IDEA GROUP PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com

ITB13684

This paper appears in the publication,
Artificial Intelligence and Integrated Intelligent Information Systems: Emerging Technologies and Applications
edited by X. Zha © 2007, Idea Group Inc.

Chapter VII

An Evolutionary Framework for Nonlinear Time-Series Prediction with Adaptive Gated Mixtures of Experts

André L. V. Coelho, University of Fortaleza (Unifor), Brazil
Clodoaldo A. M. Lima, State University of Campinas (Unicamp), Brazil
Fernando J. Von Zuben, State University of Campinas (Unicamp), Brazil

Abstract

A probabilistic learning technique, known as gated mixture of experts (MEs), is made more adaptive by employing a customized genetic algorithm based on the concepts of hierarchical mixed encoding and hybrid training. The objective of such effort is to promote the automatic design (i.e., structural configuration and parameter calibration) of whole gated ME instances more capable to cope with the intricacies of some difficult machine learning problems whose statistical properties are time-variant. In this chapter, we outline the main steps behind such novel hybrid intelligent system, focusing on its application to the *nontrivial task of nonlinear time-series forecasting. Experiment results are reported with respect to three benchmarking time-series problems and confirmed our expectation that the new integrated approach is capable to outperform — both in terms of accuracy and generalization — other conventional approaches, such as single neural networks and non-adaptive, handcrafted gated MEs.*

Introduction

Time-series analysis and forecasting constitute an important research and application area. Much effort has been devoted over the past several decades to develop and improve generic models and methodologies capable to deal more properly with the complicated requirements commonly imposed by time-series analysis and forecasting (Weigend & Gershenfeld, 1994). Examples of well-established time-series models and methodologies include: (1) linear models such as moving average, exponential smoothing, and the auto-regressive integrated moving average (ARIMA) (Box, Jenkins, & Reinsel, 1994); (2) nonlinear models, such as neural networks, fuzzy systems, and, more recently, kernel-based models (Zhang, Patuwo, & Hu, 1998; Kim, Park, Hwang, & Kim, 1995; Cao, 2003); and (3) combined linear-nonlinear models (Zhang, 2003).

One of the neural-network-related techniques that has gained much attention recently, both in the areas of nonlinear regression and prediction, as well as on pattern classification, is referred to as mixtures of experts (MEs). Such conceptual learning framework, first introduced by Jacobs, Jordan, Nowlan, and Hinton (1991) and later extended by Jordan and Jacobs (1994), Xu, Jordan, and Hinton (1995), Weigend, Mangeas, and Srivastava (1995), and Meila and Jordan (1997), amount to a family of modular architectures devised to tackle complex problems in consonance with the divide-and-conquer principle: In a first stage, an original problem is divided into several smaller and simpler sub-problems that are subsequently solved independently; afterwards, solutions to those sub-problems are seamlessly integrated to form a complete solution to the original problem.

A single, non-hierarchical ME instance is composed of an array of supervised, competing modules, known as experts, whose activities are orchestrated by a mediator module, termed as gating. In the canonical ME model (Jacobs et al., 1991), the experts are assumed to be linear in nature whereas the soft-max function is adopted as the nonlinear activation function for producing the gating module's outputs. Conversely, in the gated ME variant, conceived by Weigend et al. (1995), a single nonlinear gating module is specified to coordinate competing nonlinear experts (implemented as multilayer perceptrons — MLPs) for dealing with different regimes of a given complex dynamic process. One advantage of using a nonlinear, instead of a linear, gating lies in the possibility of generating more elaborated time / space decompositions through nonlinear decision boundaries — something very attractive to cope with the nonstationarity and overfitting issues commonly present in nonlinear time-series prediction, for instance. Moreover, due to neural networks' universal approximation capabilities (Haykin, 1999), a well-configured nonlinear gating can, at least in theory, induce any sort of optimal problem decompositions. Therefore, gated MEs have achieved noticeable performance results in the identification of nonlinear systems (Lima, Coelho, & Von