Applied Soft Computing 49 (2016) 834–844

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Neural network classifier optimization using Differential Evolution with Global Information and Back Propagation algorithm for clinical datasets

N. Leema (Research Scholar)[a], H. Khanna Nehemiah (Associate Professor)[a,*], A. Kannan (Professor)[b]

[a] Ramanujan Computing Centre, College of Engineering Guindy, Anna University, Chennai 600025, Tamilnadu, India
[b] Information Science and Technology, College of Engineering Guindy, Anna University, Chennai 600025, Tamilnadu, India

ARTICLE INFO

Article history:
Received 4 December 2014
Received in revised form 29 July 2016
Accepted 1 August 2016
Available online 30 August 2016

Keywords:
Artificial Neural Network
Differential evolution with global information
Back propagation
Optimization
Computer-Aided Diagnostic system

ABSTRACT

A Computer-Aided Diagnostic (CAD) system that uses Artificial Neural Network (ANN) trained by drawing in the relative advantages of Differential Evolution (DE), Particle Swarm Optimization (PSO) and gradient descent based backpropagation (BP) for classifying clinical datasets is proposed. The DE algorithm with a modified best mutation operation is used to enhance the search exploration of PSO. The ANN is trained using PSO and the global best value obtained is used as a seed by the BP. Local search is performed using BP, in which the weights of the Neural Network (NN) are adjusted to obtain an optimal set of NN weights. Three benchmark clinical datasets namely, Pima Indian Diabetes, Wisconsin Breast Cancer and Cleveland Heart Disease, obtained from the University of California Irvine (UCI) machine learning repository have been used. The performance of the trained neural network classifier proposed in this work is compared with the existing gradient descent backpropagation, differential evolution with backpropagation and particle swarm optimization with gradient descent backpropagation algorithms. The experimental results show that DEGI-BP provides 85.71% accuracy for diabetes, 98.52% for breast cancer and 86.66% for heart disease datasets. This CAD system can be used by junior clinicians as an aid for medical decision support.

1. Introduction

In the past few decades Artificial Neural Network (ANN) has been widely used for classification in the field of machine learning [1–3]. ANN represents an information-processing paradigm that is inspired by the way biological nervous systems process information, and is widely used to address the classification problem [4]. For most of the real world applications, multilayered, feed-forward neural networks are used, with supervised learning backpropagation (BP) algorithm. Three types of BP algorithms namely Gradient Descent, Conjugate Gradient and Quasi Newton are used for local search in ANN training. Gradient Descent BP, Gradient Descent with momentum BP, Gradient Descent with adaptive learning rate BP, Gradient Descent with momentum and adaptive learning rate BP and Resilient BP belong to Gradient Descent type. Fletcher-Reeves, Polak-Ribiere, Powell-Beale and Scaled Conjugate Gradient belongs to Conjugate Gradient type. Levenberg-Marquardt, BFGS and One Step Secant belongs to Quasi Newton type [48]. The efficiency of this algorithm depends on several parameters namely, the learning rate, number of nodes in the hidden layer and types of activation function. Without choosing appropriate network parameters, it slows down the training speed and the NN gets strucked in the nearest local minimum. The training outputs of the NN are entirely dependent on the initial weights [5–8]. The local search with faster convergence of ANN for classification has been improved by various researchers [9–11].

Particle Swarm Optimization (PSO) developed by Kennedy and Eberhart [12,13] can be applied to overcome the local minima problem occurring in any optimization problems. Though PSO has good performance in classification but it has the problem of premature convergence due to stagnation, which means that a particle remains still or cannot move from its position [14,25,31]. To overcome the stagnation problem of PSO, the diversity of the search space is increased. This can improve the global search ability of PSO in the solution space. Many improved PSOs have been proposed in the past to improve the global search of PSO [15–17].

* Corresponding author.
E-mail addresses: chandrakhanna@cs.annauniv.edu, nehemiah@annauniv.edu (H.K. Nehemiah).

http://dx.doi.org/10.1016/j.asoc.2016.08.001
1568-4946/

To obtain better performance in classification, the double search methods namely APSO, DE-BP, GA-BP have been proposed [18,19,49]. It is a combination of heuristics based local search and metaheuristics based global search method [22]. Combining ANN with PSO enhances the power of global search and reduces the local minima problems of ANN. The PSO approach gives a global search ability to NNs, which enables convergence to global optimum with increase in classification accuracy [20,21]. The combination of NN with the PSO approach is applied in medical domains in the development of CAD systems in disease diagnosis [23].

In this work, the ANN is trained using DEGI-BP to develop a CAD system for classifying clinical datasets is proposed. DEGI-BP is proposed to overcome the premature convergence of PSO and improves the accuracy of the ANN classifier. DE's modified best mutation operation increases the search space of PSO. The global best value obtained by training ANN using PSO is used as the seed for BP. Local search is performed using gradient descent BP, in which the weights of the NN are adjusted to obtain an optimal set of NN weights.

The rest of the paper is organized as follows: Section 2 presents a literature review about the existing research work. Section 3 explains the materials and methods used in the proposed work. Section 4 is devoted to the design and implementation of the proposed system. In Section 5, the experimental results are discussed. Finally, Section 6 concludes the work and provides the scope for future work.

2. Literature review

There are several works in literature that use metaheuristic optimization approaches for enhancing the classification performance of ANN. This section discusses these works, and also highlights the novelty and contribution of this work.

Jing-Ru Zhang et al. presented an optimization algorithm by combining particle swarm optimization with backpropagation for training feed forward neural network [22]. The local and global search abilities of ANN are improved by them using BP and PSO respectively. Turker Ince et al. have investigated the performance of global and local training techniques that can be applied to neural network classifiers [23]. They have experimented and compared the classification performance of the NN-BP and NN-PSO algorithms. The NN-BP algorithm gives better generalization, local search ability and provides better classification performance for some (to increase or decrease the number of hidden nodes) network configurations. They have experimented the two techniques on three benchmark datasets, namely, Wisconsin Breast Cancer (WBC), Cleveland Heart Disease (CHD), and Pima Indian Diabetes (PID) from the UCI machine learning repository. They have stated that the NN-PSO provides the global search ability and minimum classification error.

Hatem Abdul-Kader et al. developed a system to evaluate differential evolution and particle swarm optimization algorithms in training neural networks for stock prediction [24]. The authors implemented and compared two meta-heuristic approaches namely DE and PSO, which are used to train the NN. They experimented these approaches on stock market datasets obtained from different drug manufacturing companies. The experimental results show that both the algorithms converge to the global minimum and improve the classification accuracy of the NN. The advantages of these two methods are that they avoid the local minima and tune the NN parameters.

Masoud Yaghini et al. presented an algorithm by combining a global search strategy of opposition based PSO and the local search ability of BP with momentum term [25]. Opposition based learning and random perturbation during the iteration explore the search space. Constriction factor is used to ensure convergence of particles. Piotrowski et al. in their research explored the differential evolution algorithms suffered from stagnation which when applied to neural network training [26]. The authors introduced the DE with global and local neighborhood-based mutation operator, which performs better than the Levenberg-Marquardt algorithm for neural network training. They have experimented the techniques on benchmark problems and concluded that without selecting the proper mutation and population size, the DE will fall into stagnation during NN training.

Fei Han et al. developed a diversity-guided hybrid particle swarm optimization based on gradient search (DGPSOGS) [27]. The authors presented an algorithm which adaptively searches local minima with random search realized by adaptive PSO, and also performs global search with semi-deterministic search realized by DGPSOGS. Therefore, the search ability is improved. From this work, the developed hybrid algorithm has better convergence performance with better diversity compared to some classical PSOs. Arpit et al. have presented an approach for breast cancer diagnosis using genetically optimized ANN model (GONN) [28]. In this work, the NN structure and weights are optimized using the genetic programming approach for classification. They introduced new modified crossover and mutation operators with the genetic programming life cycle for reducing the destructive nature of these operators. The model was tested on WBC dataset obtained from the UCI machine learning repository. The classification accuracies of this approach are 98.24%, 99.63% and 100% for 50–50, 60–40 and 70–30 training – testing partition respectively.

Najmeh et al. in their work have presented an approach for optimization of neural network model using modified bat algorithm [29]. In this work, the NN structure and weights are optimized based on the echolocation behavior of bats. They have modified the basic algorithm by considering personal best solution in the velocity adjustment, the mean of personal best and global best solutions. These modifications are used to improve the exploration and exploitation capability of the bat algorithm. During the training process of ANN, the selection of structure, weights and biases are influenced by these modifications. The ANN parameters are tuned using the Taguchi method. The datasets used in this work are glass identification, credit card, breast cancer, thyroid dysfunction, diabetes diagnosis and iris. These classification datasets are obtained from the UCI machine learning repository. The time series datasets are Gas furnace and Mackey-Glass. They have tested the developed method to real world problem to predict the future values of rainfall data and the results show satisfactory of the developed model.

Hsiao et al. have developed liver cancer prediction model using an artificial neural network for type-II diabetes patients [45]. They have used ANN and Logistic Regression (LR) prediction model, and also used 70–30 ratio to partition the training and testing data. The dataset was obtained from the National Health Insurance Research Database (NHIRD) of Taiwan. From the database they identified 2060 cases and divided them into two groups namely, patients diagnosed with liver cancer after diabetes and patients with diabetes but no liver cancer. The ANN based predictive model provides 83.7% accuracy and LR based predictive model provides 73.7% accuracy.

Comparing with the work discussed in the literature, the proposed work is different in the following ways: The existing work in the literature has two major limitations. ANN that is optimized using BP and traditional metaheuristic approach gets stuck in a local minimum and also suffers from premature convergence due to stagnation. Hence, in this work a metaheuristic approach that combines the relative advantages of DE and PSO is proposed. The approach modifies the best mutation operation of DE using linearly non-increasing weight values, explores the search space. PSO finds the global best weights between the input to hidden and hidden to output layer. These obtained global best weights are used as a seed