Rough Set Feature Selection Using Bat Algorithm

Hemal Patel[1], Prof. Mansi Shah[2]

[1]M.E I.T
[2]Information Technology & KITRC
[1] hemalptl451@gmail.com; [2] mansi.parmar18@gmail.com

Abstract - The Swarm Intelligence Techniques and Evolunatry Algorithms of Rough Set Feature Selection are development Algorithms, which attracted much attention and appeared its ability in many applications such as Classification, Clustering etc. In this Research BAT Algorithm (BA) is Used in Place of Particle Swarm Optimization (PSO).Result Shows Using Medical Data Set that BAT Algorithm is overcome on it in terms of Accuracy Level and accelerate the time which increase Network efficiency.

Index Terms - Classification, Particle Swarm Optimization (PSO) Rough Sets, Feature Selection (FS), Bat Algorithm (BA)

1. Introduction

Data mining is the process of selecting, exploring and modelling large amounts of data in order to discover unknown patterns or relationships which provide a clear and useful result to the data analyst[1]. There are two types of data mining tasks: descriptive data mining tasks that describe the general properties of the existing data, and predictive data mining tasks that attempt to do predictions based on available data.

Data mining involves some following key steps :

I. Problem definition: The first step is to identify goals.

II. Data exploration: All data needs to be consolidated so that it can be treated consistently.

III. Data preparation: The purpose of this step is to clean and transform the data for more robust analysis.

IV. Modelling: Based on the data and the desired outcomes, a data mining algorithm or combination of algorithms is selected for analysis. The specific algorithm is selected based on the particular objective to be achieved and the quality of the data to be analysed.

V. Evaluation and Deployment: Based on the results of the data mining algorithms, an analysis is conducted to determine key conclusions from the analysis and create a series of recommendations for consideration.

TECHNIQUES OF DATA MINING

There are several major data mining techniques have been developing and using in data mining projects recently including association, classification, clustering, prediction, sequential patterns and decision tree.

(A) Association :-

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That's is the reason why association technique is also known as *relation technique*.

(B) Classification :-

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data items into groups. For example, we can apply classification in application that "given all records of employees who left the company, predict who will probably leave the company in a future period." In this case, we divide the records of employees into two groups that named "leave" and "stay". And then we can ask our data mining software to classify the employees into separate groups.

(C) Clustering :-

Clustering is a data mining technique that makes meaningful or useful cluster of objects which have similar characteristics using automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

(D) Prediction:-

The prediction, as it name implied, is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, the prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable.

(E) Sequential Patterns:-

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar patterns, regular events or trends in transaction data over a business period.

(F) Decision trees:-

Decision tree is one of the most used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers.

Classification

Classification involves predicting an outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, normally known as prediction attribute. The algorithm discovers the relationships between the attributes that would make it possible to predict the outcome. After that the algorithm is given a new data set called prediction set, which contains the same set of attributes, except for the prediction attribute is not yet known. The algorithm analyses the input and generates a prediction.

➢ In Classification collected data is usually associated with a high level of noise. There are There are many reasons causing noise in these data, among which imperfection in the technologies that collected the data and the source of the data itself are two major reasons.

Dimensionality reduction is one of the most popular techniques to remove noisy (i.e. irrelevant) and
redundant features.

Reduction Technique

Feature Selection   Feature Extraction
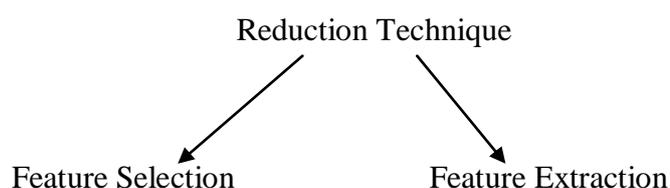
Feature extraction approaches project features into a new feature space with lower dimensionality and the new constructed features are usually combinations of original features.
Example : Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Canonical, Correlation Analysis (CCA).

314

2. Feature Selection

It   aim is to select a small subset of features that minimize redundancy and maximize relevance to the target such as the class labels in classification.
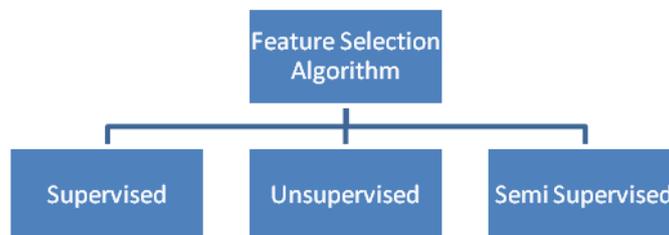
A feature selection method consists of four basic steps namely, subset generation, subset evaluation, stopping criterion, and result validation.

1) A candidate feature subset will be chosen based on a given search strategy, which is sent,
2) To be evaluated according to certain evaluation criterion.
3) The subset that best fits the evaluation criterion will be chosen from all the candidates that have been evaluated after the stopping criterion are met.
4) The chosen subset will be validated using domain knowledge or a validation set.

Feature selection selects a subset of features from the original feature set without any transformation, and maintains the physical meanings of the original features.

Feature selection for classification attempts to select the minimally sized subset of features according to the following criteria,
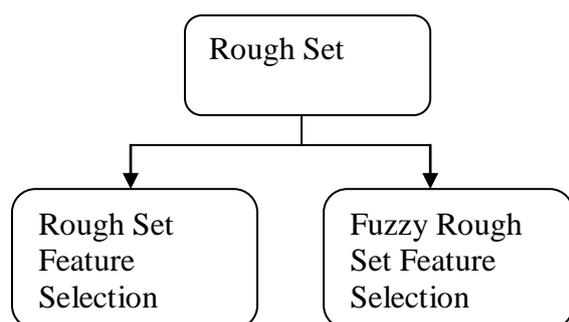
• The classification accuracy does not significantly decrease.
• The resulting class distribution, given only the values for the selected features, is as close as possible to the original class distribution, given all features.

3. Rough Set Theory

Rough set theory can be regarded as a new mathematical tool for imperfect data analysis. The theory has found applications in many domains, such as decision support, engineering, environment, banking, medicine and others.

Advantages:

• It provides efficient methods, algorithms and tools for finding hidden patterns in data.
• It allows to evaluate the significance of data.
• It allows to generate in automatic way the sets of decision rules from data.
• It is easy to understand.
• It offers straightforward interpretation of obtained results.

```
┌─────────────────┐
│    Rough Set    │
└─────────────────┘
     │         │
     ▼         ▼
┌──────────┐  ┌──────────────┐
│Rough Set │  │Fuzzy Rough   │
│Feature   │  │Set Feature   │
│Selection │  │Selection     │
└──────────┘  └──────────────┘
```

4. Rough Set Feature Selection

Rough set attribute selection is utilized when the value of attributes are discrete values. Methods in both categories use the concept of dependency for finding reducts. One of the well-known methods is the so called QuickReduct that is greedy search algorithm using dependency.

The QuickReduct algorithm calculates a reduct without finding all subsets. It starts from an empty set and each time selects a feature which causes the greatest increase in dependency degree. The algorithm stops when adding more features does not increase dependency degree.

It does not guarantee to find minimal reduct as long as it employs a greedy algorithm which is a forward search and capable of being trapped in a local optimum.

Evolutionary algorithms and swarm intelligence algorithms like PSO and ACO have been widely used for finding minimal reducts by means of rough set concepts

5. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a conventional and semi-robotic algorithm. It is based on the social behaviour associated with bird's flocking for optimization problem. A social behaviour pattern of organisms that live and interact within large groups is the inspiration for PSO. The PSO is easier to lay into operation than Genetic Algorithm. It is for the motivation that PSO doesn't have mutation or crossover operators and movement of particles is effected by using velocity function[3].

PSO, Particle Swarm consists of 'n' particles. The position of each particle stands for potential solution in D-dimensional space. Individuals, potential solutions, flow through hyper dimensional search space. The experience or acquired knowledge about its neighbours influences the changes in a particle within the swarm. The PSO algorithm involves of just three steps, which are being replicated until stopping condition, they are as follows[4].

(i) Evaluate the fitness of each particle.
(ii) Update individual and global best functions.
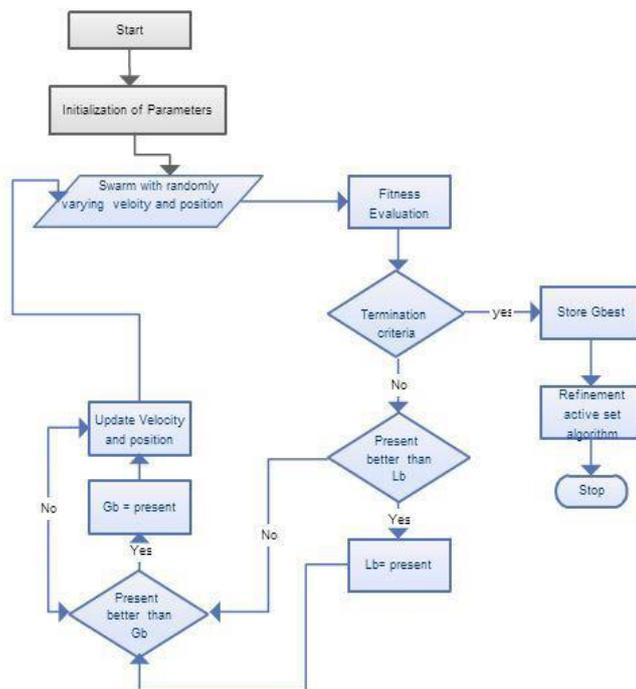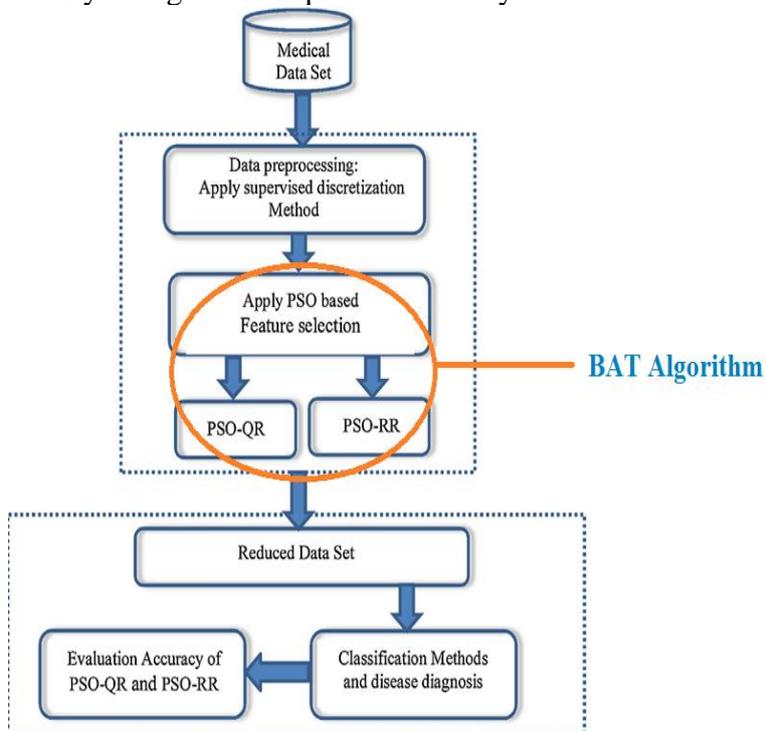(iii)Update velocity and position of each particle.

Fig 2 : Working of PSO

This is the current Methodology with using PSO Method we have to replace BAT Algorithm on the place of PSO. By Using this it improves accuracy level and time.

317

6. BAT Algorithm

> Bat-inspired algorithm is a metaheuristic optimization algorithm developed by Xin-She Yang in 2010. This bat algorithm is based on the echolocation behaviour of micro bats with varying pulse rates of emission and loudness. it has a constant frequency which is usually in the range of 25–150 kHz corresponding to the wavelengths of 2–14 mm. Echolocation works as a type of sonar: bats, mainly micro-bats, emit a loud and short pulse of sound, wait it hits into an object and, after a fraction of time, the echo returns back to their ears. Thus, bats can compute how far they are from an object. In addition, this amazing orientation mechanism makes bats being able to distinguish the difference between an obstacle and a prey, allowing them to hunt even in complete darkness.
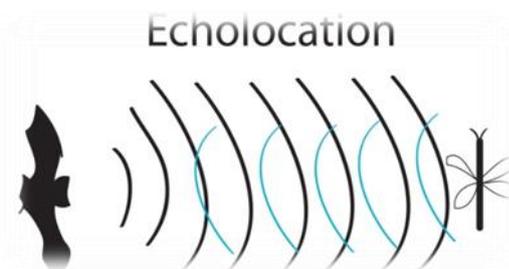
Echolocation

Some bats have evolved a highly sophisticated sense of hearing. They emit sounds that bounce off of objects in their path, sending echoes back to the bats. From these echoes, the bats can determine the size of objects, how far away they are, how fast they are travelling.

Idealized rules for Bat algorithm

1. All bats use echolocation to sense distance, and they also "know" the difference between food/prey and background barriers in some magical way.

2. A bat bi fly randomly with velocity vi at position xi with a fixed frequency $f$min, varying wavelength λ and loudness $A0$ to search for prey. They can automatically adjust the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission $r \in [0, 1]$, depending on the proximity of their target .

3. Although the loudness can vary in many ways, Yang assume that the loudness varies from a large (positive) $A0$ to a minimum constant value Ami .

Steps for Algorithm :-

- While $t < T$
- For each bat bi, do
- Generate new solutions through Equations (10),(11) and (12).
- If $rand > ri$, then
- Select a solution among the best solutions.

- Generate a local solution around the best solution.
- If $rand < Ai$ and $f(xi) < f$(xcgbest), then
- Accept the new solutions.
- Increase ri and reduce .
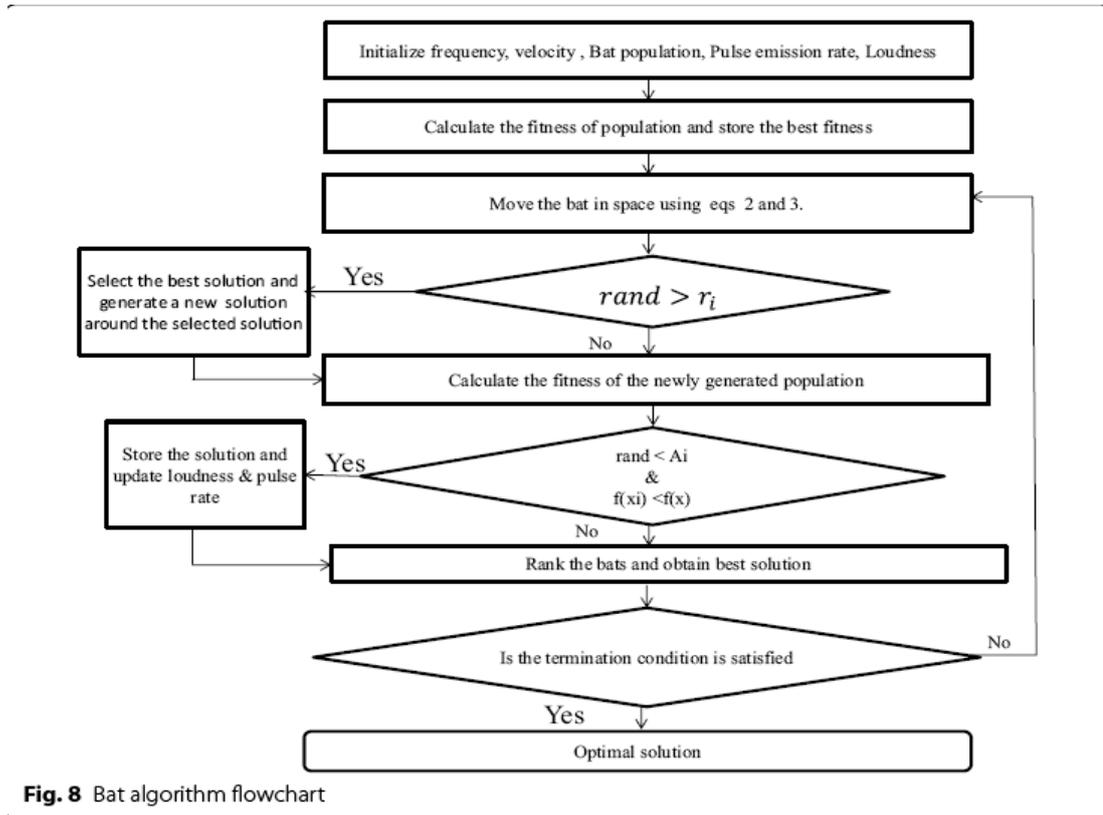- Rank the bats and find the current best xcgbest.

Flow Chart for Algorithm

Fig. 8 Bat algorithm flowchart

7. Literature Review

Sr No	Method Name	Description	Advantage
1	SPSO-QR[5]	It start with an empty set and it adds one at a time , in turn.	The dependency of subset is calculated based on dependency & decision attribute and best particle is chosen
2	SPSO-RR[5]	It start by selecting random values for each particle & velocity.	To avoid calculation of discern ability functions which can be computationally expensive without optimizations.
3	PSO[6]	It is based based on the use of multiple sub-swarms instead of one (standard) swarm.	It increase overall performance of network.

4	HGAPSO[7]	It is obtained through integrating standard velocity & update rules of PSO with selection, crossover & mutation from the GA.	It does not need to set the number of desired features a priori.

8. EXPERIMENTAL EVALUATION

In this paper, a comparison between the results of the (PSO) algorithm and the bat algorithm (BA) using Naive Byes Classifier with Medical Data Set it shows accuracy level and time both.
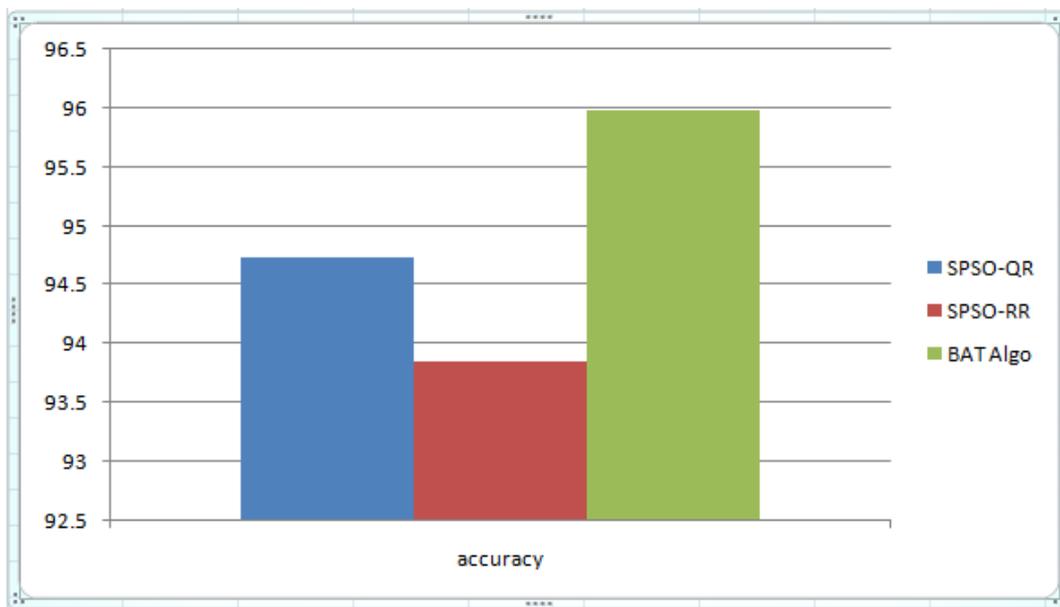

Table 1. shows the (MSE) & (SD) & Time for PSO

Data Set	MSE	Standard Deviation	Time in Second
Iris	2.43	4.87	
wine	36.36	2.00	6.5
Glass	54.30	20.03	

Table 2. shows the (MSE) & (SD) & Time for BA

Data Set	MSE	Standard Deviation	Time in Second
Iris	2.14	1.09	
wine	23.80	1.73	
Glass	18.29	1.00	3.6

9. Conclusions

Through experiments and tests it was concluded that the Bat Algorithm is better than (PSO) Algorithm for Medical dataset. Although the PSO algorithm has the power to find a Global Minimum, but its society has a slow rate of convergence in finding from optimal solution, therefore the Bat algorithm is better because its based on the principle of frequency tuning and change the emission rate of impulses which lead to the good affinity from ideal solutions, in addition to the creation process of a balance between exploration and exploitation and accelerate the training time, which led to increase network efficiency and reduce the fall errors and thus the algorithm is very efficient in multiple applications, such as image processing and clustering, etc.

10. ACKNOWLEDGEMENT

REFERENCES

[1] **A.Shameem Fatima , D.manimegalai** and **Nisar Hundewale** , "A Review of Data Mining Classification Techniques Applied for Diagnosis and Prognosis of the Arbovirus-Dengue":Ijcsi -**Volume 8, Issue 8, No.3 November 2011.**

[2] **Ritika MTech Student** , "Research on Data Mining Classification":Ijarcsse - **Volume 4, Issue 4, April 2014.**

[3] **Sivagowry.S , Dr. Durairaj.M**, "PSO - An Intellectual   Technique for Feature Reduction on Heart Malady   Anticipation   Data":Ijarcsse   -   **Volume   4,   Issue   9,   September   2014.**

[4] **Durairaj. M , Sivagowry. S**, "Feature Diminution by Using Particle Swarm Optimization for Envisaging the Heart Syndrome": Ijitcs- **January 2015**.

[5] **H.Hannah Inbarani,Ahmad Taher Azar and G. Jothi** "Supervised hybrid feature selection based on PSOand rough sets for medical diagnosis": Elsevier- **2014.**

[6] **Sivagowry.S , Dr. Durairaj.M**, "PSO - An Intellectual   Technique for Feature Reduction on Heart Malady Anticipation Data":Ijarcsse - **Volume 4, Issue 9, September 2014.**

[7] Pedram Ghamisi, Student Member, IEEE, and Jon   Atli Benediktsson, Fellow, IEEE, "Feature Selection Based on Hybridization of Genetic Algorithm and Particle Swarm Optimization":IEEE - **Volume 12, No 2 February 2015.**