Chapter 35
Hybrid Wavelet-Neuro-Fuzzy Systems of Computational Intelligence in Data Mining Tasks

Yevgeniy Bodyanskiy
Kharkiv National University of Radio Electronics, Ukraine

Olena Vynokurova
Kharkiv National University of Radio Electronics, Ukraine

Oleksii Tyshchenko
Kharkiv National University of Radio Electronics, Ukraine

ABSTRACT

This work is devoted to synthesis of adaptive hybrid systems based on the Computational Intelligence (CI) methods (especially artificial neural networks (ANNs)) and the Group Method of Data Handling (GMDH) ideas to get new qualitative results in Data Mining, Intelligent Control and other scientific areas. The GMDH-artificial neural networks (GMDH-ANNs) are currently well-known. Their nodes are two-input N-Adalines. On the other hand, these ANNs can require a considerable number of hidden layers for a necessary approximation quality. Introduced Q-neurons can provide a higher quality using the quadratic approximation. Their main advantage is a high learning rate. Universal approximating properties of the GMDH-ANNs can be achieved with the help of compartmental R-neurons representing a two-input RBFN with the grid partitioning of the input variables' space. An adjustment procedure of synaptic weights as well as both centers and receptive fields is provided. At the same time, Epanechnikov kernels (their derivatives are linear to adjusted parameters) can be used instead of conventional Gauss functions in order to increase a learning process rate. More complex tasks deal with stochastic time series processing. This kind of tasks can be solved with the help of the introduced adaptive W-neurons (wavelets). Learning algorithms are characterized by both tracking and smoothing properties based on the quadratic learning criterion. Robust algorithms which eliminate an influence of abnormal outliers on the learning process are introduced too. Theoretical results are illustrated by multiple experiments that confirm the proposed approach's effectiveness.

DOI: 10.4018/978-1-5225-2229-4.ch035

INTRODUCTION

The Computational Intelligence (CI) systems (especially artificial neural networks(ANNs)) are widely spread nowadays in a large class of identification, emulation, intelligent control and time series prediction tasks due to their universal approximating properties and their learning abilities. Since there's a number of practical tasks when a learning sample volume is restricted, a learning rate factor goes in the forefront.

At the same time, not all neural networks (first of all, the most popular multilayer architectures are learned with the error back propagation procedure) satisfy the real task conditions because of a low rate of a tuning process and a possible overfitting effect. Radial-Basis Function ANNs (RBFNs) are considered rather fast networks that are learned with the help of the second-order optimization procedures. Their output signals depend linearly on their adjusted synaptic weights.

However, RBFNs suffer from the so-called "curse of dimensionality", when a number of neurons (R-neurons) in a hidden layer runs up exponentially while an input vector signal's dimensionality is growing.

More advanced systems are such ones as neuro-fuzzy systems (NFSs) like ANFIS, Takagi-Sugeno-Kang, Wang-Mendel and others. They are characterized by improved approximating properties and abilities of linguistic interpretation compared to ANNs. But these systems are exposed to the same drawbacks, including both overfitting and the "curse of dimensionality" simultaneously.

The same thing can be said about wavelet-neuro-fuzzy systems (WNFSs) that are based on the NFSs. Their designation is to process non-stationary signals.

It may be possible to overcome the above-mentioned problems by partitioning initial tasks into a set of low-dimensional subtasks and uniting later in some way obtained results to get a required result. From a computational point of view, the most convenient method in this case is the Group Method of Data Handling (GMDH) (Ivakhnenko, 1970; Ivakhnenko, 1971a; Ivakhnenko, 1971b; Ivakhnenko, 1975; Ivakhnenko, 1986; Ivakhnenko, 1984; Madala, 1994) that demonstrated its efficiency while solving a number of practical tasks.

It seems to be reasonable to join the GMDH ideas with the CI systems to get new qualitative results in Data Mining, Intelligent Control and other scientific areas.

The GMDH-ANNs are currently well-known. Their nodes are two-input N-Adalines. Every N-Adaline contains a set of adjustable synaptic weights (which are estimated with the standard least squares method) and provides the quadratic approximation of a restored nonlinear mapping. On the other hand, these ANNs can require a considerable number of hidden layers for a necessary approximation quality.

Neural networks that can self-organize their architectures using inductive sorting-out GMDH algorithms for their neurons were proposed in (Ivakhnenko, 1995; Ivakhnenko, 1994; Ivakhnenko, 1999). For example, a Twice-Multilayered Neural Network has a conventional multilayered feed-forward architecture where each active neuron synthesizes within itself a multilayered GMDH-structure. This network demonstrated its efficiency while solving a number of practical tasks, but its usage is restricted by numerical unwieldiness.

Networks' architectures are both feed-forward and recurrent with a feedback loop (GMDH-FL). Their neurons can have polynomial, sigmoidal or radial-basis activation functions. These architectures are proposed in (Kondo, 1998; Kondo, 2002; Kondo, 2004; Kondo, 2008a; Kondo, 2008b; Kondo, 2003; Kondo, 2010). These networks can automatically organize their architectures using the heuristic self-organization method that is a type of the evolutional computation and based on the GMDH ideas. Synaptic weights of this network are tuned in a batch mode using the stepwise regression analysis. It's supposed that the neurons can have two and more inputs.