Chapter VII
Committee of Networks for Estimating Missing Data

ABSTRACT

This chapter introduces a committee of networks for estimating missing data. The first committee of networks consists of multi-layer perceptrons (MLPs), support vector machines (SVMs) and radial basis functions (RBFs). The committee was constructed from a weighted combination of these three networks. The second, third and fourth committees of networks were evolved using a genetic programming approach and used the MLPs, RBFs and SVMs, respectively. The committee of networks was collectively implemented with hybrid particle-swarm optimization and a genetic algorithm for missing data estimation. They were tested on an artificial taster as well as HIV datasets and then compared to the individual multi-layer perceptron, radial basis functions and support vector regression for missing data estimation. It was found that the committee of network approach provided improved results over the three methods acting individually. However, this improvement comes with a higher computational load than does using the individual approaches. Furthermore, it is found that evolving a committee method was a good way of constructing a committee.

INTRODUCTION

Several techniques have been introduced for missing data estimation (Abdella, 2005; Abdella & Marwala, 2005, 2006; Allison, 2000). A number of of these methods make use of machine learning approaches to accomplish this mission. In this chapter, committees of machine learning algorithms are used for missing data estimation. The principal incentive for using the committees of networks technique is from the intuitive logic that many ‘heads’ are better than one and, therefore, using many networks is thus better than using one.

Committee of Networks for Estimating Missing Data

identification in structures, whereas Marwala (2001a) implemented a probabilistic fault identification process in structures using a committee of neural networks and vibration data. Furthermore, Marwala et al. (2001) used a committee of agents and genetic programming to evolve a stock market prediction system.

In this chapter, the fact that the committee of networks approach will yield superior performance to the stand-alone networks for regression problems is extended to the missing data estimation problem and then demonstrated mathematically. After that, it is tested on an artificial taster and HIV prediction problems that were described in earlier chapters. Two approaches are pursued in constructing committee methods. These are the traditional approach proposed by Perrone and Cooper (1993), and the second approach is evolving the committee method by using a type of evolutionary programming known as genetic programming.

MISSING DATA APPROACH

In this chapter, the missing data estimation approach that was adopted entailed the use of a committee consisting of an auto-associative Multi-Layer Perceptron (MLP), Radial Basis Function (RBF) and a Support Vector Machine (SVM). These auto-associative networks were trained to predict their own input vectors and therefore were called recall networks. The results obtained from these networks were averaged in a weighted manner. Thus far, the MLP, RBF and SVM, which are the members of the committee of networks presented in this chapter, have each been individually used for missing data estimation (Nelwamondo, 2008) as was shown in earlier chapters. Pelckmans et al. (2005) proposed a technique for handling missing values in support-vector machine classifiers while Junninen et al. (2004) developed missing values imputation methods for air quality data sets. On the other hand, Chandramouli et al. (2007) used artificial neural networks for estimating missing microbial concentrations in a riverine database whereas Zhong, Lingras, and Sharma (2004) estimated missing traffic counts using factor, genetic, neural, and regression techniques. In this chapter, just as in Zhong, Lingras, and Sharma (2004), regression methods are used for estimating the missing data. The missing data estimation procedure adopted in this chapter is composed of two components i.e., regression and optimization components.

The missing-data estimation error equation was written in earlier chapters as follows:

\[
e = \left\| \left\{ X_{\hat{s}} \right\} - f \left( \left\{ X_s \right\} \right) \right\| \quad (7.1)
\]

In equation 7.1 the observed data of the complete dataset \( \{ X \} \) is \( \{ X_s \} \), the missing vector to be estimated is \( \{ X_{\hat{s}} \} \), \( \| \| \) is the Euclidean norm, \( f \) is the auto-associative network, which in this chapter is a multi-layer perceptron, a radial basis function or a support vector machine. The missing data estimation objective is, therefore, defined as the desire to identify the missing vector \( \{ X_{\hat{s}} \} \) that would minimize equation 7.1 given the known parameters \( \{ X_s \} \) and the auto-associative mathematical model \( f \). In this chapter, to approximate the missing input values, equation 7.1 is minimized using some optimization technique. In this chapter the hybrid genetic algorithm and particle swarm optimization method (HGAPSO) is used for this task. This hybrid approach is chosen over the traditional gradient-based approaches because of its simplicity of implementation and its characteristic of ensuring a higher probability of finding the global optimum solution than the traditional optimization methods (Goldberg, 1989). Furthermore, it is
21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product’s webpage: www.igi-global.com/chapter/committee-networks-estimating-missing-data/6799?camid=4v1


Related Content

Integrated Manufacturing Applications and Management Decision Making: Putting the P Back into ERP
Fergal Carton and Frederic Adam (2008). Handbook of Computational Intelligence in Manufacturing and Production Management (pp. 263-280).
www.igi-global.com/chapter/integrated-manufacturing-applications-management-decision/19363?camid=4v1a

Unexplored Hypotheses on Potency-Magnitude Relations of eWOM Messages with Intensified Comparative Expressions
www.igi-global.com/article/unexplored-hypotheses-on-potency-magnitude-relations-of-ewom-messages-with-intensified-comparative-expressions/101316?camid=4v1a

The Formal Design Models of a Universal Array (UA) and its Implementation
Yingxu Wang, Jason Huang and Jingsheng Lei (2011). International Journal of Software Science and Computational Intelligence (pp. 69-89).
www.igi-global.com/article/formal-design-models-universal-array/60750?camid=4v1a

Magnetic Remanence Prediction of NdFeB Magnets Based on a Novel Machine Learning Intelligence Approach Using a Particle Swarm Optimization Support Vector Regression
WenDe Cheng (2014). International Journal of Software Science and Computational Intelligence (pp. 72-81).