Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 22 (2013) 172 – 181

Procedia Computer Science

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Hybrid Multi-layered GMDH-type Neural Network Using Principal Component Regression Analysis and Its Application to Medical Image Diagnosis of Liver Cancer

Tadashi Kondo*, Junji Ueno, Shoichiro Takao

Graduate School of Health Sciences, The University of Tokushima, 3-18-15 Kuramoto-cho Tokushima 770-8509, Japan

Abstract

In this study, a hybrid multi-layered Group Method of Data Handling (GMDH)-type neural network algorithm using principal component-regression analysis is proposed and applied to the computer aided image diagnosis (CAD) of liver cancer. In the GMDH-type neural network, a heuristic self-organization method that is a type of evolutionary computation, is used to organize the neural network architecture. In this revised GMDH-type neural network, the optimum neural network architecture is automatically organized from three types of neural network architectures, such as the sigmoid function neural network, the radial basis function (RBF) network and the polynomial neural network architecture, by the heuristic self-organization method. Furthermore, the structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables, are automatically determined using the heuristic self-organization method. In the revised GMDH-type neural network proposed in this paper, the principal component-regression analysis is used to protect multi-colinearity which has occurred in the learning calculations of neurons, and accurate and stable prediction values are obtained. This new algorithm is applied to the medical image diagnosis of liver cancer. In this application, two types of neural network architectures fitting the complexity of the multi-detector row CT (MDCT) medical images, are automatically organized using the revised GMDH-type neural network algorithm The first neural network recognizes and extracts the liver regions from the MDCT images of the liver, and the second neural network recognizes and extracts the liver cancer regions. These results are compared with the conventional sigmoid function neural network trained using the back propagation method, and this GMDH-type neural network algorithm is shown to be useful for CAD of liver cancer.

Selection and peer-review under responsibility of KES International

Key words: Neural Network; GMDH; Evolutionary computaion; Medical image diagnosis; Liver cancer;

* Corresponding author. Tel.:+81-88-633-9024; fax:+81-88-633-9024.
*E-mail address:*kondo@medsci.tokushima-u.ac.jp

Selection and peer-review under responsibility of KES International
doi:10.1016/j.procs.2013.09.093

1. Introduction

The Group Method of Data Handling (GMDH)-type neural networks and their applications have been proposed in our early works [1]-[4]. GMDH-type neural networks can automatically organize neural network architecture by heuristic self-organization method [5],[6] and they can also determine such structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables. But, in the conventional GMDH-type neural network algorithms [3],[4], multi-colinearity occurs in the learning calculations of the neurons and the prediction values become unstable. In this study, a revised GMDH-type neural network algorithm using principal component-regression analysis is proposed. In this algorithm, multi-colinearity does not occur and the accurate and stable prediction values are obtained. This new algorithm is applied to the computer aided image diagnosis (CAD) of liver cancer. In the previous study [1], the GMDH-type neural network was applied to the medical image recognition of liver and blood vessel regions in the liver, and these regions were extracted and displayed automatically using the revised GMDH-type neural network. In this paper, the liver cancer regions are recognized and extracted automatically using new revised GMDH-type neural network algorithm. In this algorithm, the structural parameters such as the number of layers, the number of neurons in hidden layers and useful input variables can be automatically selected so as to minimize prediction error criterion defined as Akaike's information criterion (AIC) [7] or Prediction Sum of Squares (PSS) [8]. The recognition results show that the revised GMDH-type neural network algorithm is useful for CAD of liver cancer and is easy to apply practical complex problem because optimum neural network architecture is automatically organized.

2. Heuristic Self-Organization [5],[6]

Heuristic self-organization method is a basic theory of the conventional GMDH algorithm [5],[6] and was applied to the GMDH-type neural network algorithm [4]. Architectures of GMDH-type neural network are automatically organized using heuristic self-organization method. First, the procedures of heuristic self-organization method are shown because it plays very important roles for organization of GMDH-type neural network. Heuristic self-organization method is a kind of the evolutionary computation.

Heuristic self-organization method is constructed by the following six procedures:

1) Separating original data into training and test sets

Original data is separated into training and test sets. Training data is used for estimating parameters of partial descriptions which describe partial relationships of the nonlinear system. Test data is used for organizing complete description which describes complete relationships between input and output variables of the nonlinear system.

2) Generating combinations of input variables in each layer

All combinations of two input variables (x_i, x_j) are generated in each layer. The number of combinations is $p!/(p-2)!2!$. Here, p is the number of input variables.

3) Calculating partial descriptions

For each combination, partial descriptions of the nonlinear system can be calculated by applying regression analysis to training data. Output variables of partial descriptions are called as intermediate variables.

4) Selecting intermediate variables

L intermediate variables which give L smallest test errors calculated using test data are selected from generated intermediate variables.

5) Iterating calculations from procedure 2) to 4)

Select L intermediate variables are set to input variables of the next layer and calculations from procedure 2) to 4) are iterated. The multi-layered architecture is organized.