Comparison of Fully Connected Net with Particle Swarm Optimization Neural Network and PSO in the Diagnosis of Heart Disease

Johannes K, Chiang
Department of Management Information Systems
National Chengchi University,
sinoisus@gmail.com

Renhe Chi
Department of Management Information Systems
National Chengchi University
renhechi@gmail.com

Abstract

This paper proposes an Enhanced Hybrid Particle Swarm Optimization (PSO) with Artificial Neural Network (ANN), which is applied in the diagnosis of heart disease of the common features in University of California, Irvine (UCI) dataset. This UCI data includes 303 test results and consist of 13 features with two classes. One class is with health people and the other class of people are with heart disease. PSO-ANN combined Particle Swarm Optimization (PSO) and Artificial Neural Network (ANN), using ANN's escaping mechanism to enhance the deficiency of PSO slow convergence and easy to fall into the local optimal solution. The overall search ability is increased and the tracking time is reduced. This paper uses fully connected net with PSO-ANN with Python environment compares with PSO in R, the result demonstrates that the proposed model is better than PSO around 12%.

Keywords- Artificial Neural Networks (ANN), Particle Swarm Optimization, PSO-ANN, Fully Connected, Heart Disease.

I. INTRODUCTION

Cardiovascular diseases in most of the cases are the major cause of death in most countries [1]. According to the latest statistics by World Health Organization (WHO), over 37% of world deaths are effects by cardiovascular disease.

In the recent years, the study of the combination of the up to date metamorphic algorithms and data searching approaches to improve detection of the syndrome is very popular. The collecting of data mining techniques and analysis improvement which based on

clinic data can help doctors to use the data more efficient (alkeshuosh 2017). PSO algorithm is a class of optimization algorithms inspired by nature.

Including genetic algorithm, particle swarm optimization algorithm, cuckoo search algorithm, particle swarm optimization(PSO-ANN), by comparing the obtained results, it is indicated that PSO-ANN algorithm is not with the highest accuracy method. With developing models using new parameters, it is increased around 12% to reach more accurate and reliable results to diagnosis the probability of heart disease with less error. However, in our traditional method, LDA has the best accuracy rate. The results could help medical specialists, hospitals and emergencies for real practice selection.

II. PREVIOUS WORK

In this part, we analysis previous works in the area based on the Heart Disease Dataset (HDDB) and related data mining and analysis theories.

Arthor	Method	Main works
A.Dewan and M. Sharma	Fuzzy Logic system and chaos	This model is a combination of fuzzy c- means cluster algorithm with parameters tuned by genetic chaos with firefly algorithm
Mahmoodabai and S. Shaerbaf Tabrizi	Decision tree	The fuzzy expert system is used to classify data for hear disease.
A. Dewan and M. Sharma	Genetic algorithm and neural network	Using this model to identify and extract unknown knowledge related to heart disease database.
F. Ahmad, N. Ashidi Mat Isa, Z. Hussain, and M. Khusairi Osman	Gentic algorithm-Multilayer Perceptron(GA-MLP) method	This improved genetic algorithm is used to determine the optimization parameters automatically and simultaneously through evolutionary process.
C. Cortes and V. Vapnik	Support vector machine(SVM)	SVM is a linear classification methods separating data by building a hyper plan.
S. Mahmoudi, R. Rajabioun, and S. Lotfi	Binary cuckoo optimization (BCOA)	This method is converting continueous Cuckoo optimization algorithm(COA) to binary version.
Duda et al., 2000; Fukunaga, 1990;Hastie et al., 2001	Linear Discriminant Analysis	A well-know method for dismensionality reduction and classification that projects high-dimensional data onto a low-dimensional space where the data achieves maximum class.
Azhar Hussein Alkeshoush and Mariam Zomorodi Moghadam	Particle swarm optimization.	Particle Swarm Optimization (PSO) is a swarm intelligence technique, which is very similar to evolutionary computation approaches.
L.K. Gaddala and N. N. Malleswara Rao,	Artificial neural network(ANN) optimized by particle swarm optimizatopn(PSO)	PSO is a global search and population-based algorithm which has been used for training neural networks, finding neural network architectures, tuning network learning parameters, and optimizing network weights.

Table 1. Heart Disease Dataset(HDDB) related analysis theories.

Type2 fuzzy logic system (IT2FLS) and chaos algorithm for broad data sets merges of fuzzy c-means clustering algorithm with the constants tuned by genetic chaos with firefly algorithm is suggested that can cut down features correlated with the other machine learning methods such as simple bayesian, support vector machine and neural network. [2].

The decision tree administers fuzzy expert system is applied to label and lessen the size of data and contribute well-founded rules [3]. The improvement of the fuzzy enrollment functions is finished by adopting Independent Component Analysis (ICA) and upgraded ICA.

The key aspiration adopting the genetic algorithm and neural network was establishing a model to classify and draw out unknown understanding related to heart disease database [4]. With optimization, the defect of Back Propagation (BP) algorithm that obtaining stuck in local optimum is solved. The neural network enhances genetic algorithm performance.

Genetic Algorithm-Multilayer Perceptron (GA-MLP) method is the advance genetic algorithm introduced for medical detection to regulate the optimization parameters include some hidden nodes, initial weights and feature subsets of multi-layer consciousness naturally and simultaneously through evolutionary process [5]. GA-MLP algorithm has the advantageously accuracy among previous effort for diabetes, heart disease and cancer.

The recognized method for dimensionality scaling down and categorization, for binary classes, Linear Discriminant Analysis (LDA) has been demonstrated to be the same as linear regression with rating tag as output which can be formulated as a least squares problem [6].

III. PROPOSE METHOD

A. PSO Algorithm

Particles swarm optimization (PSO) is developed by Kennedy and Eberhart in1995 [7] [8]. PSO is an approach for optimization on aped of social behavior of flock of birds and/or groups of fish. Particles mimicked the creation accomplishing excellent advancement. Stand on inhabitants close to genetic algorithms (GAs). PSO is adopted solving many nonlinear function minimization, no-differentiable, discontinuous, multi model and game theory issues in science and engineering problems [9]. In PSO, a solution is described as a particle, and the population of solutions is termed a swarm of particles. PSO algorithms are simple to fulfill and reach global optimal solutions with high feasibility.

B. ANN

With capable computing power and memory, as a result educated from human brain functioning, ANN has been considered and developed for numerous distinct application motivations sometimes it is even more clever than humans in several utilizations. The information will be evaluates via a deportation operation and conveyed over a hidden layer consist of hidden neurons. Data is encrypted, evaluated and deliberated by weights in the hidden layer. The deviation is also estimated to guarantee a stable data level. The result is calculated on the output layer [10].

C. PSO in ANNs

PSO in ANNs: PSO is a global search and population-based algorithm accustomed to

training neural networks, searching neural network architectures, adjust network learning parameters, and optimizing network weights. Not based on gradient information, PSO avoids trapping in a local minimum [11]. To get the best set of weights is the function of PSO in ANN that several particles are trying to move to get the best solution. The dimension of the search space is the sum number of weights and biases.

D. Fully Connected layers in a neural networks are those layers where all the inputs from one layer are connected to every activation unit of the next layer. In the most popular machine learning type, the last few layers are full connected layers which organize the data obtained by previous layers to generate the final output. It is the second most time consuming layer second to Convolution Layer.

Feature	Description
Age	Age in years
Sex	1 = male 0 = female
Cp	Chest pain type: 1 = typical angina 2 = atypical angina 3 = non-anginal pain 4 =asymptomatic
Trestbps	Resting blood pressure (in mm Hg)
Chol	Serum cholesterol in mg/dl
Fbs	Fasting blood sugar > 120 mg/dl: 1 = true 0 = false
Restecg	Resting electrocardiographic results: 0 = normal 1 = having ST-T wave abnormality 2 =showing probable or defines left ventricular hypertrophy by Estes'criteria
Thalach	Maximum heart rate achieved
Exang	Exercise induced angina: 1 = yes 0 = no
Oldpeak	ST depression induced by exercise relative to rest
Slope	The slope of the peak exercise segment : 1 = up sloping 2 = flat 3= down sloping
Ca	Number of major vessels colored by fluoroscopy
Thal	Heart condition: 3: normal 6: fixed defect 7: reversible defect
Num	0: if less than 50% diameter narrowing in any major vessel (CAD, no) 1: if more than 50% (CAD, yes)

IV. THE USE DATABASE

In this study, University of California, Irvine (UCI) dataset is used to evaluate the proposed method [13]. This data includes test results of 303 people. Number of common features for the detection of heart disease, refer to World Health Organization standard, is 76.

All of these features are not functional and the functional features should be elected. This dataset consists of two classes, one class for healthy people and the other class for people with heart disease. In Table. 2 the features of this data set are listed.

V. RESULT

With PSO-ANN the accuracy average rate is at 72.45% which lower than previous study but the best accuracy rate we got is at 83.82% which is higher than 77.41%. We found that with ANN added up, the accuracy rate can increase around 12.20%.

				Best	Average	Average
	Previous PSO-ANN	LDA	QDA	PSO-ANN	PSO-ANN	PSO
Accuracy	77.41%	90.12%	88.89%	83.82%	72.45%	60.25%

Comparing with previous studied, for the traditional method we also added on LDA and QDA, the accurate in LDA is 90.12% and in QDA is 88.89%.

VI. DISCUSSION

The previous works using PSO-CNN has better performance than our model. Hence, we still checking for the improvement for existed PSO-ANN algorithm especially for the parameter of PSO and ANN itself.

However, there has other study shows that when the neural network with number of attributes performs much better than the characters is 13[15].

The accuracy results could hardly improve when reaches 3 hidden layers, Besides the reason not enough attribute, sample size might also not enough. We should try other heart related dataset to test this view point.

For further study in this field, we select cardio and heart_statlog_develand_hungary open data sets. By compare with different data sets with either more sample size or more features, we can have clear picture relate to cardiovascular diseases.

References

[1] Mayo Clinic. 2014. *Diseases and Conditions Heart disease*. http://www.mayoclinic.org/diseasesconditions/heart disease/ basics/ definition/ con20034056.

[2] N. Cong Long, P. Meesad, and H. Unger, *A highly Accurate Firefly Based Algorithm for Heart Disease Prediction*, Expert Systems with Applications, vol. 42, pp. 8221-8231, 2015.

[3] Z. Mahmoodabai and S. Shaerbaf Tabrizi, *A New ICA-Based Algorithm for Diagnosis of Coronary Artery Disease*, Intelligent Computing, Communication and Devices, vol. 2, 2014, p. 415-427.

[4] A. Dewan and M. Sharma, *Prediction of Heart Disease Using a Hybrid Technique in Data Mining Classification*, Computing for Sustainable Global Development (INDIACom), 2015, pp. 704-706.

[5] F. Ahmad, et al., *Intelligent Medical Disease Diagnosis Using Improved Hybrid Genetic Algorithm-Multilayer Perceptron Network*, Journal of Medical Systems, vol. 37, pp. 9934, 2013.

[6] Jieping Ye, *Least squares linear discriminant analysis*, Proceedings of the 24th international conference on Machine learning, 2007, pp.1087-1094.

[7] Russell Eberhart and James Kennedy, *A New Optimizer using Particle Swarm Theory*, Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995, p.39-43

[8] Chunkai Zhang and Huihe Shao, *Particle Swarm Optimization in Feed Forword Neural Network in Artificial Neural Networks in Medical Biology*, Springer-2000, pp 327-332.

[9] Haufen Yang, et al., *An Improved particle swarm optimization algorithm*. ScienceDirect, Volume 193, Issue 1, 1 October 2007, pp. 231-239

[10] P. Srikala and Shaik Umar Faruq, *Neural Network based Image Compression with Lifting Scheme and RLC* , International Journal of Reseach In Engineering and Technology, ISSN : 2319-1163

[11] L.K. Gaddala and N. N. Malleswara Rao, *An Analysis of Heart Disease Prediction Using Swarm Intelligence Algorithms*, International Journal of Innovactions in Engineering and Technology, Feb 2018, pp. 081-087.

[12] Nasser Mohammadi and Seyed Javad Mirabedini, *Comparison of Particle Swarm Optimization and Backpropagation Algorithms for Training Feedforward Neural Network*, Journal of Mathematics and Computer science 12 , 2014, pp. 113-123

[13] https://archive.ics.uci.edu/ml/datasets/Statlog+(Heart).

[14] C. S. Dangare and S. S. Apte, "Improved study of heart disease prediction system using data mining classification techniques," International Journal of Computer Applications, vol. 47, pp. 44-48, 2012.