Inter national Journal of Pure and Applied Mathematics
Volume 113 No. 11 2017, 208 – 216
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

A Comparison of Fuzzy Genetic and Neuro Genetic Hybrid Algorithm for the Classification of Diabetes Disease

S.Poongothai[1,*], C.Dharuman[1] and P.Venkatesan[2]
[1] Department of Mathematics, SRM University,
Ramapuram Campus, Chennai, India.
[2] Sri Ramachandra University,Chennai, India.
*Corresponding author:poongothaikannan25@gmail.com

February 8, 2017

Abstract

Computational Intelligence methods like fuzzy logic, genetic algorithms and artificial neural networks can deal with complex problems which are very difficult to explain in normal classical methods. This paper proposes a comparison of fuzzy and hybrid techniques like fuzzy genetic and neuro genetic for the diagnosis of diabetes on a Pima Indian Diabetes dataset from UCI machine learning repository. Diabetes is one of the major challenging disease to the world. We also extend the hybrid algorithm of combining genetic with neuro fuzzy. The hybrid techniques performs better than the other normal methods in classification and also with lesser time.

AMS Subject Classification:03B52, 03E72.

Key Words and Phrases: Fuzzy logic, Genetic Algorithm, Artificial neural networks, hybrid techniques, Diabetes.

1 Introduction

In medical diagnosis the most commonly used technique is artificial intelligence [1, 2, 3].Diabetes is one of the major challenging disease to the human society which is complicated to diagnose in the earlier stage. Insulin is wholy responsible for regulating glucose level in the human body. Improper production of insulin causing diabetes. It can be controlled by proper food habits and exercise since there is no medicine for complete curing the disease. In this paper dataset for diabetes is taken from UCI machine learning repository. Genetic algorithm is used to extract the feature set and combined with fuzzy to find the optimal solution [4] and also combined with neural networks. The performance of the proposed method is measured using classification accuracy, confusion matrix and 10 fold cross validation.

1.1.Related works

For the prediction and classification of medical data so many techniques are handled by many researchers. Polatet et al. proposed two types of approaches for diabetes data classification - principal component analysis and neuro-fuzzy inference, Generalized Discriminant Analysis (GDA) and least square support vector machine (LS-SVM). The accuracy for their work are 89.5% and 79.2% respectively[5]. Muni, Pal, Das [6] presented a method for multiclass classifier and introduced a new concept of unfitness for improving genetic evolution. Hasan Temurtas et al.[7] proposed a neural approach for classification of diabetes data and achieved 82.4% accuracy. Pradhan et al. combined Comparative Partner Selection (CPS) along with GP to design a 2-class classifier for detecting diabetes. Cheung used C4.5, Naive Bayes, BNND and BNNF algorithms and reached the classification accuracies 81.1%, 81.5%, 81.1% and 80.9%, respectively [8]. Ephzibah [9] presented a fuzzy rule based classification system for feature subset selection for cost reduction in diabetes diagnosis. Arcanjo et al. proposed a KNN-GP (KGP) algorithm, a semi-supervised transductive one, depends on three basic assumptions of semi-supervised learning. This system was implemented on 8 datasets of UCI repository but results were obtained for diabetes dataset are not satisfactory[10].

1.2.Database

The dataset for this paper (Pima Indian diabetes data) was taken

from UCI machine learning repository database [11]. Out of 768 cases 268 cases were in class 1(positive test for diabetes) and 500 cases were in class 0(negative test for diabetes). Nine attributes were chosen for our work, the details as given below in Table 1.

Table 1: Description of attributes used in PIMA diabetes

S.No	Attributes	Description
1	Preg	Number of times pregnant
2	Plas	Plasma glucose concentration at 2 hours in an oral glucose tolerance test
3	Pres	Diastolic blood pressure
4	Skin	Triceps skin fold thickness
5	Insu	2 hours serum insulin
6	BMI	Body Mass Index
7	Pedi	Diabetes pedigree function
8	Age	Age (years)
9	Class	Class variable (0 or 1)

2 Machine Learning Methods

2.1.Fuzzy Logic

Based on Fuzzy logic model, many diagnostic systems in medicine have been developed and performing well in the diagnosis and treatments of various diseases like diabetes, cancer and many more [12]. It is an excellent mathematical tool to handle the uncertainty arising due to vagueness. It is a set of mathematical principles based on degrees of membership for knowledge representation. Fuzzy logic is used for effective prediction of the disease. Many classifier systems available in the research world.Out of those, most widely used fuzzy classifier systems are FCM (Fuzzy C-means), FIS (Fuzzy Inference System) and FGP (Fuzzy Grid Partition). FCM is one of the data clustering technique in which a dataset is grouped into n clusters with every point in the dataset belongs to every cluster to some degree. FIS follows the rule as "If antecedent then consequent" [13]. In the simple fuzzy partition grid methods, each attribute can be partitioned by various linguistic values. Grid partitioning forms a partition by dividing the input space into several fuzzy slices, each

of which is specified by a membership function for each feature dimension [14].

2.2.Genetic Algorithms

Genetic Algorithms (GAs) is an iterative and stochastic process that operates on a set of individuals (population). Each individual represents a potential solution to the problem being solved. Initially, the population is randomly generated and every individual in the population is assigned, by means of a fitness function, a measure of its high quality with respect to the problem taken. The three main operators of GAs are selection, crossover and mutation. The genetic algorithm uses three main types of rules at each step to create the next generation from the current population (1) *Selection rules* select the individuals at random in the initial stage, called *parents* which contribute the population at the next generation. (2) *Crossover rules* combine two parents to form two children for the next generation. (3) *Mutation rules* apply random changes to individual parent to form a single child.

2.3.Neural Networks

Neural networks, oftenly called Artificial Neural Networks (ANNs) is one of the computational models that contains many simple processing units which communicates by sending signals to one another over a large number of weighted connections. The idea of ANNs originally developed from the inspiration of human brains. The special feature of these networks is its adaptive nature i.e., "learning by example" replaces "programming" for solving any kind of problems. ANNs imitate the learning process of the human brain and can process problems involving complex and non-linear data even if the data are noisy and imprecise. In this paper multilayer perceptron technique is used. A multilayer perceptron (MLP) is a feed forward artificial neural network model which maps sets of input data to a set of appropriate outputs. It consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one.

3 Results

In this paper Genetic algorithm is used to select optimal subset of the attributes using WEKA tool. The GA gave three features

namely preg, plas and pres. We have also normalized using the following formula

$$l_{norm} = \frac{l - l_{\min}}{l_{\max} - l_{\min}}$$

where l is the real data, $l_{\min}$ be the minimum value in l, $l_{\max}$ be the maximum value in l and l_{norm} is the normalized value. Once this normalization is applied to all attributes, the next thing is to find the subset of attributes which could improve the accuracy of the problem. For this purpose Genetic Algorithms is used. It reduces the attributes from 8 to 3 which are more useful to predict the disease. The reduced attributes are shown in Figure 1. Figure 2 and Figure 3 shows results of fuzzy and fuzzy with GA respectively. Figure 4 and 5 shows results of ANNs and ANNs with GA respectively. The metrics used to check the performance of pro-

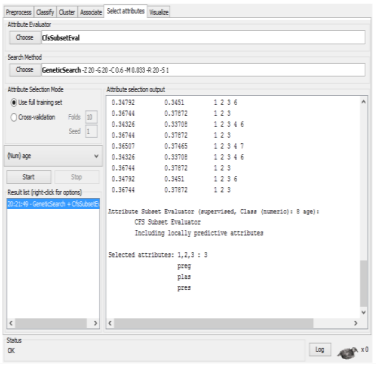

Figure 1: Features selection by GAs.

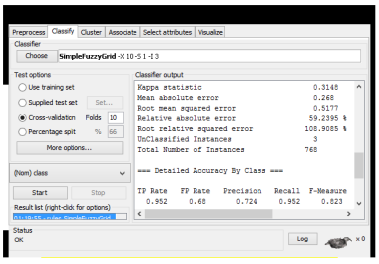

Figure 2: Results obtained by Fuzzy.

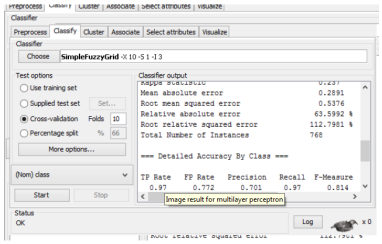

Figure 3: Results obtained by Fuzzy with GA.

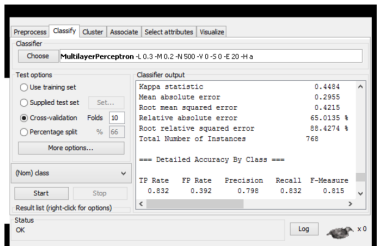

Figure 4: Results obtained by ANNs.

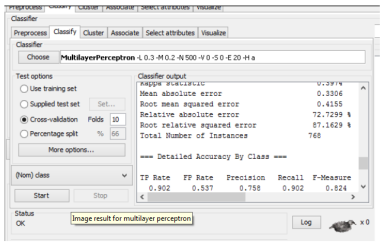

Figure 5: Results obtained by ANNs with GA.

posed algorithm are sensitivity, specificity and accuracy. Table 2 shows the performance of hybrid and non-hybrid techniques. From this table we can conclude that fuzzy and ANN performs well with Genetic Algorithms. Comparing these two (ANN and Fuzzy) Fuzzy with GA works well.

Table 2: Comparative Study of Hybrid and Non-Hybrid Techniques

Algorithms	TP Rate (%)	Mean absolute error	Time taken to build the model(sec)
Fuzzy	95	0.2681	0.28
ANNs	83	0.2955	0.86
Fuzzy with Genetic Algorithms	97	0.2891	0.06
ANNs with Genetic Algorithms	90	0.3306	0.45

4 Discussions and Conclusion

This paper shows the comparative study of hybrid and non-hybrid techniques and it is obviously seen that fuzzy system performs well with genetic algorithm than with neural networks. Since GA reduces the attributes, the time taken to solve the problem also get reduced. Fuzzy grid partition system along with GA gives promising prediction of the disease. In future, more studies with other datasets are needed to establish the higher performance of genetic algorithm based hybrid models.

References

[1] Srimani P.K. and Koti M.S. 2014, "Knowledge discovery in medical data by using rough set rule induction algorithms", Indian Journal of Science and Technology, Jul; 7(7):905-15.

[2] Overiu M., Simon D.,2010, "Biogeography-based optimization of neuro-fuzzy system parameters for diagnosis of Cardiac disease", Proceedings of the 12th annual conference on Genetic and evolutionary computation, p. 1235-42.

[3] Rajeswari K., Vaithiyanathan V., Amirtharaj P., 2011, "Prediction of risk score for heart disease in India using Machine intelligence", International Conference on Information and Network Technology; Singapore: IACSIT Press; p.19-22.

[4] Poongothai S., Dharuman C., Venkatesan P.,2016, "Fuzzy Evolutionary Computing in Biological Data Mining", Global Journal of Pure and Applied Mathematics (GJPAM) ISSN 0973-1768 Volume 12,Number 1, p. 306-309.

[5] K. Polat, S. Gunes, A. Aslan, 2008, "A Cascade Learning System for Classification of Diabetes Disease: Generalized Discriminant Analysis and Least Square Support Vector Machine", Expert systems with applications, vol.34(1), p. 214-221.

[6] Muni D.P., Nikhil R. Pal, and Jyotirmoy Das, 2004. "A Novel Approach to Design Classifiers Using Genetic Programming", IEEE Transactions on Evolutionary Computation, Vol. 8, No. 2, p. 183-196.

[7] Hasan T., Nijat Y., Feyzullah T., 2009 "A Comparative Study on Diabetes Disease Using Neural Networks", Expert system with applications, vol.36, p.8610-8615.

[8] Polata K., SalihGnesaSlaymanTosunb, Diagnosis of Heart Disease Using Artificial Immune Recognition System and Fuzzy Weighted Pre- Processing, ELSEVIER, PATTERN RECOGNATION.

[9] Ephzibah E. P.,2011, "Cost Effective Approach On Feature Selection Using GeneticAlgorithms And Fuzzy Logic For Diabetes Diagnosis", International Journal on Soft Computing (IJSC), Vol.2, No.1, p. 1-10.

[10] Filipe de L. Arcanjo, Gisele L. Pappa, Paulo V. Bicalho, Wagner Meira Jr., Altigran S. da Silva, 2011, "Semi-supervised Genetic Programming for Classification", ECCO'11, Dublin, Ireland.

[11] ForinaM., "Pima Indian Diabetes dataset", http://archieve.ics.uci.edu/ml/datasets/ Pima+Indians+Diabetes.1991.

[12] Barro S., Marin R., 2002, "Fuzzy Logic in Medicine", Heidelberg : physica-verlag.

[13] Hong T., 2008, "MCMC Algorithm, Integrated Four-Dimensional Seismic Reservoir Characterization and Uncertainty Analysis in A Bayesian Framework", ProQuest LLC, p. 31.

[14] LeoniSharmila S., Dharuman C., Venkatesan P., 2016, "A Novel Neuro-Fuzzy System for Classification", Global Journal of Pure and Applied Mathematics (GJPAM) ISSN0973-1768 Volume 12,Number 1, p. 267-270.