Evolutionary Wavelet Networks for Statistical Time Series Analysis

TOMMASO MINERVA
Department of Social Sciences
University of Modena and Reggio Emilia
Via Allegri 9, Reggio Emilia, I-42100
ITALY
tommaso.minerva@unimore.it

Abstract: In this paper I describe an evolutionary wavelet network to optimize the filtering of a statistical time series into separate contributions. The wavelet base is regarded as a neural network where the network nodes are discrete wavelet transforms, the wavelon, and the network structure and parameters are selected through evolutionary techniques. With this combined approach I can separate stochastic from structural components within an optimized framework and finally I can perform optimized predictive analysis on the time series components.

Key-words: Statistical Time Series, Evolutionary Algorithms, Signal Processing, Wavelet, Filtering, Stochastic Decomposition, Wavelet Networks

1 Introduction

In the early 90s, Zhang (Zhang and Benveniste, 1992; Zhang 1994; et al.) introduced wavelet networks as one hidden-layer net where the neurons in the hidden layer are wavelets (wavelon).

There are some advantages in using Wavelet Networks instead of feed forward neural nets mainly related to the problem that standard feed forward networks can bring to a lack of efficiency for determining the net parameters and the net topology. Moreover, the issue of the optimal network selection is still an open problem.

As shown further Wavelet Networks can be efficiently used to filter non-stationary and complex time series (such as financial t.s., environmental t.s., dynamical control systems, etc…).

If we consider the non parametric regression model involving an unknown function:

\[ f : \mathbb{R}^d \rightarrow \mathbb{R} \]

\[ Y = f(X) + \varepsilon \]

where \( X \) (input) and \( Y \) (output) are random variables in \( \mathbb{R}^d \) and \( \mathbb{R} \) respectively, \( \varepsilon \) is a noise of zero mean, \( f \) is supposed to be non linear in some functional space and \( d \) is the input dimension.

If \( O^n = \{(x_1,y_1), (x_2,y_2), ..., (x_n,y_n)\} \) is a sample of input-output pair (training data set) the problem of regression is to find a non-parametric estimator \( f^* \) of \( f \) based on the observations. In this case non parametric means that \( f^* \) belongs to some parameterized class of functions but flexible and with an unknown number of parameters in \( f^* \).

This kind of non-parametric estimators are useful for black-box non-linear modelling (such as Neural Networks). The design of the estimator, usually, consists in two main steps:

1) Setting a class of possible candidate \( f^* \) functions (usually very large);
2) Developing an algorithm to determine the parameters in \( f^* \) based on the observed data.

In the proposed design the two points are addressed at the same footing. The first, by using the scheme of wavelet networks while the second by evolutionary computational approaches.

The introduction of evolutionary wavelet networks in time series analysis can introduce some advantages respect to standard wavelet transform.

Infact we should keep into full account that in signal filtering we must deal with border effects and this is a crucial effect in predictive models on Time Series where the border, i.e. the more recent observations, is the most significant area of data.

We should also consider that it is not always possible to build orthonormal basis so the difference between
standard wavelet transforms and wavelet networks could be also based on the properties of the expansion basis. In this could be useful to consider that we can achieve more freedom if we give up the idea that the wavelet family should be a basis requiring only that the family is a redundant frame of localized functions. We can also give up the idea of a fixed regular lattice for the dilations and translations because this introduce some distortions on the ‘natural’ wavelet spectra. Moreover we prefer to use observed data to build the wavelet family (frame). The relation:

\[ f(x) = \sum_{i=1}^{N} u_i(a, b) \psi(a, x - b) = \sum_{i=1}^{N} u_i \psi(x) \]  

(2)

can be regarded as a one-hidden-layer feedforward neural network where each of the \( N \) neurons in the intermediate layer (wavelon) is a dilated and translated wavelet and \( u_i \) represents the weight of the \( i \)-th wavelon. In this relation the parameters \( u_i, a_i, b_i \) represent the wavelet expansion parameters and they have to be estimated numerically on the training data while the number of wavelon (hidden nodes), \( N \), represents another parameter and another parameter is represented by the mother-wavelet, \( \psi \) to be used in the wavelet expansion.

This 3 layers (input-wavelon-output) structure is called Wavelet Network.

We introduce also a bias parameter on each neuron to obtain eq. (3):

\[ f(x) = \sum_{i=1}^{N} u_i(a, b) \psi(a, x - b) + \theta_i \]

\[ = \sum_{i=1}^{N} u_i \psi(x) + \theta_i \]  

(3)

The network structure (sketched in figure 1) has a wavelet on each neuron and can be generalized to \( M \) inputs and \( N \) outputs.

This network structure can be considered as another universal function approximator but with some kindly properties:
- Can efficiently deal with local properties and high non-linearity respect to Feed Forward Networks or Radial Basis Networks;
- Can be demonstrated that Radial Basis Networks are a special class of Wavelet Networks;
- Can efficiently deal with a high-order input space respect to standard wavelet transforms;
- Can be more flexible respect to standard ortho-normal discrete wavelet transform;
- Can be efficiently used in wavelet filtering without border effects because one has not to evaluate the coefficient as an integral over the wavelet.

If we reconsider the eq. (3) we can identify the network parameters. The model, i.e. the wavelet decomposition, is established only when we are able to estimate the complete and optimal set of parameters: \( u_i, a_i, b_i, \theta_i \), \( N \) and \( \psi \) for each \( i=1,N \).

With this constrains the evaluation of the model, i.e. the set of parameters in equation (3) could be not so trivial and complex computational techniques have to be used.

This is not a trivial task and several approaches has been proposed in recent literature. In this paper I introduce an approach based on evolutionary algorithms.

2 Evolutionary Wavelet Networks

Zhang (1992, 1994) proposed an approach starting from a discrete candidate set of wavelet and then selecting the appropriate ones by stepwise-like algorithms, where the neural weights were evaluated by usual adaptive gradient-like algorithms. Y.He, F.Chu and B. Zhong (2002) and K. Kobayashi and T. Torioka (1998) have already proposed some evolutionary approaches to overcome the strong sensitivity to initial condition and path dependency in the stepwise approaches.

Here I generalize this evolutionary approaches to evaluate on the some footing all the parameters in a
Wavelet Network without lacking the property to separate the signal (the time series) into different frequency domain to obtain an efficient filtering.

Evolutionary Computation refers to a class of population based stochastic global search algorithms developed from the idea and principles of natural evolution. Individuals in a population encode a candidate solution and compete and exchange information with each other in order to perform a task, exactly as in natural evolution. Evolutionary operators (selection, crossover, mutation, elitism,...) evolve the population respect to a fitness function measuring the goodness of each individual (solution) respect to the optimization problem. Evolutionary approaches can be useful for dealing with large complex problems with many local optimal. They can perform global search without being trapped in local optima and do not use gradient-like operators so do not require usual regularities of the functions. All these properties indicate that evolutionary approaches can be useful in tackling our problem.

An Evolutionary scheme is represented by the following pseudo-code of a general Evolutionary Schema:

```plaintext
set i=0;
randomly generate the initial population P(0);
doi{
evaluate the fitness function of each individual in P(i);
select parents from P(i) based on their fitness value;
apply evolutionary operators (crossover, mutation, elitism, reinsertion) to parents to produce offsprings in P(i+1);
i++;
}
while (convergence criterion is satisfied)
optimal solution == best individual;
```

where \( P \) represent the population (set of candidate solutions) and \( i \) is the generation (searching iteration).

The main components in this Evolutionary Schema are represented by:

- The Individual Encoding, i.e. how the optimization/selection problem is coded in terms of Evolutionary Chromosome;
- The Fitness Function, i.e., how to evaluate the goodness of an individual respect to the optimization problem;
- Parameters of the evolutionary operators: mutation rate, selection rate, selection rule, crossover, etc…;

The main difficulty in applying evolutionary schema is to set up the genetic codification of the search/selection problem and to establish an efficient fitness function able to drive algorithm to the optimal solution.

In our problem the genetic code has to be able to determine the optimal combination of the parameters \( u_i, a_i, b_i, \theta, N \) and \( \psi \) while the fitness function has to measure the predictive goodness of the solutions.

If we fix \( N \), the number of wavelons (wavelet neurons), the genetic code is composed by \( N + 1 \) fragments.

Each one of the \( N \) fragments represents the parameters related to each wavelon: \( u_i, a_i, b_i, \theta \) while the \( N+1 \)th fragment encodes the mother-wavelet. So if we have \( N \) node in the wavelon layer than we have \( N \) fragments encoding the wavelon parameters (\( u_i, a_i, b_i \) and \( \theta \)) and a fragment encoding the wavelet family (see figure 2).

Fixed \( N \), the parameters related to each wavelon: \( u_i, a_i, \psi \),...
\( b, \theta \) are obtained by evolutionary minimization of the root mean square (RMS) error between target and output of the wavelet network. The number of wavelons, \( N \), is selected by an iterative loop minimizing the Akaike Final Prediction Error Criterion (FPE) that compare the goodness of the approximation and the number of parameters in the wavelet networks. Usually 200 individuals and 2000 generation are used in the evolutionary computation for each iteration but the algorithm stops after few hundred of generations (depending on the net complexity) and requires from few seconds (low \( N \)) to more than 4 hours (\( N > 40 \)) of CPU-time on a Mac Os X Server 2x2,8 Ghz quad-core Processor with 16 GB of RAM.

Firstly we tested the Evolutionary Wavelet Networks on several sample of non linear functions obtaining better approximations respect to other Wavelet Network reported in literature and respect to back-propagation Neural Nets. We investigate also the forecasting capabilities on the well known (used as a benchmark) sun-spot time series. The results are summarized in the table where the RMS is evaluated on a disjoint test set.

In this calculation the data were disjointed into a training set, a validation set and a test set to perform out-of-the-sample validation analysis. The weights were obtained on the training set while the fitness function was evaluated on the validation set and, finally, the predictive analysis was performed on a disjoint test set.

In table 1 I show that the Evolutionary Wavelet Network (EWN) better performs respect to standard Wavelet Networks, Wavelet Transform decomposition, Back Propagation feed-forward Neural Networks (BP-NN) and linear AR models. In the table is shown that the EWN results simpler respect to other network structure with only 21 hidden nodes while maintaining the best predictive RMS. The better of EWN is mainly obtained in treating border effects on the recent data.

The optimized Evolutionary Wavelet Network can be used as a filtering engine in stochastic time series filtering where noise and signal are separated by a wavelet decomposition [5].

### 3 Conclusions

In this paper I reported an evolutionary approach to build efficient wavelet expansion to perform predictive time series analysis based on stochastic wavelet filtering. The evolutionary approach can be used to break some regularity assumption of standard wavelet decomposition as, for example, regular wavelet lattice and regular wavelet basis.

The wavelet filtering allowed us to separate stochastic from deterministic components for time series coming from complex phenomena (such as financial markets).
as reported in a previous work (Minerva, 2010). For the deterministic component we perform a predictive analysis while the stochastic components were included on the basis of stochastic simulation techniques and, finally, we obtain a predictive distribution. The approach has been applied on benchmark datasets (sun spot) and on financial time series obtaining encouraging results also for long predictive horizon \((t+20)\).

The Evolutionary Wavelet Network has been compared to other linear (AR) and non linear models (Wavelet, Wavelet Network, Back Propagation Network) obtaining a simpler but more efficient predictive model.

References


