Intelligent Techniques for the Diagnosis of Coronary Artery Disease

Ravi Jain

This thesis submitted for the degree of

Doctor of Philosophy

SUB CRUCE LUMEN

Department of Applied Mathematics,
The University of Adelaide
South Australia

(November 1998)

Contents

Abstract

In this thesis, a genetic-programming-based classifier system for the diagnosis of coronary artery disease is proposed. It maintains good classification and generalisation performance. Based on genetic programming, a software system called Evolutionary Pre-Processor has been developed which is a new method for the automatic extraction of non-linear features for supervised classification. The central engine of Evolutionary Pre-Processor is the genetic program; each individual in the population represents a pre-processor network and a standard classification algorithm. The EPP maintains a population of individuals, each of which consists of an array of features. The features are transformations made up of functions selected by the user. A fitness value is assigned to each individual, which quantifies its ability to classify the data. This fitness value is based on the ability of a simple classifier to correctly classify the data after it has been transformed to the individual's feature space. Through the repeated application of recombination and mutation operators to the fitter members of the population, the ability of the individuals to classify the data gradually improves until a satisfactory point in the optimisation process is reached, and a solution is obtained.

Recently there has been a rising interest in using artificial intelligent (AI) techniques in the field of medical diagnosis. However, it is noted, that most intelligent techniques have limitations, and are not universally applicable to all medical diagnosis tasks. Each intelligent technique has particular computational properties, making them suitable for certain tasks over others. Integration of domain knowledge into empirical learning is important in building a useful intelligent system in practical domains since the existing knowledge is not always perfect and the training data are not always adequate. Genetic algorithms (GAs) are robust but not necessarily the most successful optimization algorithms for any particular domain. Hybridizing a GA with algorithms currently in use can produce an algorithm better than both the GA and the current algorithms. A GA may be crossed with various problem specific search techniques to form a hybrid that exploits the global perspective of the GA and the convergence of the problem specific technique. In some cases, hybridization entails employing the representation as well as the optimization techniques already in use in the domain while tailoring the GA operators to the new representation.

In this connection a hybrid intelligent system is highly desirable. Here two different hybrid techniques are also presented. In the first approach, fuzzy systems is integrated with genetic algorithms. In this approach, each fuzzy if-then rule is treated as an individual and each population consists of certain number of fuzzy if-then rules. It can automatically generate fuzzy if-then rule from training patterns for multi-dimensional for pattern classification problems. Classifiers in this approach are fuzzy if-then rules. In the second approach genetic algorithms are combined with back-propagation algorithms to enhance the classification performance. In this approach, a complete set of weights and biases in a neural network are encoded in a string, which has an associated fitness indicating its effectiveness. Each chromosome completely describes a neural network. To evaluate the fitness of a chromosome, the weights on the chromosome are assigned to the links in a network of a given archtecture, the network is then run over the training set of examples and the sum of the sqaures of the errors is returned from each example.

All approaches were tested on a real-world problem of coronary artery disease data.