INTERNATIONAL RESEARCH JOURNAL OF PHARMACY

www.irjponline.com

ISSN 2230 – 8407

Research Article

DIABETES DATA CLASSIFICATION USING WHALE OPTIMIZATION ALGORITHM AND BACKPROPAGATION NEURAL NETWORK

Rajeshkumar J *[1], Kousalya K [2]

[1]Assistant Professor, SNS College of Technology, Coimbatore, Tamilnadu, India

[2]Professor, Kongu Engineering College, Erode, Tamilnadu, India

*Corresponding Author Email: rajeshresearch88@gmail.com

Article Received on: 20/11/17 Approved for publication: 10/12/17

DOI: 10.7897/2230-8407.0811242

ABSTRACT

Diabetes also called as diabetes mellitus is a health issue which affects more people in a world. Diagnosis of diabetic problem depends on different parameters and requires experience or good algorithm to classify it optimally. Many researchers have found different classification algorithms to diagnose this health issue with promising results. In this paper, combinations of whale optimization algorithm and backpropagation neural network methodology are integrated to diagnose diabetes mellitus. This proposed method supports high convergence speed and improved accuracy. Due to this combination, local minima trapping problem which affects the quality of the solution is totally reduced. In the proposed methodology, Whale optimization technique develops new solutions in solution space and backpropagation algorithm finds the globally optimal solution. Experimental analysis compares the proposed methodology with other algorithms and finally concludes the proposed algorithm outperforms other methodologies.

Keywords: Classification, whale optimization algorithm, backpropagation neural network, diabetes data, artificial neural Network

INTRODUCTION

Diabetes is a collection of diseases which related with metabolic action. Due to this disease a person may have high blood glucose. Reason for the presence of such glucose may be inadequate insulin production or poor respond by body cells for insulin[1]. Diabetes is classified into two major types. First one is type – 1 in which the pancreas of body stops the production of insulin. In the type – 2 diabetes mellitus, body fails to use the insulin effectively which means that poor responding of body for insulin produced by the body. At the later stage of type – 2, the situation where body fails to produce enough insulin may occur. Another form of diabetes is gestational diabetes. This form of diabetes occurs on pregnant women. The women should not have the previous diabetes history. As per statistics of International Diabetes Federation, 415 million people have diabetic problem till the year of 2015[2]. So diabetes requires advanced technique for identifying the diabetes based on symptoms.

Data mining is defined as deriving meaningful knowledge or information from available huge data sets. Clustering and classification are two important data mining techniques which can be used to analysis the different sets of data and to attain meaningful conclusion at the final. Classification a supervised methodology in which assigning the different objects or groups to suitable classes[3]. In the classification algorithm, two phases are available for successful classification. First one is model construction and second one is model usage. Learning process from the available dataset is carried out by first stage. Classifier constructed using the above first phase. Classifier holds the available sample data with their corresponding labels. In the second phase, test data of the problem involved to estimate accuracy of the classifier. If the accuracy is required level, then classification rule can be used for newly entered data. Data mining techniques can improve quality of decision making, delivery method of health care and enhance the organization of disease relevant data[4]. Data processing and retrieving information from medical relevant data requires more accuracy supporting techniques to avoid major problem on human health. That is the main reason of applying data mining techniques for processing diabetes classification problem.

The main objective of the proposed algorithm is to label all available patient data based on their diabetes predicting attributes. In this paper, Pima Indians Diabetes Data set is used as training and testing data set. All the experiments are carried out using this and comparisons have been made with different algorithms. The remaining paper is organized as follows. Section 2 provides the detailed discussion on problem. It discusses with objectives, parameters and characteristics of the problem. Section 3 deals with literature survey related to diabetes classification problem and other optimization algorithm based classification problem. Section 4 will explain the working methodology, characteristics and merits of proposed algorithm. Section 5 will give detail analysis of proposed algorithm's efficiency and its comparison with other traditional algorithms. Section 6 provides the concluding remarks and possible extension works of proposed system.

PROBLEM DESCRIPTION

Objective: Labelling available data using appropriate classification algorithm with high accuracy rate
Input: Pima Indians Diabetes Data set.
Output: Dataset with labels.

Based on values of parameters each available data should be classified to relevant group and it should be assigned with some known value as a label value. Label value can clearly indicate diabetes type which that corresponding patient belongs. Problem solving methodology needs some training procedure for accurate class prediction.

RELATED WORKS

Artificial neural network (ANN) contributes more in the field of medical diagnosis. Applications of ANN on medical field were explain by Qeethara Kadhim Al-Shayea[5]. Adaptive genetic algorithm based methodology was proposed for feature selection and gene expression[6]. This method reduces the gene expression dimension and removes the redundancies in classification. Multi label method using ant colony optimization algorithm was proposed for data classification problem[7]. Gravitational search algorithm for classification problem was introduced with additional particle swarm optimization algorithm memory characteristics[8]. In this system gravitational search algorithms memory lacking problem was solved using particle swarm optimization algorithm's memory property. This combination of hybridization increases the overall efficiency of the gravitational algorithm.

Optimized radial basis functional neural network algorithm was developed to small data classification[9]. This system holds two important features such as hybrid coding and simultaneous evolving in hidden layer and it provides efficient platform for small amount of data handling applications. Improved particle swarm optimization algorithm was proposed for data classification problem with mixed attribute capability[10]. In this, new position updating mechanism was introduced to increase convergence speed and integer mapping was replaced by non numerical data method. Solving classification problem integrated the characteristics of chaotic particle swarm optimization algorithm and least square support vector machine[11]. Performance of the above method was tested on three different dataset such as binary classification data, IRIS data and three drug data. Absence of choice randomness and good efficiency in workload selection are the important findings at the end of experimental analysis.

Classification algorithm with multiple rule set was proposed using ant colony optimization algorithm[12]. Correlation and convergence was involved with new heuristic function is one the features. To classify the test data weighted voting mechanism was introduced along with the ant colony algorithm. Multi objective based optimization algorithm was developed for microarray analytic hierarchy process[13]. Two important rules were introduced such as Higher and fewer rule and forcibly decrease rule in this method. Genetic algorithm and wavelet based classification algorithm was delivered for cancer diagnosis application[14]. Different kinds of feature set were explored by this hybrid methodology. Artificial bee colony based classification method was developed for the remote sensing related data[15]. This algorithm was tested on different satellite data type such as IRS P-6 and IRS P-5.

Multi class imbalanced data classification problem was solved using binary particle swarm optimization algorithm and K-Nearest neighbour algorithm[16]. To test the efficiency of an algorithm nineteen benchmarks were considered and applied the algorithm. Microarray medical data classification method adopted the characteristics of cat swarm optimization algorithm. This method can be applied for both binary and multiclass data set.

PROPOSED METHODOLOGY

Whale optimization algorithm (WOA)[18] is a metaheuristic based algorithm which mimics the bubble-net hunting strategy of humpback whales. In this strategy, the whale follows stereotypical bubbles motivates the creation of circular or 9-shaped path while circumscribing victim during hunting. Shrinking encircling mechanism and spiral updating position are the two important mechanisms which involved in mathematical modelling of WOA. Exploration phase of this optimization algorithm is called as search for prey. In this each agent is mapped to the diabetes data problem and its parameter mapped to all WOA parameters. Initial random solutions are generated by some low level heuristic algorithms and it will considered as starting phase input of WOA.

Another important algorithm which integrated with proposed system is backpropagation neural network (BPNN)[19].BPNN is used to make the output optimally by searching good quality solution from the available solutions in solution space. Automatic weight adjusting based on the previous solutions quality will be carried out by BPNN. This process enhances the global best solution and enriches the quality of convergence speed. BPNN is a supervised learning method and back propagating principle is well suited procedure for the application where it depends on the relationship between input and output in all the cases. This is the main reason why BPNN applied for proposed data classification solving methodology. BPNN helps to increase the overall quality of a solution. Diabetic data classification requires high level accuracy. This is the one of the major reasons of using BPNN in our proposed system. The following algorithm illustrates proposed whale optimization algorithm using backpropagation neural network for solving diabetic data classification problem.

Step 1: Initialize the parameters of whale population and randomly generated diabetic data class solution to each candidate.

Step 2: Calculate fitness value for each candidate. Overall accuracy of the class is fitness function in this case.

Step 3: Identify the local best and global best solutions from available intermediate solutions.

Step 4: Update position and velocity of each whale candidate using the current local and global best values.

$$\vec{X}(t+1) = \overrightarrow{X_{rand}} - \vec{A}\vec{D} \quad (1)$$

Step 5: Feed the output weight value in BPNN. Calculate the error from the tested value samples. Adjust the weight according to the error value.

Step 6: Update the parameters of WO algorithm as well as problem domain based on feedback received from network.

Step 7: Check termination condition. Here number of iterations is considered as termination condition.

Step 8: If termination condition is false, go to step 2. Otherwise go to step 9.

Step 9: Stop the process of classification. Final solution will hold the label for all diabetic data available in data set.

In this above algorithm the number of iteration are fixed to 210. The above algorithm was tested on Pima Indian diabetes data set[20] which holds the 768 female candidates result.

EXPERIMENTAL ANALYSIS

In the experimental analysis proposed algorithms efficiency is compared with particle swarm optimization algorithm, whale optimization algorithm in terms of sensitivity, specificity and accuracy. Sensitivity is defined as the ration between true positive and actual positive. Specificity is defined as ratio between true negative and actual negative.

Sensitivity = True Positives / Positives; (2)
Specificity = True Negatives / Negatives; (3)

Figure 1 illustrates the sensitivity analysis, figure 2 illustrates specificity analysis and figure 3 illustrates accuracy comparison. From the experimental analysis, we can observe that the proposed algorithm outperforms other algorithms.

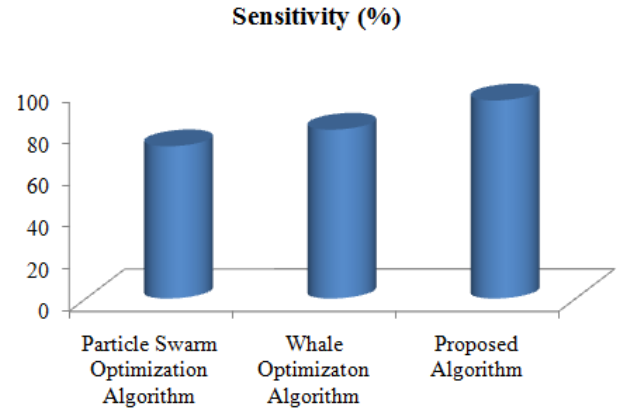

Figure 1: Sensitivity Comparison

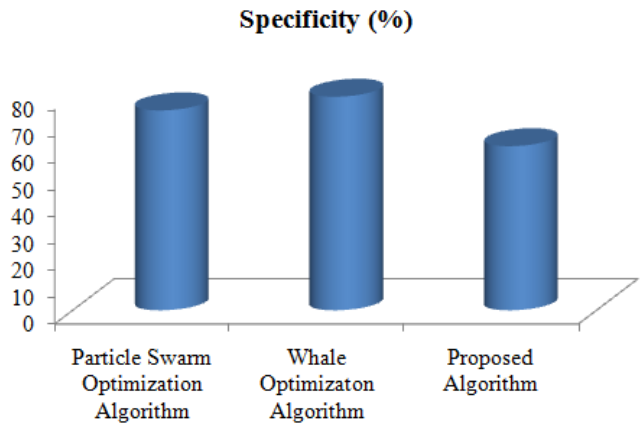

Figure 2: Specificity Comparison

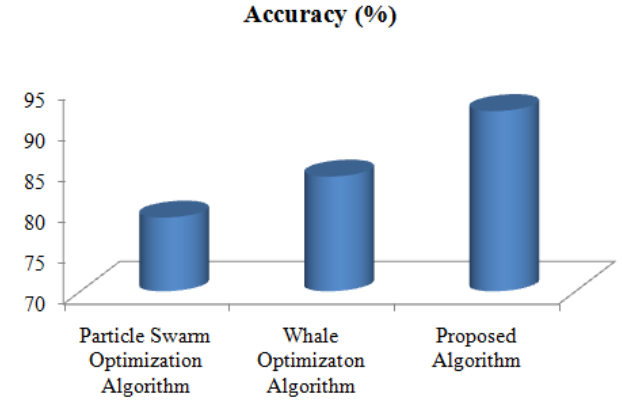

Figure 3: Accuracy Comparison

CONCLUSION AND FUTURE WORK

Whale optimization algorithm with backpropagation neural network algorithm was proposed to solve the diabetic type classification problem. Diabetes is a health issue which affects different age peoples due to different reasons. Identifying diabetes in initial stage can prevent human from different consequences. In this paper, the proposed algorithm helps system to classify diabetes based on different observations. Accuracy, convergence speed and execution time of the proposed algorithms are in efficient manner. Solution quality is very optimal than the other algorithms. In future implementation complexity of neural network can be replaced by any other simple heuristic methods which can avoid many levels of calculations.

REFERENCES

1. Type-1 diabetes. Available from: http://www.diabetes.org/diabetesbasics/type
2. Diabetes Statistics. Available from: https://www.idf.org/about-diabetes/what-is-diabetes.html
3. Han Kamber M. Data mining concepts and techniques. 2nd ed. Amsterdam, Netherlands: Elsevier Publisher 2006; 383–5.
4. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang, JF and Hua, L. Data Mining in Healthcare and Biomedicine: A Survey of the Literature. Journal of Medical Systems 2012; 36:2431 - 2448.
5. Qeethara Kadhim Al-Shayea. Artificial Neural Networks in Medical Diagnosis. International Journal of Computer Science Issues 2011; 8(2):150 – 154.
6. Lu Huijuan, Chen Junying, Yan Ke, Jin Quna, Xue Yu and Gao Zhigang. A Hybrid Feature Selection Algorithm for Gene Expression Data Classification. Neurocomputing 2017; 256:56 – 62.
7. Salabat Khan and Abdul Rauf Baig. Ant Colony Optimization Based Hierarchical Multi-label Classification Algorithm. Applied Soft Computing 2017; 55:462 – 479.
8. Binjie Gu and Feng Pan. Modified Gravitational Search Algorithm with Particle Memory Ability and Its Application. International Journal of Innovative Computing, Information and Control 2013; 9:4531 – 4544.
9. Weikuan Jia, Dean Zhao and Ling Ding. An Optimized RBF Neural network Algorithm Based on Partial Least Squares and Genetic Algorithm for Classification of Small Sample. Applied Soft Computing 2016; 48:373 – 384.
10. Nabila Nouaouri and Mounir Boukadoum. Improved Global-Best Particle Swarm Optimization Algorithm with Mixed-Attribute Data Classification Capability. Applied Soft Computing 2014; 21:554 – 567.
11. Fang Liu and Zhiguang Zhou. A New Data Classification Method Based on Chaotic Particle Swarm Optimization and Least Square Support Vector Machine. Chemometrics and Intelligent Laboratory Systems 2015; 147:147 – 156.
12. Zhengping Liang, Jiangtao Sun, Qiuzhen Lin, Zhihua Du, Jianyong Chen and Zhong Ming. A Novel Multiple Rule Sets Data Classification Algorithm Based on Ant Colony Algorithm. Applied Soft Computing 2016; 38:1000 – 1011.
13. Jia Lv, Qinke Peng, Xiao Chen and Zhi Sun. A Multi Objective Heuristic Algorithm for Gene Expression Microarray Data Classification. Expert Systems with Applications 2016; 59:13 – 19.
14. Thanh Nguyen, Saeid Nahavandi, Douglas Creighton and Abbas Khosravi. Mass Spectrometry Cancer Data Classification Using Wavelets and Genetic Algorithm. FEBS Letters 2015; 589(24):3879 – 3886.
15. Jayanth J, Shivaprakash Koliwad and Ashok Kumar. Classification of Remote Sensed Data Using Artificial Bee Colony Algorithm. The Egyptian Journal of Remote Sensing and Space Science 2015; 18(01):119 – 126.
16. Guo Haixiang, Li Yijing, Li Yanan, Liu Xiao and Li Jinling. BPSO-Adaboost-KNN Ensemble Learning Algorithm for Multi-Class Imbalanced Data Classification. Engineering Applications of Artificial Intelligence 2016; 49:176 – 193.
17. Mohapatra P, Chakravarty S, and Dash PK. Microarray Medical Data Classification Using Kernel Ridge Regression and Modified Cat Swarm Optimization Based Gene Selection System. Swarm and Evolutionary Computation 2016; 28:144 – 160.
18. Mirjalili, S. The whale optimization algorithm. Advances in Engineering Software 2016; 95:51–67.
19. Rumelhart, DE, Hinton GE, and Williams RJ. Learning Internal Representations by Error Propagation. Journal of Parallel Distributed Processing: Explorations in the Microstructure of Cognition 1986;
20. UCI Repository of Bioinformatics Databases, Website: http://www.ics.uci.edu/~mlearn/MLRepository.html.

Cite this article as:

Rajeshkumar J and Kousalya K. Diabetes data classification using whale optimization algorithm and backpropagation neural network. Int. Res. J. Pharm. 2017;8(11):219-222 http://dx.doi.org/10.7897/2230-8407.0811242

Source of support: Nil, Conflict of interest: None Declared