Diagnosing Prostate Cancer Using Backpropagation Neural Network and Greedy Decision Procedure

Anilkumar K.Gopalakrishnan and Thitipong Tanprasert
Parallel and Distributed Computing Lab (PDCL), Faculty of Science and Technology,
Assumption University, Soi 24, Ram Khamhaeng Road, Hua Mak,
Bang Kapi, Bangkok 10240, Thailand.
anil@scitech.au.edu, nui@scitech.au.edu

Abstract

A novel procedure for diagnosing prostate cancer (PC) based on Back propagation Neural Network (BPNN) is proposed. Elderly men with symptoms such as urinary retention, urinary hesitancy, urinary dribbling, burning urination, hematuria, etc. are considered as primary attributes. Prostate-specific antigen (PSA) level and Gleason score are the secondary attributes. Initial dataset is generated based on the clinical database. The BPNN assigns symptom levels of a set of patients based on their primary attributes. A greedy decision procedure predicts tumor stages of patients based on their strong symptom levels and secondary attributes. The simulation shows that the proposed procedure is an effective way for diagnosing prostate tumor stages.

Key Words: Backpropagation Neural Network, Prostate Cancer, Prostate-specific antigen, Gleason score, Symptom level, Greedy decision procedure

1. Introduction

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining can be used to uncover patterns in data but is often carried out only on *samples* of data. The mining process will be ineffective if the samples are not a good representation of the larger body of data. Data mining cannot discover patterns that may be present in the larger body of data if those patterns are not present in the sample being "mined". The discovery of a particular pattern in a particular set of data does not necessarily mean that a pattern is found elsewhere in the larger data from which that sample was drawn. An important part of the data mining process is the verification and validation of patterns on samples of generated dataset. This paper normalizes the collected clinical dataset [1] for the diagnosis of prostate cancer without losing the generality of the clinical data.

The neural networks are capable of adapting human thoughts in the form of numerical values. Typically, a neural network is initially "trained" or fed large amounts of data and rules about the data relationships. The BPNN used in this paper is trained with back-propagation algorithm and train the initial data (seen data) until its *Mean Squared Error* (MSE) is reduced to a value less than 0.01. In a typical 3-layer BPNN, the computation time will be asymptotically $\Theta(ih + ho)$, where i, h, and o are the number of input neurons, hidden neurons and the output neurons, respectively [2].

There is a greedy decision procedure is applied along with the BPNN to generate fast predictions based on the outputs from the neural network. In a 'greedy method' the idea is to build up a particular solution procedure which should produce the best immediate or local solutions for a problem. Moreover, greedy algorithms are quicker since they do not consider the details of the possible alternatives [3].

When considering the prostate cancer, there are more chances of 100% recovery from its aggressiveness. Even if there are effective medical therapies exists now a days towards the cancer,