1º WORKSHOP NACIONAL EM REDES NEURONAIS E 1ª ESCOLA DE REDES NEURONAIS

Nome do Coordenador: Manoel Fernando Tenorio

Instituição: Purdue University

Departamento: Electrical Engineering

Endereço para correspondência: Purdue University
School of Electrical Engineering MS/EE BLDG
West Lafayette, Indiana 47907-1285
USA

Telefone: Fax: 317-494-6440

Composição da Equipe (Além do coordenador)

Antonio G. Thome; William HSu; Craig Codrington; Sho Kuwamoto; Alex Yali; John Kassebaum

Infraestrutura Básica Disponível (Hardware/Software)

- Sun Worksation 3/50 and Sparc-1*
- MacII Microcomputer SE and FXII*
- Matlab/signal processing software package
- Simulab, Mathematica, MacSpin, Statview, ETC.*
- All engineering computing network facilities, with a maspar and the titan mini supercomputer

Cooperações Técnico-Científicas Existentes (Nacionais e Internacionais):

- Instituto De Logica, Filosofia, Tenica da Ciencia - ILTC

(*) Parallel Distributed Structures Laboratory - PDSL

ECONOMIC SIGNAL ANALYSIS, MODELLING AND ESTIMATION THROUGH A SIMILARITY BASED APPROACH

This project uses a new approach to analyze, capture and model those more complex structured, possibly non-linear and nonstationary time series. Spectral analysis and digital filter theory are used to break the original series into more coherent and power compatible components. Each resulting component is then treated individually as a totally independent time series. Finally the individual results are combined to provide future estimations for the original signal.

To decompose the original series not based on the traditional elements, such as trend, seasonality and irregular fluctuations, but based on their coherence in terms of orthogonal elements and compatible power (integral of the square of the element values), generates a more homogeneous series to be treated and modeled individually. Based on this approach, a series can be decomposed in as many subseries as necessary to better separate and capture the distinct underlying patterns.

In forecasting problems, it is generally agreed that if the original function can be well represented by a decomposition then each component can more readily be modeled due to a better signal-to-noise ratio. When creating predictions with each individual component of the series, one needs to keep in mind that the final model can only be represented by the combination of the submodels created. In the nonorthogonal case, the fusion of these components can be very complex and nonlinear due to the amont of redundant information. It may also be the case that the fusion, in practice, may lead to a worse prediction than the separate components. Such instability occurs mainly due to the high correlation between the components. In the orthogonal case, one may not end up with the most compact representation of the series, and may even lose the ability to make local judgements, since the representation gives global information on a transformed space. Nevertheless several key advantages are present in this class of representation schemes. First, the components are orthogonal and therefore have zero cross-correlation. Second, the fussion is linear since each predictor adds independent information about the dynamics. Third, the prediction process can be treated separately and each component that is added increases the accuracy of the overall prediction. Thus one can trade the ability and effort to model for accuracy.

Among several different prediction techniques that can be used over the orthogonal resulting series, one, based on a similarity approach is particularly interesting for economic time series. Through a similar process as used by human experts, the market can be decomposed into clusters of similar patterns of behavior whose outcomes, in the desired prediction horizon, are consistent among the elements in the same cluster. This property is here named *data consistency*. This transforms the prediction problem into a pattern recognition problem.

The reference database is built by defining clusters embodying the similar states selected according to the similarity operator. Such clusters divide the data hyperspace into hyperspheres represented by its centroid and a common radius. Once the cluster centers have been defined, they perform a number to symbol transformation, in which the trajectory of the original time series can be represented by a string of cluster names, with minimum information loss. Prediction is accomplished by determining which cluster centers are close to the present state.

An immediate implementation of this algorithm via a feedforward neural network is possible. First, the hidden layer weights are assigned to be the cluster centers. Second, the hidden units, which represent the clusters, compute the similarity measure between the input vector and the cluster centers. This measure is multiplied by the output weights of the hidden nodes, which are the prototype value for the cluster outcome. The output unit then integrates all these estimates and normalizes them.

This technique has shown to be promising for economic time series prediction. Our main conclusion from this study is that the prediction accuracy depends not only on power and data consistency, but it also depends on the sampling rate-to-prediction horizon ratio, as it was witnessed by the variances of the 3 components of a weekly based oil price time series used on the experiments. For a better prediction of the third component (that one with highest frequency), as well as improvements on the other 2, a daily signal would be necessary. We are now pursuing the design of on-line decomposition and predictive schemes.