

Fusion of Classification with Hybrid Optimization Technique to Predict Diabetes

B.Gomathi, R.Sujatha, T.Padma

Abstract: The main objective of this paper is to predict diabetes which is growing like an epidemic in India. The key focus is to envisage diabetic and normal patient using classification approach. Fusion of SVM enhanced with hybrid optimization of PSO-BAT algorithm is proposed. Classification techniques used namely Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), Random Forest (RF) are compared with our novel approach SVM enhanced with hybrid optimization of PSO-BAT algorithm. The accuracy is increased using the combination technique. The benchmark diabetic dataset, PIMA Indian Diabetes Dataset from UCI machine learning repository is utilized for the research. To improve the efficiency more, classifiers such as Precision, Recall and f-measure is used.

Keywords: BAT, Classification, Fusion, PSO, SVM

I. INTRODUCTION

Data Fusion is grouping of more than one method. The bang of data mining towards information classification and society is tremendous because it involves integration of various techniques like neural network, machine learning, intelligent computing, pattern recognition, fuzzy logic, etc. Machine learning is a sub system of data science, which mainly focus on designing algorithms that can learn from training data and make prediction for test data. In this research Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), Random Forest (RF) classifiers are compared with our fusion model.

PIMA Indian Diabetes Dataset from UCI machine learning repository is used in this research. The cause for diabetes is increase in sugar level in blood and the symptoms include pain in legs, increase in frequency of urine discharge, tiredness, headache etc.

II. LITERATURE REVIEW

Fusion is referred as combination is used to locate feature of the software[1]. Fusion of Artificial Bee colony with fuzzy data was done on health care data to enhance the efficiency of IR[2]. Liu et al., [3] implemented RFID (radio frequency identification) to BFO for scheduling network. Swim length was adjusted dynamically and hence it is called as self-adaptive bacterial foraging optimization (SABFO).

Revised Manuscript Received on February 04, 2020.

Gomathi B, Assistant Professor, Department of Computer Science, Shri Shankarlal Sundarbai Shasun Jain College for Women, Chennai, India. gomathiphd2k15@gmail.com

Sujatha R, Assistant Professor, Department of Computer Science, PSG College of Arts & Science, Coimbatore, India. sujama2k5@gmail.com

Padma T, Professor, Department of Computer Applications, Sona College of Technology, Salem, India. padmatheagarajan@gmail.com

Sastri et al., [4] have introduced concept of hybridization which is nothing but combination and named it as velocity modulated bacterial foraging optimization technique (VMBFO). The essence of VMBFO is hybridization of BFO and PSO which helps to decrease convergence time. Initially all bacteria uses PSO and after this session they are allowed to search randomly by using BFO.

III. PROPOSED SYSTEM

The proposed classification model is boosted with accuracy to predict the diabetic patient. In this model, we have compared different classifiers with SVM enhanced with PSO-BAT hybrid algorithm. The major focus is to increase the accuracy by using fusion technique on a benchmark well renowned diabetes dataset acquired from PIMA Indian Diabetes Dataset from UCI machine learning repository.

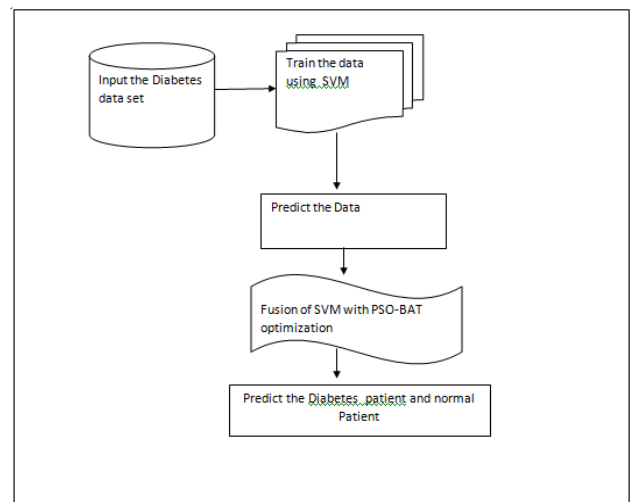


Fig. 1. Fusion model using Classification

The function of the classifier is to predict diabetes and normal person from the dataset loaded. The two subsets are training set and the test set and the former holds set of rows titled with class labels and the test set contains rows with unknown labels.

A. SVM

SVM or Support Vector Machine is a supervised learning technique and is used for data analysis. The idea of SVM is to show examples, with clear superior gap between various categories. SVM can be used for both classification and regression analysis [5]-[10].

B. Training of SVM Classifier

The inputs required to train are data, the groups and the kernel function. The function uses SVM train function where matrix is called data and each row is named as observation and column denotes one feature. Groups are column vectors and kernel function depicts linear separation of data by hyperplane.

Pseudo Code for SVM

```

Input: A and B loaded with training labeled data,  $\alpha \leq 0$ 
or
 $\alpha \leq$  partially trained SVM
C  $\leq$  some value (20 for example)
repeat
for all {ai, bi, aj, bj} do
Optimize the values ai and aj
end for
until condition is met
Output: maintain only the support vectors (ai > 0)
    
```

C. Particle Swarm Optimization

PSO is a optimization technique depending on how swarms are moved [11]. Here the agents also named as particle move in the space to find the solution which is best. It is compulsory for Particles to adjust its flying and also consider about other particles that are flying. Every particle to maintain pbest1 and gbest1. Pbest1 indicate personal best gbest1 is the global best which is nothing but the best value obtained so far by any particle who is neighbors to that particle. The basic idea is to move the particle from pbest1 to gbest1 locations. The pseudo code is illustrated below.

Function PSO()

Input: initialize position and velocity

For each particle position (xi) evaluate fitness;

If fitness(xi) is better than fitness(pbest1) then

pbest1=xi;

Set best of pbes1t as gbest1;

Update agent velocity and position;

Stop the process after evaluating gbest1 the optimal solution;

End

The advantage of PSO is selection used in GA is not required. Secondly life time of particle is entire run and it does not take into account survival of fitness. The major limitations are it cannot deal with multi dimension and always fine tuning is necessary to get result.

D. BAT algorithm

One of the extensively used metaheuristic algorithms used for global optimization is Bat algorithm. Mohd Nadhir et al [12] used bat algorithm for human pose estimation. In this paper bat algorithm is fused with PSO in order to overcome the limitations in PSO. The advantages of BAT include Frequency tuning, Automatic zooming and control on parameters [12].

Pseudo Code for BAT

Input: set the objective function

Initialize the population

Set frequency, pulse rate, loudness

While (x < maximum number of transactions)

Adjust frequency and generate new solutions
and update velocity

If (rnd > yi)

From best solution select the solution

Calculate local solution from the best selected solution

End if

Generate new solution based on random fly

If (rnd < loudness, pulse rate, frequency)

Accept new solution

End if

Current best is found by ranking bats

End while

Fusion Algorithm for Prediction

- Input the PIMA dataset
- Preprocess the Data
- Feature extraction through principle component analysis (PCA)
- Apply SVM technique
- SVM with PSO and BAT(Fusion is done)
- Achieve trained model with highest accuracy.

IV. EXPERIMENTAL RESULTS

The Diabetes dataset has total attributes 9 and total no of observations 768. A data frame with 768 observations on the following 9 variables. The bench mark data set is compared with the classifiers Multilayer Perceptron (MLP), Sequential Minimal Optimization (SMO), Random Forest (RF), The Proposed hybrid fusion algorithm SVM_PSO_BAT results in better accuracy, precision, recall and F-measure. The below Table I shows comparison of precision, Recall and F-Measure.

Table I: Comparison of Precision, Recall and F-Measure

Algorithms	Precision (%)	Recall (%)	F-Measure (%)
SMO	84	69	67
MLP	74	64	64
RF	83	69	67
SVM_PSO_BAT	86	71	70

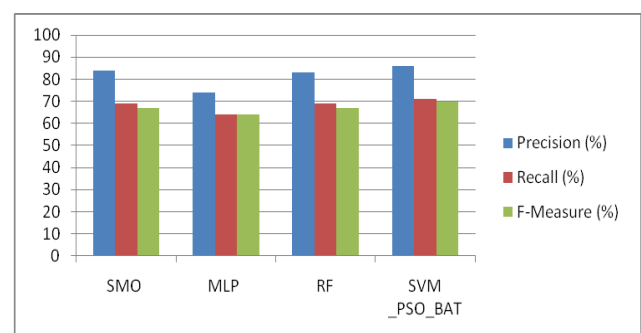


Fig. 2. Comparison of Precision, Recall and F-Measure

Table II: Comparison of accuracy of various classification algorithms.

Algorithms	Time taken in seconds	Accuracy
SMO	0.09	69%
MLP	2.18	64%
RF	0.1	69%
SVM_PSO_BAT	0.01	72%

The accuracy of the various classification algorithms is given in the Table II.

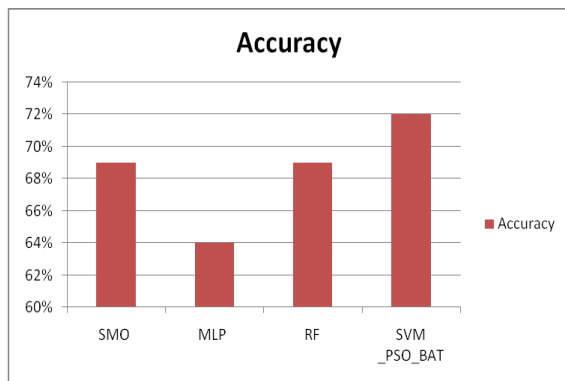


Fig.3. Accuracy Comparison

The Fig. 3. Illustrates the comparison accuracy of algorithms

V. CONCLUSION

Implementing fusion yields better result than used individually. The accuracy level of SVM fused with hybrid algorithm produces better result than other classification algorithm. The precision, recall and F-measure of the novel approach is better when compared to the existing approaches. In future we can use other optimization algorithms like cuckoo search to enhance the prediction.

REFERENCES

1. Bogdan Dit, Meghan Revelle, Denys Poshyvanyk, "Using Data Fusion and Web Mining to Support Feature Location in Software", 18th IEEE International Conference on Program Comprehension, 2010, pp.14-23.
2. Gomathi and P.Sakthivel, "Enhancing Information Retrieval process using Data Fusion by ABC Weighted Based Fuzzy Retrieval in Health care Analytic Software", Journal of Medical Imaging and Health Informatics Vol .No.6, no.3, June 2016,pp. 863-868.
3. Liu W, Chen H, Chen H, and Chen M, "RFID Network Scheduling Using an Adaptive Bacteria Foraging Algorithm", Journal of Computer Information Systems (JCIS), Vol. 7(4), 2011, pp. 1238-1245
4. Sastri GSVR, Pattnaik SS, Bajpai OP, Devi S, Sagar CV, Patra PK and Bakwad KM, "Bacterial Foraging Optimization Technique to Calculate Resonant Frequency of Rectangular Microstrip Antenna", Int. J. RF Microwave Computer Aided Eng., Vol. 18, 2008, pp. 383-388.
5. Lee, Y, Lin, Y, and Wahba, G. (2004), "Multicategory Support Vector Machine: Theory, and Application to the Classification of Microarray Data and Satellite Radiance Data", Journal of the American Statistical Association 99 (465),pp. 67–81.
6. Joachims, Thorsten, "Transductive Inference for Text Classification using Support Vector Machines", Proceedings of the 1999 International Conference on Machine Learning (ICML 1999), pp. 200-209.
7. Drucker, Harris, Burges, Christopher J. C., Kaufman, Linda, Smola, Alexander J, and Vapnik, Vladimir N. (1997), "Support Vector Regression Machines", in Advances in Neural Information Processing Systems 9, NIPS 1996, MIT Press, pp. 155–161

8. Suykens, Johan A. K., Vandewalle, Joos P. L., "Least squares support vector machine classifiers", Neural Processing Letters, vol. 9, no. 3, Jun. 1999, pp. 293–300.
9. Ferris, Michael C. and Munson, Todd S. (2002). "Interior-point methods for massive support vector machines". SIAM Journal on Optimization 13 (3), pp. 783–804.
10. Mohd Nadhir Ab Wahab, Samia Nefti-Meziani, and Adham Atyabi, "A comprehensive Review of Swarm Optimization Algorithms", Published online 2015 May 18. doi: 10.1371/journal.pone.0122827
11. Abdel-Rahman, E. M., Ahmad, A. R., Akhtar, S., (2012). A metaheuristic batinspired algorithm for full body human pose estimation, in: Ninth Conference on Computer and Robot Vision, pp. 369–375.

AUTHORS PROFILE



B. Gomathi, currently working as Assistant Professor in the Department of Computer Science at Shri Shankarlal Sundarbai Shasun Jain College for Women, Chennai. She received Ph.D Degree in Computer Science from Bharathiar University and her research includes data fusion, Internet of things and Big data. She is life member of IAENG and has published several papers in international journals.



R. Sujatha, is pursuing as Assistant Professor in the Department of Computer Science, PSG College of Arts & Science, Coimbatore, India. She was honored the Master degree in Computer Application, M.Phil. degree in Computer Science from Periyar University, India and Ph.D degree in Computer Science from Anna University, Chennai, India. Her area of interests includes Information Retrieval, Machine Learning, Big Data and Data Analytics. Published several papers in reputed National and International journals. She has participated in more than 25 Workshops, National seminars and FDP.



T. Padma, received the Master's degree in Computer Applications from the University of Madras, India in 1992, M. Tech degree in Information Technology from Allahabad Agricultural University, India in 2004 and the MPhil and Ph. D degrees in Computer Science from Mother Teresa Women's University, India in 2003 and 2010 respectively. Currently she is working as Professor, Department of Computer Applications, Sona College of Technology, Salem, India. Her research interests include Machine Learning, Data Mining and Data Analytics. Dr. Padma is a fellow of the Computer Society of India (CSI) and life member of the Indian Society for Technical Education (ISTE). She is a certified Six Sigma Black Belt. She was the recipient of the Shayesta Akhtar Memorial National Award for the Best Women Engineering College Teacher of the ISTE for her contributions in 2015. Her Biography was included in the 2010 Edition of the Marquis Who's Who in the World, New Jersey, USA. She serves as Editor and Reviewer for a number of well-known National and International journals. She has over 60 publications in journals and chapters in books. She is serving as subject expert in the Board of studies of various institutions.