Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 3, March 2015, pg.87 – 92

SURVEY ARTICLE

A Survey on Particle Swarm Optimization and Rough Set Theory in Feature Selection for Heart Disease Prediction

Durairaj.M[1], Sivagowry.S[2]

[1& 2] Department of Computer Science, Engineering and Applications, India

[1] durairaj.bdu@gmail.com, [2] sivagowry87@gmail.com

Abstract— ***Particle Swarm Optimization (PSO) is a heuristic comprehensive optimization method which is centred on Swarm Intelligence. This algorithm is widely used and rapidly developed for its easy implementation. Rough Set Theory (RST) is a mathematical tool which deals with the uncertainty and vagueness of the decision system. RST is productively applied in many fields. Both the methods are used to recognize the reduct set of all attributes in the decision system. This paper converses and compares both the methods in attribute reduction in theoretical aspects.***

Keywords—Particle Swarm Optimization, Rough Set Theory, Feature Reduction, Heart Disease, Prediction

I. INTRODUCTION

It was estimated by the Institute of Medicine that the medical error costs up to $11 to $19 billion US dollars [1]. This has been declined since then. Changes in the lifestyle, work culture and food habits lead to Cardio Vascular Disease (CVD) in developing countries. World Health Organization (WHO) reported that 30% of total global death is due to CVD [2, 22]. This estimation was done in 2008. It is also analyzed that the number of CVD Patients will increase from 300 to 600 million by 2020 [3]. By 2030, almost 25 million people will die from CVDs, mainly from heart disease and stroke [4].

The human decision (i.e.) the doctor or medical practitioner's decision should be precise while diagnosing of heart disease, since marginal error causes major casualty. But when the data set is large and vague, the human error is inevitable because of stress or work load. The decision which is made based on improper information from imprecise data lead to exorbitant mistake. It may cause disastrous consequences such as loss of life, casualties, etc. In this situation, Data Mining can be used as an intellectual symptomatic tool in health care for an efficient and more accurate decision making.

Feature selection is a method of eliminating features with little or no information. A large number of features with high dimension space consume considerable time in decision making since time plays crucial role in traumatic hours. It also degrades the classification accuracy [5]. More accuracy leads to correct prediction or decision making. In machine learning and statistic, the feature selection is considered as a problem of global combinatorial optimization [6]. Feature Selection techniques has three main advantages when constructing predictive models. The major benefits of the feature selection are that it improves the model interoperability, abridge the training time and over fitting which is for enhancing the generalization.

A good feature selection mechanism is needed to amplify the processing rate with predictive exactitude and to avoid incomprehensibility [7]. The problem size and the resultant search space for any learning algorithm can be reduced by using the Feature Selection. In general, two important goals of Feature Selection are to minimize the classification error and number of attributes [8]. There are number of Feature Selection algorithms available for Data Mining. The complicated task is to find the suitable algorithm which would select the feature relevant to the task without affecting the quality of the result. The Swarm Intelligent techniques and Rough Set Theory are the techniques that are mostly used for Feature reduction. But Particle Swarm

Optimization (PSO) technique is found to be effectual since it computes the individual best of each particle and finally compares all the individual particles to get the global best. It is also palpable that the Redunt feature obtained by PSO would not affect the accuracy. Rough Set also provides certain advantage in Feature reducing which doesn't affect the accuracy of the data set.

The paper is organized as follows: Section II describes about Particle Swarm Optimization algorithm with its pseudo-code. Section III describes about the Rough Set Theory and Section IV gives details about the literature that support PSO and RST in medical data processing. And Section V concludes the paper.

II. PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO) 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 [20, 21]. 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 [9]. In PSO, each particle adjusts its own flying memory and its companion's flying involvement in order to flying in the search space with velocity. In the basic PSO, Particle Swarm consists of 'n' particles. The position of each particle stands for potential solution in D-dimensional space, individuals, potential solutions and flow through hyper dimensional search space. The experience or acquired knowledge about its neighbour's influences the changes in a particle within the swarm. The PSO algorithm involves just three steps. These three steps are replicated until the stopping conditions are met. The steps are as follows:

(i) Fitness evaluation of each particle
(ii) Update individual and global best functions.
(iii) Update velocity and position of each particle.

The best-fit particle of the entire swarm [10] influences the position of each particle. Each individual particle i ∈ [1… n] where n > 1, has current position in search space x_i, a current velocity v_i and a personal best position $P_{best,i}$ where i is the smallest value determined by objective function f. By using the $P_{best,i}$, the global best position G_{best} is calculated, which is the buck value obtained by comparing all the $P_{best,\,i}$.

The $P_{best,i}$ is calculated by using the formula

$$p_{best,i} = \begin{cases} P_{best,i} \; if \; f(x_i) > P_{best,i} \\ x_i \; if \; f(x_i) \le P_{best,i} \end{cases} \qquad \text{…. …. …. … …. …. …. (1)}$$

The formula used to calculate Global Best Position G_{best} is

$$G_{best} = \{\min\{P_{best,i}\}, where \; i \in [1, \ldots\ldots, n] \text{ where } n > 1 \qquad \text{….. …. … …. .. (2)}$$

Velocity can be updated by using the formula

$$\text{…….} \quad V_i^{t+1} = wv_i\,(t) + c_1 r_1 [x_i(t) - x_i(t) + c_2 r_2 [g(t) - x_i(t)] \qquad \text{…… ……. …. (3)}$$

Where $v_i(t)$ is the velocity, w, c_1 and c_2 are user supplied co-efficients. The r_1 and r_2 are random values $x_i(t)$ is the individual best solution, $g(t)$ is the swarm's global best candidate solution. $wv_i\,(t)$) is known as inertia component. The inertia component is responsible for keeping the particle moving in same direction, it was originally heading. Inertia component value lies between 0.8 and 1.2. Lower the values of inertia component, it speeds up the convergence of swarm to optima. But higher value encourages the exploration of entire search space. c1r1 [(^{x}i(t)-x_i(t)] is known as cognitive component. It pretence as a particle's memory and it verge to return to the region of search space, where it experiences high individual factors. c2r2 [(g(t)-x_i(t)] is known as social component, which causes the particle to move to the best region the swarm has found so far.

The pseudo code of the Particle Swarm Optimization Algorithm is as given below.


```
Algorithm PSO
Input
   m: the swarm size; c1, c2 : positive acceleration constants; w: inertia weight
   MaxV: maximum velocity of particles
   MaxGen: maximum generation
   MaxFit: maximum fitness value
Output:
Pgbest: Global best position
Begin
   Swarms {xid, vid} =Generate (m); /*Initialize a population of particles with random positions and velocities on S dimensions */
   Pbest(i)=0; i = 1,….,m,d=1,…..,S
   Gbest = 0; Iter = 0;
   While (Iter<MaxGen and Gbest < MaxFit)
  {For (every particle i)
    {Fitness(i)=Evaluate(i);
      IF(Fitness(i)>Pbest(i))
        {Pbest(i)=Fitness(i);pid=xid; d=1,…..,S}
      IF(Fitness(i)>Gbest(i))
        {Gbest(i)=Fitness(i);gbest=i;
    }
   For (every particle i)
    {For(every d){
       vid = w*vid + c1*rand()*(pid-xid)+c2*Rand()*(pgd-xid)
        IF (vid > MaxV)  { vid = MaxV;}
        IF (vid <- MaxV)  { vid =- MaxV;}
        xid = xid+ vid
    }
    }
 Iter=Iter+1;
}/* rand() and Rand() are two random functions in the range [0,1]*/
Return P_{gbest}
End
```

III. Rough Set Theory

The concept of Rough Set was first introduced by Z.I. Pawlak in 1982 [10]. Any external parameter is not required by Rough Set Theory to analyze and conclude about data set. It is used to find the indispensable features. The discernibility matrix is used to recognize dispensable and indispensable Features [11].

A. *Basics*:

Let I=(U,A) be an Information System. Here, U is an non-empty set of objects called universe and A is non-empty finite set of attributes such that $_a$:U→V_a for every $_a$∊ A. V_a is a set of values that attribute $_a$ may take. If P⊆A, there is an associated equivalence relation:

$$IND\ (P)-\{(x,y)\in U \times U\ \forall a \in P,\ f_a(x)=f_a(y)\} \quad (4)$$

The partition of U generated by IND (P) is denoted by U/P. If (x,y) ∊ IND (P) , then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted by $[x]_P$. Let X ⊆ U, the P-lower Approximation $\underline{P}$ X and P-upper approximation $\overline{P}$ X of set X can be defined as:

$$\underline{P}X - \{x \quad U|[x]_P \subseteq X\} \quad (5)$$

$$\overline{P}X - \{x \quad U|[x]_P \cap X \neq \phi \quad (6)$$

Let P, Q⊆ A be equivalence relations over U, then the positive, negative and boundary regions can be defined as :

$$POS_P(Q) = \bigcup_{x\in U|Q} \underline{P}X \quad (7)$$

$$NEG_P(Q) = U - \bigcup_{x\in U|Q} \overline{P}X \quad (8)$$

$$BND_P(Q) = \bigcup_{x \in U|Q} \overline{P}X - \bigcup_{x \in U|Q} \underline{P}X \quad (9)$$

The goal of attribute reduction is to confiscate redundant attributes so that the abridged set provides the same superiority of cataloguing as the original. The set of all reduct is defined as:

$$Red(C) = \{ R \subseteq C \mid \Upsilon_R(D) - \Upsilon_C(D), \forall B \subseteq R, \Upsilon_B(D) \neq \Upsilon_C(D) \quad (10)$$

A dataset may have many attribute reducts. The set of all optimal reducts is:

$$Red(C)_{min} = \{ R \in Red \mid \forall R' \in Red, |R| \leq |R'| \quad (11)$$

B. *Quick Reduct Algorithm*:

It calculates the dependency of each attribute, and the best candidate is chosen based on the calculation. The best features are added to the dataset until the dependency of the reduct candidate is equal to the consistency of dataset [13].

Algorithm QR:
QR (*C*, *D*)
C, the set of all conditional features;
D, the set of decision features.
$R \leftarrow \{\}$
do
  $T \leftarrow R$
$\forall x \in (C-R)$
if $\Upsilon_{R \cup \{X\}}(D) > \Upsilon_T(D)$
    $T \leftarrow R \cup \{x\}$
  $R \leftarrow T$
until $\Upsilon_R(D) > \Upsilon_C(D)$
return R

IV. Feature Selection Based on PSO and RST

The following literature supports explain the way both PSO and RST are used in Feature Selection. The experimental results described in literatures show the satisfactory results.

A modified Multi Swarm PSO [8] was studied to crack discrete problems with number of sub swarm and a multi swarm scheduler that can scrutinize and control each sub swarm using the rule. Improved Feature Selection (IFS) algorithm was incorporated with Multi Swarm PSO, Support Vector Machine (SVM) and F-Score for effective Future Selection. Recital of the anticipated method was compared with Standard PSO, Genetic Algorithm and Grid search. The method proposed in this work performed drastically in terms of classification accuracy rate.

A feature selection strategy [9] was proposed and it was based on rough sets and PSO. PSORSFS algorithm and other feature selection algorithm were implemented using Matlab. PSO obtained optimal solution so quickly. It was observed that inertia weight and maximum velocity have an imperative brunt in the performance of PSO.

PSO was combined with K-Nearest Neighbor (PSOK-NN) classifier [14] to determine Coronary artery disease by using exercise stress testing data. It was composed of two steps. At the first step, a particle was generated which demonstrates the whole sample optimally in training data for both health and unhealthy patients. Then the class of test sample was determined by using the distance of the test sample generated by utilizing K-Nearest Neighbor algorithm. The accuracy obtained while using PSOK-NN was 92.49%.

BenxianYue, et. al., [15] illustrated that PSO can be used as a solution to find the appropriate features with Rough Set Theory. The proposed method observed the change of positive region as the particles proceeded through the search space. The performance of the algorithm is evaluated with Genetic Algorithm. The fitness of each individual is evaluated during search procedure. When compared with Genetic Algorithm, it is perceived that PSO requires shorter time to obtain better results.

In [8], a hybrid filter-wrapper feature subset selection algorithm is presented which is based on PSO. Maximum relevance minimum redundancy PSO (mr^2PSO) uses the mutual information available from the filter model to weigh the bit selection probabilities in discrete PSO. This method integrated mutual information filter model within the PSO based wrapper model. Mutual dependency of two variables is measured by the quantity called mutual information. The mutual information is expressed as a composite measure of feature relevance and redundancy. The probability of the features is weighed by using feature relevance and redundancy, which enhances the convergence rate and solution quality of feature subset selection. A stratified 10-fold cross validation is recommended.

In [19], it is recommended that hill climbing rough set approach is inadequate to find optimum solution. It is not feasible for complete search even in medium-sized data set. The PSO interacts with individuals in the population to find the optimal region in complex search space. The main advantages of PSO over Genetic Algorithm (GA) is that it doesn't require cross over or mutation operation and also inexpensive in terms of both memory and run time. Experimental results of this work show that PSO is efficient for rough set based feature selection.

In [13], a computer aided diagnosis system has been introduced for the heart valve disease by using binary PSO and Support Vector Machine (SVM) algorithm in conjunction with K-nearest neighbor and leave-one-out cross validation. The most weighted feature is selected by using binary PSO algorithm. SVM is applied to classify the outcome into two major classes as healthy and having heart valve disease. In this work, the proposed method helps to optimize the feature selection process. It can also be used an ideal pre-processing tool since it increase the classification accuracy.

A new feature selection based on hybrid method which combines Rough Set Theory hybrid and Bee Colony Optimization (BCO) is proposed [12]. This method is applied on the medical data set and minimal reduct set is obtained. The proposed method is compared with Quick Reduct, Entropy based Reduct and with hybrid technique which combines Genetic Algorithm, Particle Swarm Optimization, Ant Colony Optimization and Rough Set. The solutions provided by Quick Reduct and Entropy based Reduct was close to minimal reduct set. But the solution is not optimal. The GenRSAR, AntRSAR, PSO-RSAR are also performing well but there is no consistency in the result. This is because these algorithms dealt with random parameters. Bee-RSAR exhibits constant and better performance on medical data set.

To offer better conditions on subsequent analysis and to reduce the complexity of data, the data is pre-processed. Rough Set Theory was applied in three pre-processing steps such as [16] Discretization, Feature Selection and Instance Selection. Experimental results have shown that the Rough Set Theory considers only feature with large dependency. K-nn classifier is applied to validate the result.

Similar objects are determined by using Expectation Maximization (EM) Clustering algorithm [17]. The features generated by this method are compared with the Fuzzy Rough Feature Selection and Tolerance based Feature Selection. The important aspect to reduce the irrelevant and redundant features is that they lead to slow learning and low accuracy. Rough Set Theory used to reduce the input dimensionality based on data dependencies. Function of the approximation is used to calculate the measure of dependency. The proposed method produces a smaller number of attributes compared with other Feature Selection algorithms. It also improved the average accuracy of classifiers like J48, JRIP and CART.

Quick Reduct [18] algorithm is used to reduce the number of genes from gene expression data. The Quick Reduct algorithm is used to obtain minimal data set. Rough Set Theory is used to pre-process the data to mine suitable rules. For post processing, Formal concept analysis is used for mining rules. The rules are mined to extract knowledge and the most important factors that affect the decision making [16]. The aim of using Rough Set Theory is that it will help in prediction whereas Formal Concept Analysis is used in describing the data. This model is believed to be useful for decision making in the medical field.

PSO [23, 24] is used for Feature Reduction. Here, the patients are classified as diseased and non-diseased with the aid of Artificial Neural Network. The parameters such as Regression, Performance plot, Confusion Matrix and ROC Values are used to analyze the performance. The performance of the whole network is increased after applying PSO.

V. CONCLUSION

Particle Swarm Optimization and Rough Set Theory both plays a major role in Feature Reduction. Rough set theory overcomes PSO in a way such that unlike PSO the RST does not depend on any external parameter. But in PSO, the convergence to the optimal solution occurs in a short time. It is expected that combining both these methods into one hybrid method would yield some encouraging results, which is to be carried out as future extension.

REFERENCES

[1] www.nytimes.com/2009/07/29/opinim/29hall.html.

[2] Huan Liu and Hiroshi Motoda, Rudy Setiono and Zheng Zhao. "Feature Selection: An Everlasting Frontier in Data Mining", *JMLR: The 4th Workshop on Feature Selection and Data Mining*, 2010.

[3] World Health Organization. Strategic priorities of the WHO Cardiovascular Disease programme. Available online at URL: http://www.who.int/whr/200. Last accessed February 2006.

[4] Chen A.H., "HDPS: Heart Disease Prediction System*", Computing in Cardiology*, ISSN 0276-6574, pp 557-560, 2011.

[5] Latha Parthiban and R.Subramanian, "Intelligent Heart Disease Prediction System using CANFIS and Genetic Algorithm*", International Journal of Biological and Life Sciences*, Vol 3(3), pp157-160, 2007.

[6] Elbedwehy M.N, "Detection of Heart Disease using Binary Particle Swarm Optimization", *Proceeding of the Federated Conference in Computer Science and Information System*, 2012, pp 177-182.

[7] Bing Xue," Multi Objective Particle Swarm Optimization for Feature Selection", *GECCO '12*, 2012.

[8] Alper unler, Alper Murat and Ratna Babu Chinnam, "m2PSO: A maximum relevance minimum redundancy feature selection method based on Swarm Intellignce for SVM Classification", *Elsevier*, 2011, pp 4625-4641.

[9] Xiangyang Wang, Jie Yang, Xialong Tens and Weijan Xia, Richard Jension, " Feature selection basedon Rough Set and Particle Swarm Optimization", *Pattern Recognition Letters*, 2007, pp: 459-471.

[10] http://en.wikipedia.org/wiki/Rough_set

[11] Frida Coaquira and Edgar Acuna, " Application of Rough Set Theory in Data Preprocessing for Knowledge Discovery", Proceedings of the World Congress on Engineering and Computer Science, 2007.

[12] N. Suguna and Dr. K. Thanuskodi, " A Novel Rough Set Reduct Algorithm for Medical Doman based on Bee Colony Optimization", Journal of Computing, Vol 2(6), pp 49-54, 2010.

[13] Yuanning Liu, Gang Wang, Huiling Chen, Hao Dong, Xiaodong Zhu andSujing Wang " An improved Particle Swarm Optimization for Feature Selection", *Journal of Bionic Engineering*, Vol 8(2), 2011.

[14] Ismail Babaoglu, Oguz Findik, Erkan Ulker and Nazef Aygul, " A Novel Hybrid Classification Method with PSO and K-nn algorithm for diagnosis of Coronary artery disease using exercise stress test data", *International journal of Innovative Computing*, Volume 8(5), May 2012.

[15] Benxian Yue, Weihong Yao, Ajith Abraham and Hongbo Liu, " A New Rough Set Reduct Algorithm based on Particle Swarm Optimization", *IWINAC '07*, LNCS 4527, pp 397-409, © Springer Verlog, 2007.

[16] Mona Nagy Elbedwedhy, Hossam M.Zawbaa, Naveen Ghali and About Ella Hassanien, " Detection of Heart Disease using Binary Particle Swarm Optimization ", *Proceedings of Federated Conference on Computer Science and Information System*, pp 177-182, © IEEE, 2012.

[17] Faridah Fazayeli, Lipo Wang and Jacek Mandziuk, " Feature Selection based on Rough Set Theory and Expectation Maximization Clustering Algorithm", RSCTC 2008, PP 272-282, 2008.

[18] K. Anitha and Dr. P. Venkatesan, " Feature Selection by Rough-Reduct Algorithm", IJIRSET, Vol. 2 (8), 2013.

[19] B.K. Tripathy, D.P. Acharjya and V.Cynthya, " A Framework for Intelligent Medical Diagnosis using Rough Set with Formal Concept Analysis", International Journal of Artificial Intelligence and Applications, Vol. 2 (2), 2011.

[20] Sivagowry.S, Dr. Durairaj. M. (2014). PSO-An Intellectual Technique for Feature Reduction on Heart Malady Anticipation Data. *International journal of Advanced Research in Computer Science and Software Engineering*, *4*(9), 610-621.

[21] Sivagowry.S, Dr. Durairaj.M. "A Pragmatic Approach of Preprocessing the Data Set for Heart Disease Prediction." *International Journal of Innovative Research in Computer and Communication Engineering* 2.11 (2014): 6457-6465.

[22] Sivagowry, S., M. Durairaj, and A. Persia. "An empirical study on applying data mining techniques for the analysis and prediction of heart disease." *Information Communication and Embedded Systems (ICICES), 2013 International Conference on*. IEEE, 2013.

[23] Durairaj. M, Sivagowry. S,"Feature Diminution by Using Particle Swarm Optimization for Envisaging the Heart Syndrome", IJITCS, vol.7, no.2, pp.35-43, 2015. DOI: 10.5815/ijitcs.2015.02.05

