

A Novel PSO-MLP framework for Feature Selection and Classification of Gene Expression Data

Lingaraj Sethi

*Department of Computer Science and Engineering
College of Engineering and Technology, Bhubaneswar, Odisha, India*

Amit Kumar Jha

*Department of Computer Science and Engineering
K.K College of Engineering and Management, Dhanbad, Jharkhand, India*

Hiresh Kumar Chandrakar

*Department of Computer Science and Engineering
Krupajal Engineering College, Bhubaneswar, Odisha, India*

Abstract- An Artificial Neural Network (ANN) provides solutions to various complex problems. Due to the massively parallel processing capability, ANN exhibits generalization capability which can be utilized to learn the patterns discovered in the data set. This can be represented in terms of a set of rules which can be used to find the solution to a classification problem. The high dimensionality of the datasets degrades the learning ability of the ANN. Hence, to minimize this risk we have used Principal Component Analysis (PCA) which provides a feature reduced dataset to the Multi Layer Perceptron (MLP), the classifier used. Again, since the weight matrices are randomly initialized, hence, in this paper we have used Particle Swarm Optimization (PSO) method to update the weight values of the weight matrix. From the experimental evaluation, it was found that using PSO with the MLP classifier provides better classification accuracy as compared to when the classifier is used alone.

Keywords – Classification, Artificial neural network, Multi-layer perceptron, Principal component analysis, Particle swarm optimization

I. INTRODUCTION

Progress in the domain of machine learning and data mining have helped the biomedical researchers to improve the quality of healthcare [1]. Today biomedical informatics has number of applications to solve various real world problems. One such problem is the classification of gene expression data. Classification [2] is defined as the task of identifying the sub-classes to which new observations may belong on the basis of data containing observations whose sub-classes are known. Zuyi wang *et al.* [3] have described diagnostic classification as the task of assigning a particular unknown sample to a known disease class based on the expression levels of gene expression data. Classification techniques on the gene expression data can be implemented by one of the various methods such as decision tree, artificial neural network, rough sets and bayesian methods [2]. In this paper, we have used MLP as the classifier. MLP's are trainable algorithms [4] that can learn to solve complex problems because of their massively parallel processing capability, fault tolerance, self-organisation and adaptive capability, which guarantees of high classification accuracy. The performance of a classifier is highly degraded when applied with gene expression data due to the curse of dimensionality of such datasets. Hence, feature reduction techniques [5] are applied on these datasets which selects the relevant features from the dataset. In this paper, PCA has been used for feature reduction. This feature reduced dataset is used to train the MLP. A typical neural network consist of a couple of hundred of weights whose value must be found to produce an optimal solution. Hence, in this paper we have employed a bio-inspired optimization technique called particle swarm optimization (PSO) [6] which optimizes the synaptic weights between the neurons. The algorithm is based on a set of potential solutions which evolves to find the global optimum of a real-valued function (fitness function) defined in a given space (search space). The PSO algorithm uses a population of particles which move in a multidimensional space that represents the space of solutions for the

problem. The layout of this paper is as follows; section II deals with proposed algorithm, in section III experimental evaluations and results are described and finally, section IV deals with conclusion and future work.

II. PROPOSED ALGORITHM

A. Particle Swarm Optimization(PSO) as a process –

Particle swarm optimization (PSO) is a population-based search technique and motivated by the social behavior of organisms such as bird flocking and fish schooling. It is originally proposed by Kennedy and Eberhart [7] for continuous problems and then was extended to discrete problems by Kennedy and Eberhart. It is well suited for combinatorial optimization problems in which the optimization surface possesses many local optimal solutions. The underlying phenomenon of PSO is that knowledge is optimized by social interaction and thinking is not only personal but also social. The particles in PSO resemble the chromosomes in genetic algorithm. However, PSO is usually easier to implement than the GA as there are neither crossover nor mutation operators in the PSO and the movement from one solution set to another is achieved through the velocity functions. We refer the reader to [8-11] for a recent review of the applications and variations of the PSO. PSO is based on the principle that each solution can be represented as a particle in a swarm. Each particle has a position and a corresponding fitness value evaluated by the fitness function to be optimized. The particles iterate (fly) from one position to another according to their most recent velocity vector. This velocity vector is determined according to the particle's own experience as well as the experience of other particles by using the best positions encountered by the particle and the swarm. Specifically, the velocity vector of each particle is calculated by updating the previous velocity by following two best values. The first best value is the particle's personal best value (pbest) (i.e., the best position it has visited thus far) and is tracked by each particle. The other best value is tracked by the swarm and corresponds to the best position visited by any particle in the population. This best value is called the global best (gbest). The effect of personal best and global best on the velocity update is controlled by weights called learning factors. Through the joint self and swarm-based updating, the PSO achieves local and global search capabilities where the intensification and diversification are achieved via relative weighting. The process is described by two different method described as follows:

Global best PSO (gbest)

For the global best PSO, the neighborhood for each particle is the entire swarm. The social networking employed by gbest PSO reflects the star topology, where the social component of the velocity equation reflects the information obtained from the entire swarm. In this case, the social component is the best position found by the swarm, represented as $\hat{y}(t)$. For gbest PSO, the velocity of particle i is calculated as:

$$V_i(t+1)=V_i(t)+C_1R_1(t)[V_i(t)-X_i(t)]+C_2R_2(t)[Y_i(t)-X_i(t)] \quad (1)$$

Where,

- (t) Is velocity of particle i in a given dimension at time t ,
- (t) Is the position of particle i in a given dimension at time t ,
- C_1 and C_2 are positive acceleration constants,
- $R_1(t)$, $R_2(t)$ are random values in the range [0, 1], generated at time t ,
- (t) Is the best position so far found by particle i .

Local best PSO (lbest)

The local best PSO, lbest, uses a ring social network topology, where smaller neighborhoods are defined for each particle. The social component reflects the information exchanged within the neighborhood of the particle. Thus, the velocity update equation is modified as follows:

$$V_i(t+1)=V_i(t)+C_1R_1(t)[Y_i(t)-X_i(t)]+C_2R_2(t)[Y_i(t)-X_i(t)] \quad (2)$$

Where,

- $Y_i(t)$ is the best position found by the neighborhood of particle i in a given dimension.

The two versions of PSO algorithms are similar in the sense that the social component of the velocity updates causes both to move towards the global best.

B. Proposed Model –

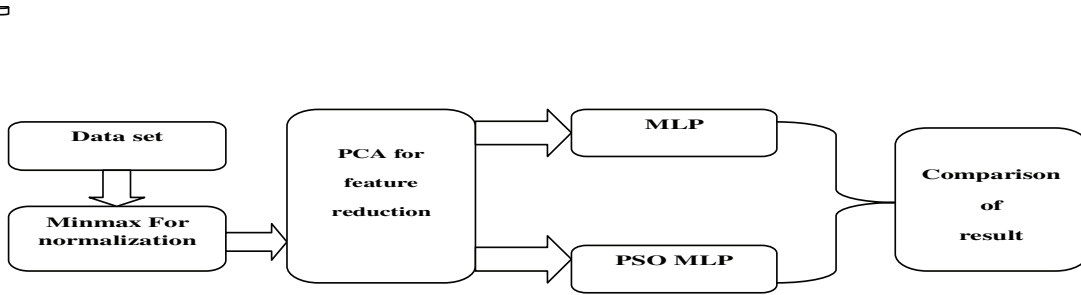


Fig 1: Proposed model

In this proposed model, the gene expression dataset is normalized by using min-max normalization. This normalized dataset is provided to PCA for feature reduction, which reduces the dimension of the dataset. This reduced dataset is provided as input to the two classifiers MLP and PSO-MLP. Finally, the accuracy of the two individual classifiers are measured and compared.

III. EXPERIMENTAL EVALUATION AND RESULT ANALYSIS

We have used two benchmarked datasets downloaded from UCI machine learning repository [12] in our experiment which are described in table 1. In this work, we have used MATLAB version 7.10, release name- R2010a. The experiment was carried on Intel core i3 processor, 2.4 GHZ, 32 bit 1GB RAM, 1GB disk space for MATLAB, 3-4 GB for ideal installation.

TABLE I: Description of datasets

Data set Name	Dimension
Yeast	1484*17
Pima Indian Diabetes	768*9

The total experimental evaluation has been carried out in the following steps:

Step1: Collection of datasets: Two data sets as describe in table 1 has been collected and processed for further processing. Yeast data set has 1484 instances and 17 samples, whereas, Pima Indian Diabetes data set contains 768 instances and 9 samples as shown in fig.2 and fig.3 for Pima Indian Diabetes and Yeast respectively.

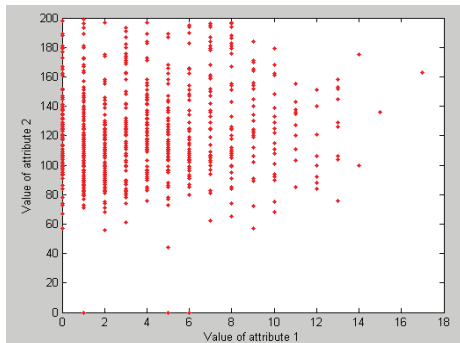


Fig 2: Pima Indian Diabetes data set

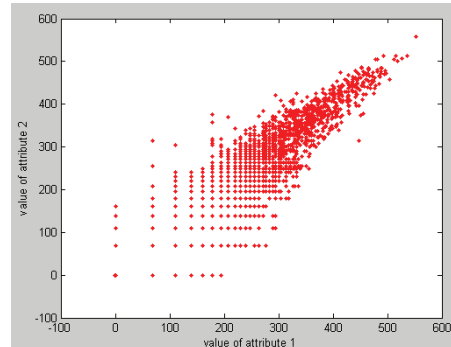


Fig 3: Yeast data set

Step2: Normalization of datasets: Data normalization [13] is an important step in the knowledge discovery process, can be even considered as a fundamental building block of data mining. The attribute data is scaled to fit in a specific range. There are many type of normalization available; we have used one technique called Min Max Normalization.

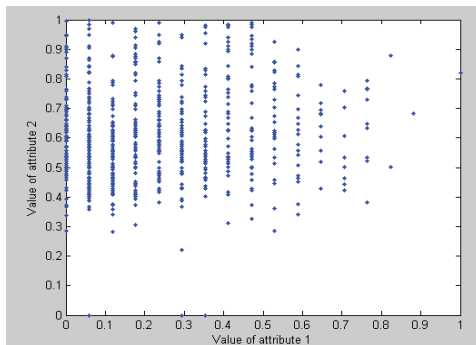


Fig 4: Pima Indian Diabetes data set
(after min-max normalization)

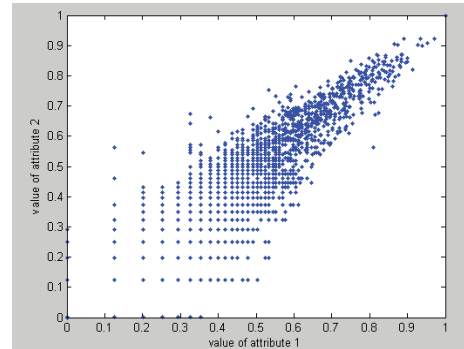


Fig 5: Yeast data set
(after min-max normalization)

Step3: Feature reduction: For high-dimensional datasets dimension reduction is usually performed prior to applying clustering and classification in order to avoid the effects of the curse of dimensionality. Feature reduction is the process of reducing the number of random variables under consideration. Here, we have used PCA for reduction of features.

PCA [14] is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The result after implementing this technique on both the data seta is shown in fig 6 & fig 7.

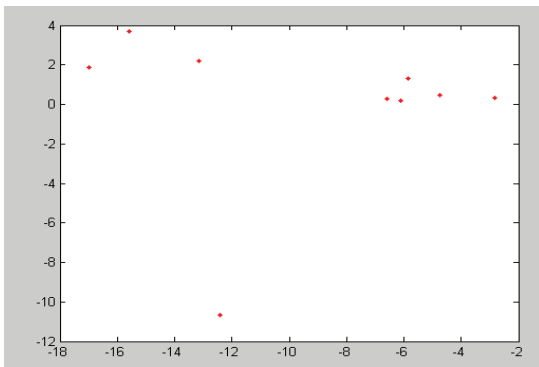


Fig 6: Pima Indian Diabetes data set
(after reduction using PCA)

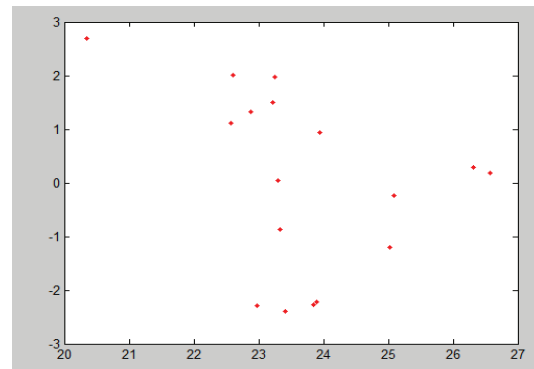


Fig 7: Yeast data set
(after reduction using PCA)

Step4: MLP for classification: MLP is a feed-forward artificial neural network model[15] that maps sets of input data onto a set of appropriate outputs. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back-propagation for training the network. MLP has been applied on the normalized and reduced data sets with the corresponding parameters.

The data set is divided into two parts; out of which 75% is used for training the network and 25% for testing. The parameters of the MPL has been initialized to Eta=0.6, alpha=0.5. The network has been trained using random weights of V and W. The output of the network has been calculated using: $OO=1/\tan$ sigmoid. The error has been

computed using: $\text{Error}(i) = (1 - \text{OO}) * (1 - \text{OO})$. The weights have been updated using PSO. The PSO updating function is described in the next step. Then, mean square error has been computed and the fig. 8 and fig. 9 shows the error curve of PCA reduced Pima Indian diabetes data sets and Yeast data set respectively.

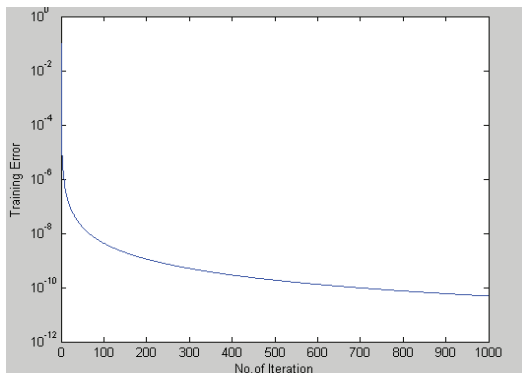


Fig 8: Error curve for Pima Indian Diabetes data set

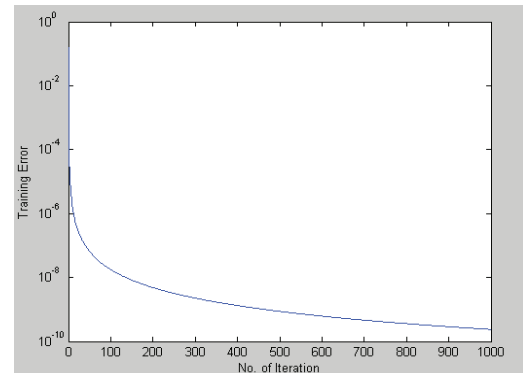


Fig 9: Error curve for Yeast data set

Step5) PSO for updating weights: PSO is one of the heuristic optimization algorithms which is based on swarm intelligence. The following are the steps used for updating the weight of the MLP using the algorithm.

- Step1) Read the weight matrix v and w and change in weight matrix $\text{del}v$ and $\text{del}w$ from MLP training.
- Step2) Initialize parameters of PSO.
- Step3) Repeat
- Step4) for each particle $i = 1 \dots n$; do
- Step5) //set the personal best position
- Step6) if $f(x_i) < f(y_i)$ then
- Step7) $y_i = x_i$;
- Step8) End
- Step9) // set the global best position
- Step10) If $f(y_i) < f(\hat{y})$ then
- Step11) $\hat{y} = y_i$;
- Step12) End
- Step13) End
- Step14) for each particle $i = 1 \dots n$; do
- Step15) Update the velocity using equation (1)
- Step16) Update the position using equation

$$x_i(t+1) = x_i(t) + v_i(t+1)$$
- Step17) End
- Step18) until stopping criteria is satisfied.

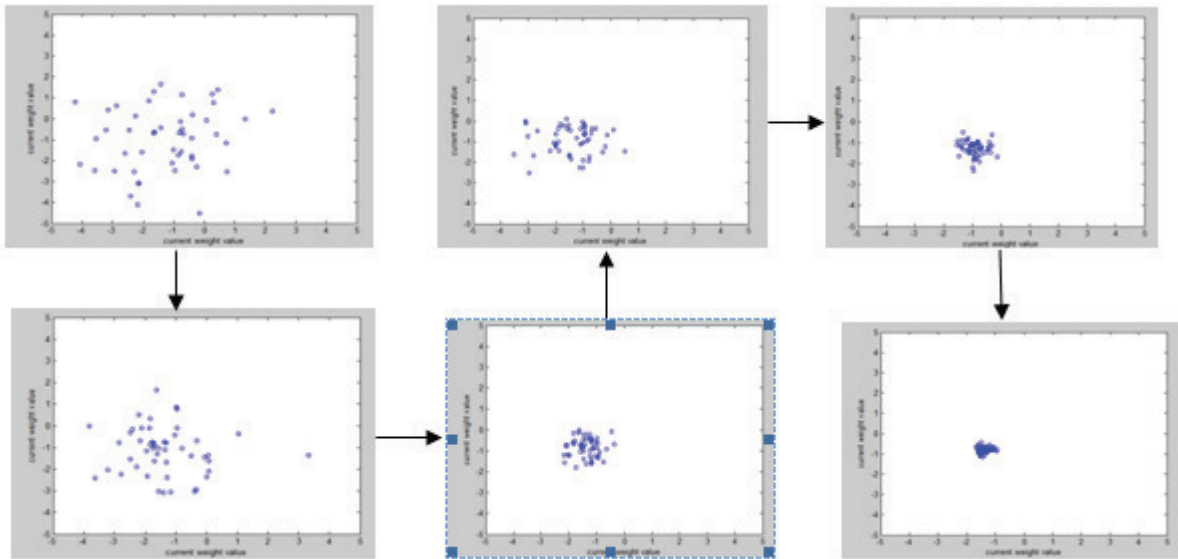


Fig 10: Updation of weights using PSO for PCA reduced Yeast data set

TABLE II: Accuracy Comparison

Datasets	% of accuracy achieved using MLP	% of accuracy achieved using PSO-MLP
Yeast	56	78.3
Pima Indian Diabetes	63.1	82.6

IV. CONCLUSION

In this paper, PSO technique has been proposed which combines the steps of dimensionality reduction to generate an optimized MLP trained network. Using the proposed algorithm a given dataset was classified in such a manner that the accuracy of classification was found to be higher once we optimize our classifier. The experimental results shows that the proposed algorithm provides better accuracy as compared to when the classifier is solely used. Although rate of accuracy achieved is higher but the computational cost involved is an area of worth investigating, therefore future work may try reduce this cost.

REFERENCES

- [1] Shelly Gupta, Dharminder Kumar, Anand Sharma, Performance analysis of various data mining classification techniques on healthcare data, *International Journal of Computer Science & Information Technology (IJCSIT 2011)*, Vol 3, pp .155-169.
- [2] Peiman Mamani Barnaghi, Vahid Alizadeh Sahzabi, Azuraliza Abu Bakar, A Comparative Study for Various Methods of Classification, *International Conference on Information and Computer Networks (ICIN 2012)*, vol. 27, pp. 62-66.
- [3] Zuyi Wang¹, Yue Wang, Jianhua Xuan, Yibin Dong, Marina Bakay, Yuanjian Feng³ Robert Clarke, Eric P. Hoffman, *Optimized multilayer perceptrons for molecular classification and diagnosis using genomic data*, Oxford University Press 2006, vol. 22, pp.755–761.
- [4] John Paul T. Yusiong , *Optimizing Artificial Neural Networks Using Cat Swarm Optimization Algorithm*, I. J. Intelligent Systems and Applications, 2012; pp. 69-80.
- [5] Jianwei Niu, Yiling He, Muyuan Li, Xin Zhang, Linghua Ran, Chuzhi Chao, Baoqin Zhang, A Comparative Study on Application of Data Mining Technique in Human Shape Clustering: Principal Component Analysis vs. Factor Analysis, *5th IEEE Conference on Industrial Electronics and Applications, 2010*, pp 2014-2018. .

- [6] R. Hassan, B. Cohanim, O. Weck, A comparison of particle swarm optimization and the genetic algorithm, in: The Proceedings of 46th AIAA/ASME/ ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2005, pp. 18–21.
- [7] J. Kennedy, R. Eberhart, Particle swarm optimization, in: The Proceedings of the 1995 IEEE International Conference on Neural Network, 1995, pp.1942–1948.
- [8] H.J. Escalante, M. Montes, L.E. Sucar, Particle swarm model selection, Journal of Machine Learning Research 10 (2009) 405–440.
- [9] J D.W. Van der Merwe, A.P. Engelbrecht. Data Clustering using Particle Swarm Optimization. In: Congress on Evolutionary Computation. 2003, pp 215-220.
- [10] Y. Lu, S. Wang, S. Li, C. Zhou. Particle swarm optimizer for variable weighting in clustering high-dimensional data. Mach. Learn. 2011, 82: 43-70.
- [11] Y Lu, S Wang, S Li, C Zhou. Text Clustering via Particle Swarm Optimization. In: IEEE Swarm Intelligence Symposium. 2009, pp 45-51.
- [12] <http://archive.ics.uci.edu/ml/>
- [13] G. Piatetsky-Shapiro and P. Smyth. From data mining to knowledge discovery: an overview, Advances in Knowledge Discovery and Data Mining, AAAI/MIT, pp. 1-34, 1996.
- [14] Jolliffe, I.T. Principal Component Analysis. Springer, New York, 1986.
- [15] Rossi, A. L. D Carv. Bio-inspired parameter tuning of MLP network for gene expression data, Eighth international conference on Hybrid Intelligent systems, 2008.