefore, to solve these problems, 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.

Neural Networks

For time series predictions, the most popularly used neural networks are clearly time delay neural networks (TDNN; Weigend, Huberman, and Rumelhart, 1990) and recurrent neural networks (RNN; Elman, 1990). The time delay neural networks can be analyzed by using standard methods and more the results of such analysis can be applied for time series predictions directly, but they may not be sufficient to characterize the patterns of highly dynamic time series. On the other hand, the recurrent neural networks are suited for applications that refer to the patterns of genuinely time dependent inputs such as time series predictions due to their dynamic feature.

While in the dynamic context the recurrent neural networks can outperform the time delay neural networks, they occasionally are difficult to be trained optimally by a standard backpropagation algorithm due in part to the dependence of their network parameters (Kuan and Hornik, 1991).

In this study, The basic model we experiment with is Backpropagation neural network (BPN) models which have a parsimonious 4 input nodes, 4 hidden-nodes and 1 output node with single wavelet filter, i.e. highpass, lowpass, or bandpass filter within the network structure. The other model we experiment with is BPN models which have 8 input nodes, 8 hidden-nodes and 1 output node with multiple filters.

Genetic Algorithms

Genetic algorithms were introduced by John Holland and his group at the University of Michigan, of which Ken De Jong was the first to apply these algorithms to parameter optimization problems.

A genetic algorithm is used to search the weight space without use of any gradient information (Whitley and Hanson, 1989; Montana and Davis, 1989). A complete set of weights is coded in a binary string (or chromosome), which has an associated fitness that depends on its effectiveness. For example the fitness could be given by -E where E is 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 (Figure 4). The crossover operation combines part of one string with part of another, and can in principle bring together good building blocks - such as hidden units that compute particular logical functions - found by chance in different members of the population. The way in which the weights are coded into strings and the details of the genetic operators are both crucial in making this effective.

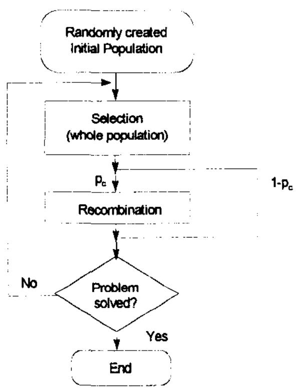


Figure 4. Genetic algorithm procedure

Genetic algorithms perform a global search and are thus not easily fooled by local minima. The fitness function does not need to be differentiable, so we can start with threshold units in Boolean problems, instead of having to use sigmoids that are later trained to saturation.

On the other hand there is a high computational penalty for not using gradient information, particularly when it is so readily available by back-propagating errors. An initial genetic search followed by a gradient method might be an appropriate compromise. Or a gradient decent step can be included as one of the genetic operators (Montana and Davis, 1989). There are also large costs in speed and storage for working with a whole population of networks, perhaps making genetic algorithms impractical for large network design.

Previous studies applied genetic algorithms to signal analysis. Their purpose was to extract a set of features characterizing the example of such a decomposition is the Fourier transform, which decomposes on a basis of harmonic functions. However, in the case of non-stationary signals, i.e., signals whose characteristics change with time, the Fourier transform does not yield a useful characterization of the signal.

We use genetic algorithms to build a new optimization method of decomposed univariate time series for financial forecasting model as Neural Network.

Neural Networks Training by Genetic Algorithm (GANN)

In the design of a neural network, a candidate parameter set of all weights and thresholds can be encoded by, for example, an integer string. Such a string is termed a chromosome in the GA context and a digit of the string is termed a gene. Initially, many such chromosomes are randomly generated to form a population of candidate designs. In this initial population, existing or known good designs can be conveniently included, which usually leads to a faster convergence. The GA uses three basic operators termed selection, crossover, and mutation to evolve, by the NP approach, a globally optimized network parameters set. The size of the population and the probability rates for crossover and mutation are termed control parameters of the GA.

Research Model Architecture

In general, for a one-dimensional distrete-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 a new hybrid time series forecasting model architecture as shown in Figure 5.

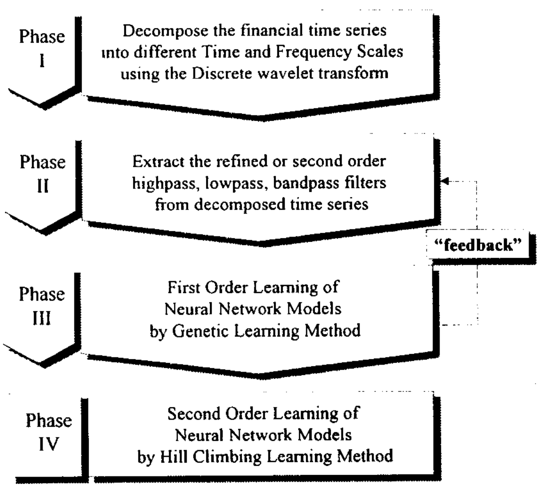


Figure 5. Proposed Research Framework

Specially, as shown in Figure 6, we suggest our hybrid neural network model as a extended neural network architecture compared to the prior model. That is, We consider a significant scale component generation automatically from the original data within our model. This function achieves at multi-scale extraction layer in Figure 6.

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 filter, highpass filter, and bandpass filter is proposed to control multiple scales of resolution within our research framework.

Each of the 10 scales was then multiplied by a weighting factor (0-1) and the weighted transform inverted back to a time-series (Figure 6).

In summary, our model architecture has an advantage over any other model architecture. That is, This hybrid learning system can solve the present wavelet thresholding problems in extracting the significant information from the original data by the optimal filter design.

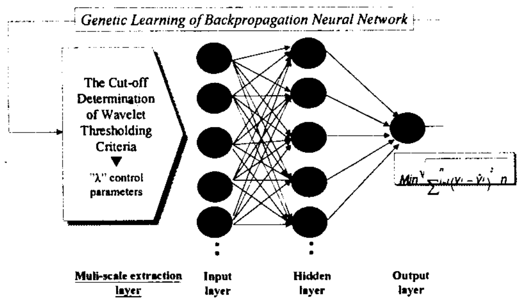


Figure 6. A Hybrid Neural network Model Architecture

$$Min\left(\sqrt{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2/n}\right) \quad (3)$$

$$\hat{y}_i = f(IWT_i)$$

$$IWT_i = \sum_{j=\lambda_1}^{\lambda_2} DS_i(j)$$

$$s.t.\ X_i = \sum_{j=1}^{N} DS_i(j)$$

$$1 \le \lambda_1 \le \lambda_2 \le N$$

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,

$f(\cdot)$ = Neural network model,

IWT_i = ith case of refined or second order wavelet transformed filter of population at (t)th day,

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

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

N = the length of decomposition levels of pyramid or tree algorithms,

n = the population size.

Among the various cost measures that one can pick for finding adaptive time-frequency decompositions, we pick an evolutionary data driven criteria. The benefits of this are as follows. Since our wavelet thresholding cost measure is more general than the others because of using final model performance measure as a cost measure within a unified model framework. In fact, these are a special case of general cross-validation measure (Figure 8).

Wavelet Periodogram (Scalogram)

The scalogram (Rioul and Flandrin, 1992) refers to the absolute value of the wavelet coefficients $W_{s,l}$, usually plotted logarithmically and 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.

The scalogram is the discrete wavelet transformation (DWT) counterpart to the well-known notion of periodogram in the spectral analysis of time series. 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. The scalogram of the discrete wavelet transform of a time series is the key too used to decompose the series into cycles of different frequencies. But, as shown in Figure 6, by the distribution shape of scalogram it is mostly difficult to extract the multi-cyclic structure from the original data like a granger shape curve.

From Figure 7 we extract lowpass filter, highpass filter, and bandpass filter from a decomposed time series (i.e. band 1-10). For example, the highest band, band 1, corresponds to high-frequency components in the data. These components correspond to signals with very short periods.

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

Decomp. Series	*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

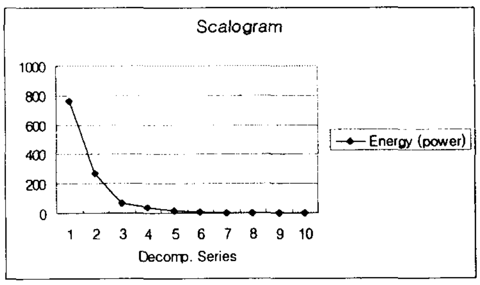


Figure 6. Scalogram of the wavelet coeffcients of daily Korean Won / US Dollar returns

In addition, we effectively extract a bandpass filter to the data by eliminating the highest and lowest bands.

Therefore, we combine the total bands (1-10) of decomposed output from the DWT into generating 3 separate bands, i.e. lowpass, highpass, and bandpass filter.

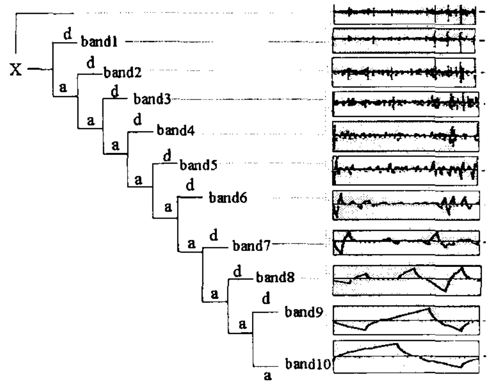


Figure 7. Tree algorithm using daily Korean Won / US Dollar returns (X: original series, a: approximation components, d: detailed componenets, band 1 (the highest highpass filter) → band10 (the lowest lowpass filter))

Various techniques for optimal bandwidth selection have been studied in present studies (Moulin *et al.*, 1992; Wahba and Wold, 1975; Wahba, 1980; Brillinger, 1975; Priestley, 1981; Jenkins and Watts, 1968). They produce estimates that have a good overall bias variance tradeoff. However, the bias versus variance tradeoff is generally not optimal locally.

Selection of Weighting Factors for WT Scales

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 seizure examples can be calculated (i.e. a measure of relevance). The 'relevance' of each scale to all the seizure examples can then be translated into weighting factor. For example, if a scale represents no features of any of the seizure examples, its weighting factor would be 0, while a weighting factor of 1 would be applied to a scale that represents features in all the examples. If a scale represents features I only some examples, the weighting factor would have an intermediate value. Using the weighted scales, the inverted WT results in a filtered time-series.

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

GA based Design for Hybrid Neural Network Model

The idea of combining Genetic algorithms (GAs) and Neural Network (NN), i.e. neuro-genetic approach came up first in the late 1980s, and it has generated a intense field of research in the 1980s.

Neural network is one tool that has generated a great deal of interest because it addresses the nonlinear nature of the financial markets. GAs offer a general-purpose tool for performing search and optimization functions. GAs, in addition to stand alone systems, are an excellent tool that can be used with other technologies including neural networks, machine learning systems and genetic programming.

In this study, for use with the GA, every weight has been coded by two digits with values in the range [-4, 4]. The thresholds are also treated as weights, with an input value of 1. In addition, our suggested multi-scale extraction layer's weights are added to the present neural network model.

Genetic algorithms basically are used to automatically determine the wavelet thresholding cut-off parameters including the learning parameters of neural networks. That is, the wavelet thresholding parameters are adjusted to optimize the performance of the financial forecasting over the entire samples (i.e. training samples)

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 genetic operators such as crossover and mutation which are described in the previous section are used to search for the optimal weight set solutions. Several parameters must be defined for the above operators, and the values of these parameters can greatly influence the performance of the algorithm. 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.

In neuro-genetic approach, the learning of a neural network is formulated as a weights optimization problem, usually using the inverse-mean-square error as a fitness measure. The basic concept behind this technique is as follows. A complete set of weights is coded in a string, which has an associated "fitness" representing its effectiveness. Starting with a random population of such strings, successive generations are constructed using genetic operators to construct new strings out of old ones such that better strings are more likely to survive and to participate in crossover operations. Unlike the back-propagation learning rule, GAs perform a global search and are thus not easily fooled by local minimum. The utilization of the linkage among population searches makes the GA a good global search method.

Experiments

In this section, we evaluate our framework using a case of the daily Korean Won / U.S. Dollar exchange rates are transformed to the returns using the logarithm and 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 order 4 for neural network forecasting models.

The set of signal features that are extracted should be independent of the time-varying noise field and the sensor dynamics. Since the feature vectors are typically composed of combinations of broadband and narrow band energy estimates, the signal spectrum should be whitened across the entire band. Thus, the first step is to use an adaptive time-domain whitening filer to decorrelate the data from the long-time ambient noise, interference, and sensor characteristics, while passing short-duration signals relatively unchanged.

It was observed in our previous study that discriminant parameters obtained using wavelet transforms yield better performance than those using autoregressive (AR) or moving average (MA) modeling, but not better than those using spectral coefficients.

The experimental results of our hybrid neural network architecture are showed in Table 2.

At first step, we compare two learning methods within hybrid system. That is, our model is trained using two learning methods, i.e. only a genetic learning method and a combined learning method by genetic and hill climbing learning (GA-NN) and then each method is compared. As shown in Table 2, Combined GA-NN method has better performance than one genetic learning method.

Secondly, according to filter type, the performance is different among them. The model using highpass, lowpass, and bandpass filter at once has better performance than models using partial filters.

Table 2. The Model Performance Using Test Samples

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

a: highpass (1-5), lowpass (2-10), bandpass (1-5),
b: highpaa+lowpass+bandpass filters, c: Random Walks,
d: Genetic Algorithms, e: Hill Climbing Learning,
h: Genetic Algorithms (λ) + Hill Climbing Learning.

In this study, 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 has better than any other models in terms of forecasting performance (Table 3). At this experiments we use a benchmark model to compare with our model's performance as follows. Namely, we use three wavelet threshold algorithms, i.e. bestbasis, cross-validation, and best level technique.

Table 3. The model performance comparison between different wavelet filtering methods and a genetic approach using test samples

Wavelet Threshold Algorithms	*Filter Types*	*Learning Methods*	*Performance (RMSE)*
BestBasis	LP[a]&HP[b]	HC	1.74329
Cross-validation	LP&HP	HC	1.676247
Best Level	LP&HP	HC	1.746597
GA[d]	LP&HP& BP[c]	GA-HC	**1.119327**

a: Lowpass filter, b: Highpass filter,
c: Bandpass filter, d: Genetic Algorithms.

Concluding Remarks

In this study, we have described a new framework for modeling and analyzing signals at multiple scales in time series forecasting problems.

In conclusion, this paper illustrates a decision support to build a hybrid forecasting system using genetic algorithms. By adjusting the input parameters to the appropriate values, multiple scale of resolution is implemented easily by discrete wavelet transform techniques. Once the financial time series has been segmented into areas with relative homogeneous value levels, the transformed information is learned as a more refined filter to extract the desired structure in neural network models.

We also conclude that our hybrid system of wavelet transformations and neural networks is much better than other models in increasing forecasting performance. That reason is as follows.

Our system finds the optimal filter parameter to extract significant input features of the forecasting model by machine knowledge, i.e. from a combined data driven approach and so improves the present different filtering criteria power of wavelet analysis in viewpoint of the neural network model optimization or performance.

References

[1] Abramovich, F. and Y. Benjamini, "Adaptive Methods of Wavelet Coefficients as Multiple Hypothesis Testing Proceedure," 1994.

[2] Alsberg, B.K., 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.

[3] Ariño, M. A. and B. Vidakovic, "On wavelet scalograms and their applications in economic time series," *Discussion paper 95-21*, ISDS, Duke University, 1995.

[4] Aussem, A., 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.

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

[6] 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.

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

[8] Coifman, R.R. 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.

[9] Coifman, R.R., Y. Meyer, S. Quake, and M.V. Wickerhauser, "Wavelet Analysis and Signal Processing," in *Wavelets and Their Applications*, J.S. Byrnes, J.L. Byrnes, K.A. Hargreaves, and K. Berry, eds., Kluwer Academic: Boston, 1994, pp.363-380.

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

[11] Daubechies, I., "Time-frequency localization operators: A geometric phase space approach," *IEEETIT* 34, 1988, pp.605-612.

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

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

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

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

[16] Donoho, D.L., 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.

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

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

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

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

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

[22] Katul, G. and B. Vidakovic, "The partitioning of attached and detached eddy motion in the atmospheric surface layer using Lorentz wavelet filtering," 1996.

[23] Kuan, C.M. and T. Liu, "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks," *BEBR Working Paper*, 92-0128, University of Illinois, Urbana-Champaign, 1992.

[24] Lin, S., 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.

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

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

[27] Mittermayr, C.R., 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.

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

[29] Morettin, P.A., "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.

[30] Moulin, P., "Wavelet Thresholding Techniques for Power Spectrum Estimation," *IEEE Transaction on Signal Processing*, Vol. 42, No.11, November 1994, pp.3126-3136.

[31] Müller, U.A., et al, "Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components", Preprint by the Olsen and Associates, 1995.

[32] Nason, G.P., "Wavelet Shrinkage using cross-validation," *Journal of the Royal Statistical Society*, Series B, vol. 58, 1996, pp.463-479.

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

[34] Rumelhart, D.E., 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.

[35] Shin, T. and I. Han, "Integration of Joint Time-Frequency Filtering Methods and Neural Network Techniques for Foreign Exchange Return Forecasting," 한국전문가시스템학회 추계학술대회, 1998, pp.205-217.

[36] Shin, K., 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(경영정보학회 추계학술대회), 1998.

[37] Taswell, C., "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.

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

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

[40] Walczak, B. and D.L. Massart, "Noise suppression and signal compression using the wavelet packet transform," *Chemometrics and Intelligent Laboratory Systems* 36, 1997, pp.81-94.

[41] Weigend, A.S., B.A. Huberman, and D.E. Rumelhart, "Predicting the Future: A Connectionist Approach," *Intl. J. Neur. Sys.* 1, 1990, pp.193-209.

[42] Weigend, A., *et al.*(eds.), *Time Series Prediction : Forecasting the Future and Understanding the Past*, Addison-Wesley Publishing Company, 1994.

[43] Whitely, D. and T. Hanson, "Optimizing Neural Networks Using Faster, More Accurate Genetic Search," *Proc. of the third Intern. Conf. on Genetic Algorithms*, San Mateo, 1989, pp. 391-396.