A Modified Particle Swarm Optimization Algorithm Used for Feature Selection of UCI Biomedical Data Sets

Talha Ali Khan
School of Biomdeical Engineering,
University of Technology Sydney,
Australia
talha.khan@uts.edu.au

Khawaja Zain-ul-abideen
School of Biomdeical Engineering,
University of Technology Sydney,
Australia
khawaja.zain-ul-abideen@student.uts.edu.au

Sai Ho Ling
School of Biomdeical Engineering,
University of Technology Sydney,
Australia
Steve.Ling@uts.edu.au

Abstract— In biomedical and health care applications, classification examination is extensively used to help diagnose health problems, decision making and enhance standards of patient care. Feature selection is a significant data pre-processing method in classification problems. Training of the data is achieved by using a subset dataset from UCI biomedical database. If the training dataset comprises inappropriate features, classification analysis resulted in inaccurate and incomprehensible results. In data mining, feature subset selection is data pre-processing phase that is of enormous importance. In this paper, for selecting a minimum number of features K-Nearest Neighbour (KNN) classifier is presented with a modified particle swarm optimisation (MPSO) to obtain good classification precision. Proposed method is applied to three UCI medical data sets and is compared with other feature selection approach available in the literature. Results demonstrate that the feature subset recognized by the presented MPSO with KNN neighbour classifiers give better results and accuracy as compared to the other techniques.

Keywords—Modified PSO, KNN, Feature Selection, UCI dataset.

I. INTRODUCTION

Problems in real world scenarios such as data mining contain a huge amount of features. Nevertheless, not all these features are important, large quantities of them comprise of unnecessary and insignificant data that minimizes the performance of an algorithm, for example classification algorithm. The primary purpose of feature selection is to select a limited amount of required subsets of features from huge amount of data. The main advantage of feature selection is that by eliminating unnecessary and insignificant features it minimizes the size of the data, enhances the classification algorithms' performance, reduces the computational time and elucidates the model [1]. Since the modern world is more digitalized day-by-day resulting in use of the computers becoming very common, along with huge amount of data being compiled and gathered. To process such large amount of data in short period is not possible. Therefore, data mining is gaining popularity while tackling real world problems.

Data pre-processing is a significant step in data mining as the quality of data is essential for the feature selection [2]. In health care or medical applications improving the quality of the database, ameliorate the diagnosis process. Steps involved in data pre-processing are data integration, data cleaning, data transformation and reducing size of data. In a database, some features might be unnecessary, as their information might be present in other features. Due to these extra features, computational time is increased that also affects the diagnosis accuracy. It is good to remove the unwanted data from the dataset before learning, as it is not useful for diagnosis. The main purpose of feature selection is to identify minimum set of features such that the resulting probability distribution of the data classes is approximately nearer to the original distribution achieved by using all features. In data reduction phase, it minimizes the amount of features, eliminates noise, and eradicates unnecessary data. This process has a huge impact of reducing the computational time; improve data mining algorithm efficiency in terms of result comprehensibility and predictive precision.

Due to large search space, feature selection is a strenuous problem in which the total number of feasible solutions are 2^n for a dataset with n features. It became more complicated as the number of features are increasing [3]. In most scenarios, an exhaustive search for the best feature subset of a given dataset is almost inconsistent. Different types of searching methods have been used for feature selection these includes random search, greedy search complete search and heuristic search. Still, many of the current feature selection techniques endure from large computational cost and stagnation in local optima. As a result, an effective global search method is required to elucidate feature selection problems. Particle Swarm Optimization method has latterly obtained much attention from the researchers as it is famous for its global search capability.

Therefore, in this paper a modified PSO is presented with KNN as a classifier used in the classification process. The resulted subset of features is provided as input to KNN classifier. The results obtained are compared with the other technique in Table III.

II. FEAUTURE SELECTION

Eradicating valuable data from randomly big data groups or data streams is a hot topic these days. Scientists and specialists recognize that the feature selection is a fundamental element to efficacious data mining. In the procedure of feature selection, inappropriate and repetitive attributes or noise in the data might be a problem in many conditions as they are not important and significant w.r.t the class concept such as microarray data analysis. Once the number of samples is not as much of as compared to the features, at that time machine learning becomes challenging, for the reason that the search area will be sparingly populated.

As a result, the model will not capable to distinguish precisely between noise and appropriate data. Depends on the evaluation principles, feature selection methods are commonly categorized into 2 types: I) wrapper method and II) filter methodology. The key variant between them is that wrapper method contain a classification/learning technique in the feature subset assessment step. The classification method is used as a "black box" by a wrapper method to assess the goodness (classification performance) of the particular features. A filter feature selection method is autonomous and do not require any classification algorithm. The filter algorithms are computationally less expensive and more common as compared to wrapper algorithms. Nevertheless, filters algorithms overlook the performance of the selected features on a classification algorithm; on the other hand, wrappers calculate the feature subsets depends upon the classification performance that generally effects in the improved performance obtained by wrappers as compared to the filter algorithm for a specific classification algorithm. Two foremost goals for the feature selection are to maximize the classification precision and decrease the number of features. Some time there are contradictory objectives. As a result, feature selection can be solved as a multi-objective problem so to find out a set of the solution that is a balance between these two objectives. The overall process for feature selection has four important phases as shown in Figure 1.
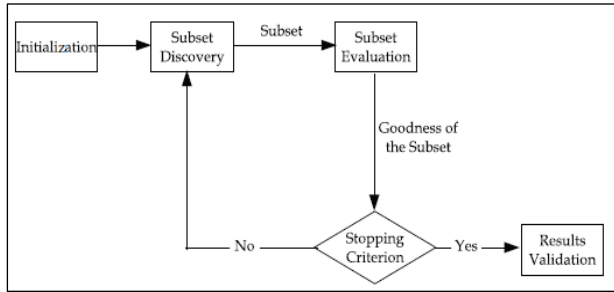

Figure 1 Four Key steps for feature selection process

a) Subset generation:

In a heuristic search in which each state requires a candidate subset for calculation in the search space. Two elementary problems decide the nature of the subset generation process. In the first one successor generation selects the search initial point, which impacts the search direction. The starting points to select the search at each state, compound, weighting, forward, backward, and random techniques may be measured. Another is that the search organization is in authority for the feature selection process with a precise approach, such as random search, sequential search or exponential search.

b) Assessment of subset

An evaluation standard must be set for the newly generated subset for evaluation. As a result, numerous evaluation criteria have been suggested in the literature to define the goodness of the candidate subset of the features. Depends upon their reliance on mining techniques, evaluation criteria

can be characterized into types: independent and dependent criteria.

c) Stopping Benchmark

To finalize the selection process stoping conditions must be set. Feature selection process ends at validation process.

d) Results validation

This stage is not the part of feature selection process however feature selection approach must be certify by performing various tests and evaluations with earlier conventional results or assessment with the results of similar approaches by using medical datasets or real world datasets.

III. MODIFIED PARTICLE SWARM OPTIMIZATION ALGORITHM (MPSO)

Standard PSO is dependent upon two equations: velocity and position of the agents. A modified PSO is proposed in this paper by transforming the velocity and position update equations.

A. Amendment in position and velocity Equations:

MPSO utilizes the same parameters and design criteria as Standard PSO. The key factor that influence the performance of the smart PSO is the addition of the new term in the position and velocity equations. The idea behind this new term is to reduce the distance between each particle over time which can result in an increase in the velocity of the particle [4]. The new velocity equation will be

$$v_{id} = wv_{id} + c_1 R_1 (p_{id} - x_{id}) + c_2 R_2 (p_{gd} - x_{id}) + w \left(\frac{c_1}{c_2} \right) (p_{id} - p_{gd}) \qquad (1)$$

Dividing the two acceleration constants in (1) helps to make the resulting value neither too big nor very small as both C_1 and C_2 have an enormous influence on the agent's mobility. The position update equation is given by

$$x_{id}(k + 1) = wx_{id}(k) + v_{id} \qquad (2)$$

B. Other Significant Parameters selection:

Other factors can also influence the performance of the MPSO. These factors are as follows:

1) MPSO Searching Criteria:

PSO uses 2 different types of searching criteria I) Global best and II) Local best. Global best method is used in this paper, so the whole swarm agents have wider access of information.

2) Velocity Clamping and Particle Penalization method:

This technique [5] helps to enhance the efficiency of the MPSO by keeping the agents velocity within the range of $[V_{max}, V_{min}]$ and penalization keep the particle with in the search domain if the sum of the agents position and velocity resulting in the new position lies outside the domain. The maximum and the velocity condition can be define as

$$V_{max} = lambda * (Max_{ss} - Min_{ss}) \qquad (3)$$
$$V_{min} = lambda * (Max_{ss} - Min_{ss}) \qquad (4)$$

Where Max_{ss} and Min_{ss} are the boundary limits for each test evaluation function and are different. The velocity of a particle is clamped by a parameter known as *lambda*. The method measures the velocity of the ith particle in d^{th} dimension with V_{max}, V_{min} and change the particle velocity by using the conditions as follows.

$$\begin{Bmatrix} V_{i,d} > V_{max} \ then \ V_{i,d} = V_{max} \\ V_{i,d} < V_{min} \ then \ V_{i,d} = V_{min} \end{Bmatrix} \qquad (5)$$

The penalization approach is as follows

$$V_{i,d} + X_{i,d} > Max_{ss} \ or \ V_{i,d} + X_{i,d} < Min_{ss} \qquad (6)$$
$$then \ V_{i,d} = 0 \ and \ X_{i,d} = -X_{i,d}$$

As defined in (6) if the particle lies outside the boundary, the position vector direction is reversed and velocity sets to zero.

3) C₁ and C₂ Values selection:

The acceleration constant in the velocity update equation controls the movement of the particle towards local and global best position. Larger values resulted in a divergence from the optimal solution and smaller values restricted the motion of the particles. Therefore, the values of the C_1 and C_2 must be selected so that it satisfies the condition $C_1 + C_2 \leq 4$. If the values do not satisfy the condition, then PSO usually does not converge. Hence, in the paper the values used in MPSO for C_1 and C_2 are set as 2.1 and 1.9 respectively.

4) Inertia constant selection:

It evaluates the influence of the earlier velocity on the present update. Bigger values of the 'w' aid in the global search whereas smaller values improve the local search. Keeping in mind the significance of "w", in MPSO the inertial weight is reduced linearly from existing iteration to the subsequent iteration. Factors are used *i.e.* w_{max} and w_{min}. To control inertial weight following equation is used:

$$w = w_{max} - \left(\frac{w_{max} - w_{min}}{FE}\right) * i \qquad (7)$$

where FE is the function evaluation and i is the current function evaluation. In this paper constriction factor has also been presented [6]. (8) measure the constriction factor as

$$chi = \frac{2}{\Phi - 2 + \sqrt{\Phi^2 - 4 * \Phi}} \qquad (8)$$

Where Φ=4.1, constriction factor is used to adjust the inertial weight by the following relation.

$$w = chi * \left(0.0005 + w * \left(\frac{FE - (i - 30)}{FE}\right)\right) \qquad (9)$$

IV. UNIVERSITY OF CALFORNIA IRVINE (UCI) MACHINE LEARNING REPOSITORY DATA SETS

UCI medical datasets is used for the feature selection process. Data sets specifications are demonstrated in Table I.

Table I Datasets used and their description

Name	Number of attributes	Classes	Number of Instances	Missing Values
Breast Cancer	9	2	699	YES
Heart	13	2	270	NO
Dermatology	34	6	366	YES

The goodness of position is measured by using the following fitness function

$$Fitness = \alpha * \gamma + \beta * (T - S)/S \qquad (10)$$

where α, β are constants that control the relative significance of classification precision and the size of feature subset, T is the total attributes in the dataset, S is the number of features chosen in that specific subset, γ is the classification precision of the features selected comparative to the preliminary classification accurateness with all the features. $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$. Here $\alpha = 0.8$ and $\beta = 0.2$.

For the classification purpose, KNN is used. The efficiency of the classifier has been calculated based on the accuracy that is illustrated as

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (11)$$

where FP= False Positives, FN= False Negatives, TN= True Negatives and TP = True positives,

V. RESULTS AND DISCUSSION

Three data sets are taking from UCI data set as shown in Table I. Each data set instances in the experiments are divided randomly into two sets: 70% as the training set and 30% as the test set. During the training procedure, each agent demonstrate one feature subset. After the training procedure, the selected features are analyse on the test set to achieve the testing classification accuracy. A learning algorithm in the evolutionary training process is required to evaluate the classification performance of the selected feature subset. The learning algorithm used in this paper is K-nearest neighbour (KNN), K = 5 in KNN to clarify the evaluation procedure. Table II shows the MPSO results with KNN as a classifier. The results highlighted that the MPSO decreases the average number of selected attributes and thus rises the accuracy. The pseudo code of MPSO is given as:

Pseudo Code of MPSO for feature Selection

Start

Split *Dataset* into a Training set and a Test set;

Particle velocity and position randomly initialise

while *function evaluation is not met* do

Evaluate fitness of each particle by (10) ;

for *i=1* to *population size* do

Update the *pbest* of particle *i*;

Update the *gbest* of particle *i*;

for *i=1* to *population size* do

for *d=1* to *number of available features* do

Update the velocity of particle *i* by *(1)*;

Update the position of particle *i* by *(2)*

Measure the classification accuracy of the selected feature subset on the test set;

return the position of *gbest* (the selected feature subset);

return classification accuracy;

Table II Classification performance on UCI Data sets

Dataset	Number of Original Features	Average number of features selected using MPSO	Accuracy with reduced features
Breast Cancer	9	3	82.6
Heart	13	4.4	85.13
Dermatology	34	15.2	98.86

Table III shows the comparison of the results obtained from MPSO with those in [7]. The results demonstrated that the MPSO gives better performance than the PSO used in [7]. The average number of the features are also reduced and the accuracy of the three data sets are much improved than the compared PSO.

Table III Comparison of the results

Dataset	Accuracy with reduced features by MPSO	Accuracy with reduced features by PSO [7]
Breast Cancer	82.6	76.22
Heart	85.13	83.70
Dermatology	98.86	98.63

VI. CONCLUSION

Examining the fitness function for MPSO that depends upon feature selection method for choosing a lesser number of features and accomplishing greater classification accurateness is the primary objective of this paper. The purpose is effectively attained by using function in (13); the fitness function is observed and compared with a paper based on wrapper feature selection method on same three data sets. MPSO with the presented fitness function can effectively

decrease the number of features required and accomplish higher classification efficiency.

REFERENCES

[1] J. Tang, S. Alelyani, and H. Liu, *Feature selection for classification: A review.* 2014, pp. 37-64.

[2] A. Jović, K. Brkić, and N. Bogunović, "A review of feature selection methods with applications," in *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2015, pp. 1200-1205.

[3] S. Fong, "Opportunities and Challenges of Integrating Bio-Inspired Optimization and Data Mining Algorithms," 2013, pp. pp.385-401.

[4] T. A. Khan, S. H. Ling, and A. S. Mohan, "Advanced Particle Swarm Optimization Algorithm with Improved Velocity Update Strategy," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2018, pp. 3944-3949.

[5] M. Alhussein and S. I. Haider, "Improved Particle Swarm Optimization Based on Velocity Clamping and Particle Penalization," in *2015 3rd International Conference on Artificial Intelligence, Modelling and Simulation (AIMS)*, 2015, pp. 61-64.

[6] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, 1998, pp. 69-73.

[7] H. Harb and A. S. Desuky, *Feature Selection on Classification of Medical Datasets based on Particle Swarm Optimization.* 2014, pp. 14-17.