Breast Cancer Data Prediction by Dimensionality Reduction Using PCA and Adaptive Neuro Evolution

R. R. Janghel, Indian Institute of Information Technology and Management, Gwalior, India
Ritu Tiwari, Indian Institute of Information Technology and Management, Gwalior, India
Rahul Kala, Indian Institute of Information Technology and Management, Gwalior, India
Anupam Shukla, Indian Institute of Information Technology and Management, Gwalior, India

ABSTRACT

In this paper a new approach for the prediction of breast cancer has been made by reducing the features of the data set using PCA (principal component analysis) technique and prediction results by simulating different models namely SANE (Symbiotic, Adaptive Neuro-evolution), Modular neural network, Fixed architecture evolutionary neural network (F-ENN), and Variable Architecture evolutionary neural network (V-ENN). The dimensionality reduction of the inputs achieved by PCA technique to an extent of 33% and further different models of the soft computing technique simulated and tested based on efficiency to find the optimum model. The SANE model includes maximum number of connections per neuron as 24, evolutionary population size of 1000, maximum neurons in hidden layer as 12, SANE elite value of 200, mutation rate of 0.2, and number of generations as 100. The simulated results reflect that this is the best model for the prediction of the breast cancer disease among the other models considered in the experiment and it can effectively assist the doctors for taking the diagnosis results as its efficiency found to be 98.52% accuracy which is highest.

KEYWORDS: Breast Cancer, Data Prediction, Fixed Architecture Evolutionary Neural Network (F-ENN), Principal Component Analysis (PCA), Symbiotic, Adaptive Neuro-Evolution (SANE)

INTRODUCTION

The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as cancer research, gastroenterology, heart diseases, brain tumors etc. (Sengur, 2007). Fully automatic malignant and benign breast cancer classification is of great importance for research and clinical studies.

Since the neural networks are sensitive to the number of inputs, as large number of inputs may lead to under-training, immensely
large training time, loss of generality, inability to model ideal functional surfaces, etc.. In such a context, the solution lies in making effective dimensionality reduction techniques for reducing the number of input attributes, whereby having the least loss of information.

The PCA performs the task of dimensionality reduction. It takes as its input a database that consists of a large number of attributes, and mines out the most interesting attributes or combination of attributes. The resultant attributes may be better suited for solving the concerned problem, and have a smaller dimensionality (Sengur, 2007).

The different models of soft computing are available in the literature and comparisons there comparison has been done. The Hybrid approaches form a very exciting field of work and research for large and complex data sets. In these systems we combine Neural Networks, Evolutionary Algorithms, Fuzzy Logic and heuristics in numerous ways to make a much more efficient system (Janghel et al., 2010).

Evolutionary Computing techniques are search algorithms based on the mechanisms of natural selection and genetics. That is, they apply Darwin’s principle of the survival of the fittest (Darwin, 1859) among computational structures with the stochastic processes of gene mutation, recombination, etc. Central to all evolutionary computing techniques is the idea of searching a problem space by evolving an initially random population of solutions such that better - “fitter” – solutions obtained.

The evolutionary algorithm searches for the most productive decision strategies using only the infrequent rewards returned by the underlying system. Together evolutionary algorithms and neural networks offer a promising approach for learning and applying effective decision strategies in many different situations.

Co-evolutionary algorithms are great problem solving tools in such scenarios, where the complete problem of optimization of a highly dimensional fitness landscape may be broken down into multiple sub-problems that together constitute the main problem. The sub-problems are simpler to solve or optimize, and hence aid in providing effective components for the solution of the main problem. It is however important for the different sub-problems to co-operate with each other, in order to enable effective evolution that ultimately results in locating global minima.

Novel neuro-evolution mechanisms called SANE (Symbiotic Adaptive Neuro-evolution) are efficient for sequential decision learning. Unlike most approaches, which operate on a population of neural networks, SANE applies genetic operators to a population of neurons. Each neuron’s task involves establishing connections with other neurons in the population to form a functioning neural network. Since no one neuron can perform well alone, they must specialize or optimize one aspect of the neural network and connect with other neurons that optimize other aspects. SANE thus decomposes the search space, which creates a much more efficient genetic search. Moreover, because of the inherent diversity in the neuron population, SANE can quickly revise its decision policy in response to shifts in the domain.

In this paper, we first use PCA for dimensionality reduction of the breast cancer database. The original database consisting of 30 attributes usually becomes very large and complex hence; the dimensionality is been reduced by PCA before giving to the input of SANE. In this problem SANE is a representative of the class of co-evolutionary neural networks. Then we compare the performance of SANE with modular neural network (MNN), fixed architecture evolutionary neural network (F-ENN) and variable architecture evolutionary neural network (V-ENN).

LITERATURE REVIEW

Fully automatic malignant and benign breast cancer classification is of great importance for research and clinical studies. Recent work has shown that classification of breast cancer is possible via supervised techniques such as artificial neural networks and support vector machine (SVM), and unsupervised classification tech-
Developing Professional Knowledge and Confidence in Higher Education
www.igi-global.com/chapter/developing-professional-knowledge-confidence-higher/70111?camid=4v1a