International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 4, April-2015, pp: (1-6), Impact Factor: 1.252, Available online at: www.erpublications.com

A Particle Swarm Optimization Algorithm Based Classifier For Medical Data Set

Baydaa S. B. Alyas
Department of Software Engineering
College of Computer Science and Mathematics, Mosul University, Mosul, Iraq

Abstract: In this paper we have proposed a PSO based classification model for multidimensional real dataset. Here we have used multidimensional real medical datasets (thyroid and renal data sets) for classification using PSO to observe the accuracy of classification. Evaluation of the performance of PSO based classifier has been made by computing the correct rate. The obtained results indicate that PSO-Classifier is an effective technique for classification and can be used successfully on more demanding problem domain.

Keywords: Classification, Correct rate, Euclidean distance, Particle Swarm Optimization, Swarm Intelligence.

Introduction

Swarm Intelligence (SI) is an innovative distributed intelligent paradigm for solving optimization problems that originally took its inspiration from the biological examples by swarming, flocking and herding phenomena in vertebrates. The Particle Swarm Optimization Algorithm (PSO) is one of swarm intelligence optimization algorithms, which were proposed by Eberhart and Kennedy (1995) [4]. The idea is originated from the exchange and sharing of information among bird individuals in the process of searching food. Each individual can benefit from discovery and flight experience of the others [8].

PSO algorithm is easy to fall into local optimization, to cause premature convergence problem. This is mainly because the proportion of tracking global best position is too much. The accumulating of all particles' tracking leads to track much faster. Therefore, population diversity will have a rapid decline. This causes large amounts of particles to turn into similar particle, even only a few particles or one particle. Namely, the main reason is rapid decline of population diversity. Thus, the iterations curve decline a great deal more quickly in the beginning, but much more slowly in the later. This is particularly obvious in the vicinity of the optimal solution[11].

In the literature, Several computer-aided classification algorithms have been proposed for this task. The authors in[9] set the PSO parameters (C1 and C2) in the way such that PSO based classifier. This will give the best result for classification of multidirectional real dataset (cancer data, diabetics data). In [5] the author proposed PSO-SVM approach for an automatic ECG beat classification. This approach presents methods for improving SVM performance in two aspects: feature selection and parameter optimization. The simulation results indicate that the PSO-SVM method can correctly select the discriminating input features and also achieve high classification accuracy (97.21%). In [7] the authors proposed a patient-adapting heart beat classifier system based on linear discriminates. In [3] the authors proposed new morphological features; also they used temporal features together their morphological features. They adopted multi-layer perceptron (MLP) neural networks with various training algorithms as their classifier. In [6] the authors present a method for electrocardiogram (ECG) beat classification based on particle swarm optimization (PSO) and radial basis function neural network (RBFNN). This method gives the best classification performance with the smallest size of network.

Particle Swarm Optimization based classifier

Particle swarm optimization (PSO) is a stochastic based search algorithm widely used to find the optimum solution. PSO incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged . PSO is a population-based optimization tool, which could be implemented and applied easily to solve various function optimization problems. As an algorithm, the main strength of PSO is its fast convergence, which compares favorably with many global optimization algorithms like Genetic Algorithms (GA) , Simulated Annealing (SA) and other global optimization algorithms. For applying PSO successfully, one of the key issues is finding how to map the problem solution into the PSO particle, which directly affects its feasibility and performance [8][10].

PSO works on the social behavior [4] of particle where the individuals, referred to as particles, are grouped into a swarm. Like other population based optimization methods the particle swarm optimization starts with randomly initialized population for individuals. Each particle in the swarm represents a candidate solution to the optimization problem. In a PSO system, each particle knows its local best value so far (lbest) and its position. This means, each particle is "flown" through the multidimensional search space, adjusting its position in search space according to its own experience and that of neighboring particles. Moreover, each particle knows the global best value so far in the group (gbest) among lbests. A particle therefore makes use of the best position encountered by itself and the best position of its neighbors to position itself toward an optimal solution. The effect is that particles "fly" toward an optimum, while still searching a wide area around the current best solution. The performance of each particle (i.e. the "closeness" of a particle to the global minimum) is measured according to a predefined fitness function which is related to the problem being solved [8][10]. Namely, each particle tries to modify its position using the following information [2][9]:

- current positions
- current velocities
- distance between the current position and lbest
- distance between the current position and gbest

This modification can be represented by the concept of velocity. Velocity of each particle can be modified by the following equation [1]:

$$v_{id}(t+1) = w * v_{id} + c_{1*} r_{1*} (P_{id} - X_{id}) + + c_{2*} r_2 r_{1*} (P_{id} - X_{id}) \quad (1)$$

Where, v_{id}: velocity of particle

X_{id} : current position of particle

C_1 & C_2 : acceleration coefficient represent the degree of tracking individual optimal and global optimal respectively.

P_{id}: local best solution (pbest) of particle i,

P_{gd} : global best solution (gbest) of the group(all particles).

r_1 & r_2: random variables (uniformly distributed random numbers) between 0 and 1.

w : is the inertia weight

Using the above equation, a certain velocity, which gradually gets close to pbest and gbest can be calculated. The current position (searching point in the solution space) can be modified by the following equation [1]:

$$X_{id}(t+1) = X_{id}(t) + v_{id}(t+1) \quad (2)$$

Classification is a popular knowledge discovery technique to find previously unknown, valid patterns and relationships in large data set. Classification can be done in two phase first one is training phase and second is testing phase [9]. The classification requires less computation time during the training phase. Consider data points within an n-dimensional space where each of the n-dimensions corresponds to one of the n-features. Closeness (belongingness) of the data point can be measured using relative distance. We minimize the distance between two similar classified data point, while maximizing the distance between data points of different classes. To measure the relative distance between data points we are using Euclidean distance ED by using the following equation [2][9]:

$$\text{Euclidean Distance} : ED(X,Y) = \left(\sum_{i=1}^{N} (x_i - y_i)^2 \right)^{1/2} \quad (3)$$

Where X and Y are two data points of the dataset and i is the no of attributes. x and y are two attribute values of X and Y.

The process of classifier trained by particle swarm optimization is described as followings:

(1) Initialize the positions and velocity of all particles.

(2) Evaluate the fitness of each particle according to the desired optimization.

(3) Update local best and global best, and update the velocity and position of the particles according to local best and global best.

(4) Finally, stop-criterion is judged. The process is given in Fig.1.

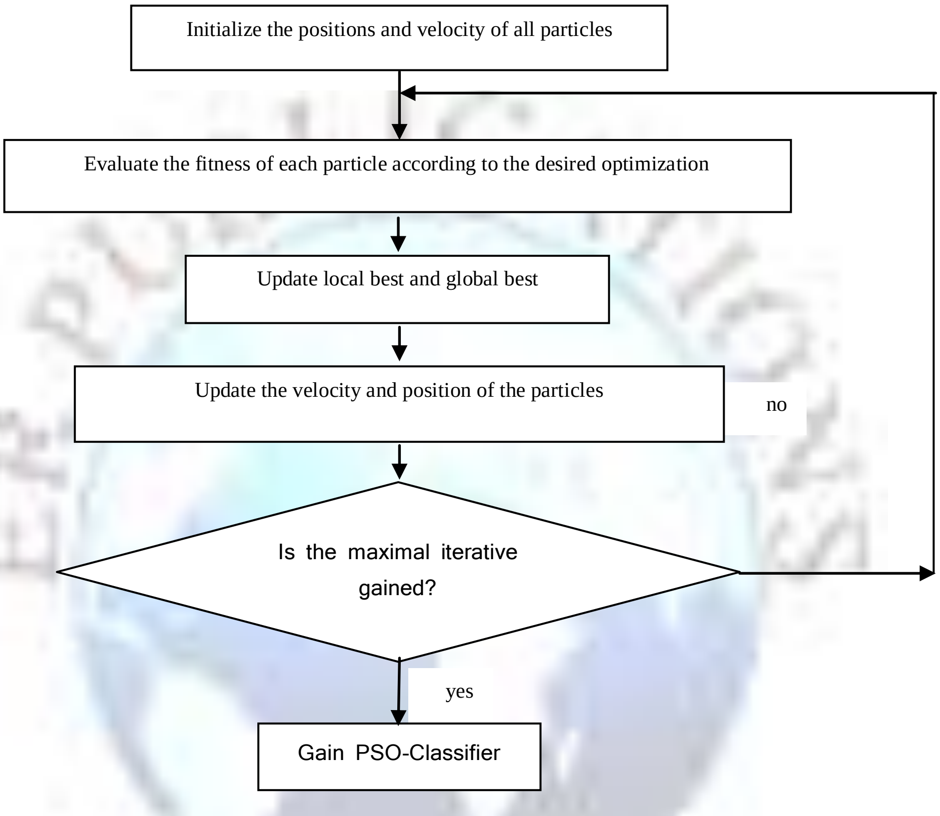

Figure 1. The process of classifier trained by particle swarm optimization

After the particles are generated, they are evaluated according to their capability of combining classifiers outputs in a way that the predicted class is correct. We evaluate particles using the fitness function correct rate to measure the classification performance of PSO-Classifier defined as follows:

$$\text{Correct rate} = \frac{\text{correct number}}{\text{correct number} + \text{error number}} \quad (4)$$

Experimental Results

In order to evaluate the ability of our algorithm PSO-Classifier to classify medical data set, we have tested it using thyroid gland and renal data sets. These medical data sets are collected from Al-Jamhoree Teaching Hospital under the supervision of doctor for classification. The thyroid dataset contains 733 patients for training and 293 for testing with 44 attributes. First three attributes are examinations and the others attributes are signs and symptoms of this disease as shown in Table 1:

Table 1 : Signs and symptoms of thyroid disease.

Attribute Information
Thyroid-Stimulating Bland Chromatin Hormone (TSH)
Triiodothyronine (T3)
Thyroxin (T4)
Goitre
Vomiting and diarrhea
Pigment increase in the palm of the hands and feet
Weakness and difficulty of reproduction
Bouts of panic and headache
Chest pains
More frequent bowel movements
Namwalazafar very quickly
Weak libido
Trouble sleeping
Abdominal pain
Fine hair
Bulging eyes and with the advent of visual problems
Nervousness and Anxiety
Increased sweating
Thyroid Antibodies
Paresthesia
Imbalance
Dry or coarse skin
Coarse hair
Hoarseness with difficulty in speech
Difficulty in hearing
Slow in thinking and frequent forgetfulness
High blood pressure
High cholesterol
An increase in the menstrual in women
Menstrual disorder or interruption in women, and may lead to impotence in men
Irregular heartbeat and tachycardia
Shortness of breath
Tremors in the hands and fingers
Weight gain despite poor appetite
Feeling cold more than others and the inability to withstand cold weather
Pain, fatigue and muscle weakness
Feeling depressed
Constipation
Slowing the heart rate
Reduced respiratory rate
Anemia
Weight loss despite increased appetite
Increased sensitivity to heat
Osteoporosis fractures occur

We are training thyroid data set in two phase. In the first training phase, we classify this data set into three classes (Normal , Hyperthyroidism and Hypothyroidism) and five classes (Normal , Hyperthyroidism , T3-thyotoxicosis , Primary Hypothyroidism and Secondary Hypothyroidism) in the second training phase. Table 2 shows the classification accuracy of PSO-Classifier.

Table 2 : Classification accuracy of PSO-Clasifier for thyroid data set

	Three classes		Five classes	
	train	**test**	**train**	**test**
Correct rate	94.4837	93.2821	95.5425	94.2051

Also we PSO-classifier method for the renal data set that contains 260 patients for training and 75 for testing with 23 attributes. First three attributes are examinations and the others attributes are signs and symptoms of this disease as shown in Table 3.

Table 3 : Signs and symptoms of renal disease.

Attribute Information
Blood Urea (BU)
Serum Uric Acid (SUA)
Serum Creatinine (SC)
Vomiting and/or diarrhea, which may lead to dehydration
Feeling tired and/or weak
Difficulty sleeping
Change the color and the amount of urine daily
Swelling in the hands, feet, abdomen, or face
Appetite loss
Shortness of breath due to extra fluid on the lungs (may also be caused by anemia)
Bones and chest pains
Abnormal heart rhythms or failure of the heart muscle
Irritating itch
Muscle paralysis
Darkening of the skin
Headache and blurred given and can sometimes be a result of the high in blood pressure
Sexual Dysfunction and menstrual disorders in women
A decline in sexual desire and menstrual disorders in women
Feeling confused and anxious
A bad taste in the mouth , Metallic Taste/Ammonia Breath
Feeling Cold
Dizziness and Trouble Concentrating
Leg/Flank Pain

We classify this data set into four classes (Normal, The Mild Renal Impairment (MRI), Renal Failure (RF) and Gout (Gou)). Table 4 shows the classification accuracy of PSO-Classifier for renal disease data set.

Table 4 : classification accuracy of PSO-Classifier for renal data set

	Four classes	
	train	**test**
Correct rate	98.4837	97.5633

In the PSO-Classifier algorithm for two data sets, we choose 50 particles. Because these datasets is a high dimensional problem space, increasing the particle number in the algorithm can increase the chance for finding the optimal solution. In the PSO algorithm, the inertia weight w is initially set as 0.72 and the acceleration coefficient constants c1 and c2 are set as 1.4. Also, the Maximum number of iterations is 100.

Conclusion

In the paper, particle swarm optimization Based Classifier is presented For classification medical Data Sets (thyroid data , renal data).The PSO is a is an efficient global optimizer for classification. The advantages of the PSO-Classifier are very few parameters to deal with and the large number of processing elements, so called dimensions, which enable to fly around the solution space effectively. The experimental results indicate that the PSO-Classifier method can achieve high classification accuracy (94.2051% and 97.5633%, see Table 2,4). This high efficiency is achieved with 44 and 23 features for thyroid and renal data sets respectively. From the obtained experimental results, it is obvious that our PSO-classifier reaches best correct rate and is possessed of the fastest convergence and the highest stabilities of results.

References

[1]. S.C. Chu, H.C. Huang, J.F. Roddick, and J.S.Pan , "Overview of Algorithms for Swarm Intelligence," © Springer-Verlag Berlin Heidelberg ,P. Jędrzejowicz et al. (Eds.): ICCCI Part I, LNCS 6922, pp. 28–41,2011.

[2]. L.Y.Chuang, Y.D. Lin, and C.H.Yang, "An Improved Particle Swarm Optimnization for Data Clustering," Proceedings of the International MuliConference of Engineers and Computer Scientists ,volI,IMECS March 14-16,Hong Kong,2012.

[3]. A. Ebrahimzadeh and A.Khazaee, "Detection of premature ventricular contractions using MLP neural networks: A comparative study, Measurement," vol.43, pp. 103–112, 2010.

[4]. J.Kennedy and R.Eberhart, "Particle swarm optimization, Proc.IEEE Int. Conf. Neural Networks," pp. 1942– 1948, 1995.

[5]. A.Khazaee, "Heart Beat Classification Using Particle Swarm Optimization," I.J. Intelligent Systems and Applications, 06, 25-33 Published Online in MECS. vol 06, pp.25-33,2013.

[6]. M. Korurek and B. Dogan,ECG, "Beat classification using particle swarm optimization and radial basis function neural network," 2010 Elsevier Ltd. All rights reserved, Expert Systems with Applications , 37, pp.7563-7569,2010.

[7]. S.Osowski, T.Markiewicz and L.T.Hoai, "Recognition and classification system of arrhythmia using ensemble of neural networks, Measurement," vol. 41, pp. 610–617,2008.

[8]. K. Premalatha and A.M. Natarajan, "A New Approach for Data Clustering Based on PSO with Local Search, Computer and Information Science,", vol. 1, No. 4,pp 139-145, 2008.

[9]. S.Mahapatra, A.K.Jagadev and B.Naik, "Performance Evaluation of PSO Based Classifier for Classification of Multidimensional Data with Variation of PSO Parameters in Knowledge Discovery Database," International Journal of Advanced Science and Technology ,vol. 34, pp.27-44,2011.

[10]. J.O.A. Mohamed and R.Sivakumar, "Ant-based Clustering Algorithms: A Brief Survey," International Journal of Computer Theory and Engineering1793-8201, vol. 2, No. 5, October, pp787-796, 2010.

[11]. J.Zhang ,Y. Wang and J.Feng "Parallel Multi-Swarm PSO Based on K-Medoids and Uniform Design ," Research Journal of Applied Sciences, Engineering and Technology 5(8): 2576-2585, © Maxwell Scientific Organization ,2013.