International Journal of Computer Applications (0975 - 8887)
Volume 161 - No.6, March 2017

RGAP: A Rough Set, Genetic Algorithm and Particle Swarm Optimization based Feature Selection Approach

Anupriya Gupta
Computer Engineering Department
Shri G.S Institute of Technology and Science
Indore-452003 (M.P.) India

Anuradha Purohit
Computer Technology and Applications Department
Shri G.S Institute of Technology and Science
Indore-452003 (M.P.) India

ABSTRACT

Feature selection plays an important role in improving the classification accuracy by handling redundant or irrelevant features present in the dataset. Various soft computing based hybrid approaches like neuro-fuzzy, genetic-fuzzy, rough set-neuro etc. are proposed by researchers to perform feature selection. The existing approaches gives higher complexity and computational cost with low classification accuracy. Hence to improve the complexity and classification accuracy, a hybrid approach based on Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Rough Set Theory (RST) to perform feature selection is proposed. In the proposed approach, GA is used as a searching algorithm. To explore search space more efficiently, GA is combined with a PSO based local search operation. Rough Set Attribute Reduction (RSAR) method based on RST is used to compute core reducts. The proposed algorithm is tested on various benchmark datasets. Satisfactory improvements in terms of complexity and classification accuracy have been achieved.

Keywords

Feature Selection, Particle Swarm Optimization, Genetic Algorithm, Rough Set Theory.

1. INTRODUCTION

Feature selection is always an area of interest for the researchers in order to deal with small or large data sets for data mining tasks. Feature selection aims to choose a small number of relevant features to achieve similar or even better classification performance than using all features. Filter and wrapper are two main categories of performing feature selection. Filter method use proxy measure to score a feature subset. On the contrary wrapper methods uses predictive model and are computationally intensive, but usually provide best relevant feature subset [3]. There are many approaches available for performing feature subset selection like piece wise linear network, PLOF'S, PCA, MCES, graph based clustering, soft computing etc. Soft Computing is one of the most widely used approach for feature selection.

Soft computing approach is an innovative approach which does not refer to a single field computation but has many components. For more optimized and efficient results hybrid approaches are developed by researchers, combining different soft computing techniques like artificial neural network, fuzzy inference system, approximate reasoning and optimization methods such as evolutionary computation, swarm optimization, rough sets etc. [5]. The empirical results shows that these hybrid components provide most appropriate approaches, to deal with incomplete and imperfect knowledge. Hence yields more appropriate results as compared to single approaches.

Xiang yang wang et.al in [4] have proposed a new optimal feature selection technique based on rough sets and particle swarm optimization (PSO). In the proposed approach RST based positive region method is used to compute the core reducts. Further PSO was applied to explore the search and the feature selection process. To validate the result and evaluating fitness of the particle RST based fitness function is used. The proposed approach suffers from premature convergence problem due to PSO.

Pradipta Maji in [6] has performed feature selection using rough set theory on fuzzy data sets. In this regard, a novel dimensionality reduction method based on fuzzy-rough sets is presented. To compute the relevant features RST based discernibility matrix method is used for dimensionality reduction technique, which is applied to the fuzzy datasets. To explore the search space another RST based measure of significance method is used. The proposed approach provides efficient result but gives high complexity and computational cost.

Si-Yuan Jing in his paper [1] has discussed a hybrid approach "HGARSTAR", by combining genetic algorithm and rough set theory for performing feature selection. Initially the core features are computed using RST based positive region method. A novel local search operation based on rough set theory is embedded in genetic algorithm to enhance search for better results. Further to fine tune the search the significance of each feature is computed using measure of significance method of RST and to validate the result again rough set based fitness function is used. The proposed approach suffers from high complexity and extensive computational cost due to RST based local search.

In this paper, a hybrid approach "RGAP: A Rough Set, Genetic Algorithm and Particle Swarm Optimization based Feature Selection Approach" is proposed to perform feature selection. In the proposed approach, GA is used as searching technique and to explore the search space more effectively, GA operators are combined with PSO based local search operation. For obtaining more optimized results, RST based RSAR method is used to compute core reducts.

The rest of the paper is organized as follows: Section 2, discusses preliminaries of various techniques used for the proposed method.

Section 3, describes the proposed approach and section 4, presents the experimental results carried on various benchmark datasets. The main conclusion of the work proposed is outlined in section 5.

2. PRELIMINARIES

2.1 Genetic Algorithm

Genetic Algorithms (GAs) are adaptive heuristic search algorithms based on the evolutionary ideas of natural selection and genetics. Genetic algorithm was proposed by Holland in 1970 based on the Darwins theory of Survival of the fittest. GAs are part of evolutionary computing, a rapidly growing area of artificial intelligence [16]. To solve problems in GA search space (all feasible solutions) is explored. The GA algorithm used to perform search and exploration is as follows:

Algorithm of GA:

Step 1: Generate the initial population of individuals.

Step 2: Evaluate the fitness of each individual in the population.

Step 3: Select one or two individuals from the population with a probability based on fitness to participate in genetic operations(Breeding).

Step 4: Create new individuals by applying genetic operations with specified probabilities.

Step 5: If acceptable solution is found or some other stopping condition is met return the best solution else repeat from step 3.

2.2 Rough Set Theory

RST is a mathematical approach developed by Z. Pawlak in 1982 [9]. It expresses vagueness through boundary region. If the boundary region is empty then set is crisp else rough. An indiscernibility relation (R) in RST is described as $R \subseteq U \times U$, where U is the universe of discourse. Let X be a subset of U, and to characterize the set (X) with respect to R, the basic concept of RST can be defined as follows [9]:

(1) Indiscernibility Relation
For an arbitrary set $P \subseteq X$, an indiscernibility relation is defined as given in equation (1).

$$IND(P) = (x, y) \in (UXU) : \forall a \in P, a(x) = a(y) \quad (1)$$

An indiscernibility relation partitions the universe U into disjoint subsets. Let $U/IND(P)$ denote the family of all equivalent classes generated by $IND(P)$. The equivalence classes $U/IND(C)$ and $U/IND(D)$ will be called condition and decision equivalent classes, respectively.

(2) Lower Approximation
The lower approximation of a set X with respect to R is the set of all objects, which can be for certain classified as X with respect to R (are certainly X with respect to R). It can be represented as given in equation (2).

$$B_*(X) = \cup(x \in U) \quad [B(x) : B(x) \subseteq X] \quad (2)$$

Where $B_*(X)$ represent the lower approximation of subset X.

(3) Upper Approximation
The upper approximation of a set X with respect to R is the set of all objects which can be possibly classified as X with respect to R (are possibly X in view of R). It can be represented as given in equation (3).

$$B^*(X) = \cup(x \in U) \quad [B(x) : B(x) \cap X \neq] \quad (3)$$

Where B*(X) represent the lower approximation of subset X.

(4) Boundary Region
The boundary region of a set X with respect to R is the set of all objects, which can be classified neither as X nor as not-X with respect to R. It can be represented as given in equation (4).

$$BN_B(X) = B_*(X) - B^*(X) \quad (4)$$

(a) If $BN_B(X) = \phi$, Then X is crisp (exact) with respect to B.
(b) If $BN_B(X) \neq \phi$, Then X is referred to as rough (inexact) with respect to B.

2.3 Particle Swarm Optimization

Particle Swarm Optimization was introduced by Rousell Eberhart and James Kennedy in 1995 [15]. It is a swarm intelligence based global optimization algorithm. In PSO, a best solution for a given problem is selected which can be represented as a point or surface in an n-dimensional space. The algorithm of PSO is as follows:

Algorithm of PSO:

Step 1: Initialize particles, initial velocity and position randomly.

Step 2: Calculate fitness of each particle.

Step 3: Update local best.

Step 4: Update global best.

Step 5: Update velocity and position.

Step 6: If acceptable solution is found or some other stopping condition is met return the best solution else repeat from step 2.

3. PROPOSED APPROACH

An approach to handle irrelevant and redundant features in the datasets is proposed by combining Rough set theory (RST),Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). In the proposed approach RST is used to calculate the core feature using Rough Set Attribute Reduction (RSAR) method. Genetic Algorithm is used with its selection, mutation and crossover operator to search the space. Particle Swarm Optimization is used to enhance searching ability of GA. The flow diagram of the proposed approach combining these three techniques is as shown in Fig 1. The hybrid process is carried out using following steps:

(1) Compute Core Features
(2) Initialization of Population
(3) Fitness Evaluation of Individuals
(4) Enhancement Process
(5) Generation of New Population
(6) Termination of Process

The steps followed are describe in detail as follows:

(1) *Compute Core Features*
To compute the core features Rough Set Attribute Reduction (RSAR) method of RST is used. In this algorithm, the core features are computed by determining dependency degree of features. The steps designed to compute core features are as follows:
Step 1: Construct an information system $S = (U, C, D, A)$ where $A \subseteq C$, C is condition attribute and D is decision attribute

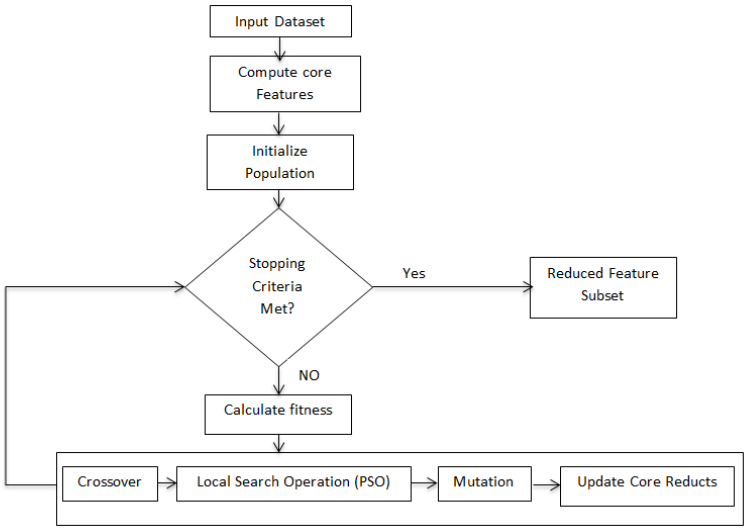

Fig. 1. Flowchart of Proposed Approach

Step 2: Compute indiscernibility relation of the information system.
Step 3: Calculate lower approximation for the subsets in the given indiscernibility relation using the following formula:

$$A_*X = \{x \in U[x]_{IND(P)} \subseteq X\}$$

Where $[x]_{IND(P)} = \{y \in U : a(y) = a(x), \forall a \in P\}$is the equivalence class of x in $U/IND(P)$.
Step 4: Find the positive region of the subsets by applying the formula:

$$POS_P(D) = \bigcup_{X \in U/IND(D)} (P_*X)$$

Step 5: Compute dependency degree of attributes in the subsets by using the given formula:

$$\gamma_P = \frac{|POS|_P(D)}{(|U|)}$$

Where $|A|$ is the cardinality of a set A.
The core features are extracted by intersecting all the subset with dependency degree equals to 1. The core reducts computed in this section are further used for improving fitness of the population.

(2) *Initialization of Population*
After the computation of core reducts, population is initialized using binary encoding technique of GA. In this technique, strings containing 0 and 1 are generated, where 0 represent that the feature is not selected and 1 represent the selection of the feature. Length of the strings that are generated depends on the number of features present in the dataset. The number of strings are generated according to population size. If the population size is given as 10, then 10 strings will be generated and each string will be applied to each attribute present in the dataset and fitness is evaluated for each subset generated by the string.

(3) *Fitness Evaluation of Individuals*
Fitness function is used to evaluate the quality of result generated. To evaluate the significance of features classifier independent fitness function of RST (i.e. dependency degree) is adopted. Fitness function used to calculate fitness of each individual is given in the following formula:

$$Fitness(A) = \alpha.\gamma_A(D) + \beta.(1 - (|A|)/(|C|))$$

Where $A \subseteq C$, $\gamma_A(D)$ is the classification quality of condition attribute set A with respect to decision set D, $\alpha$and β are parameters varies corresponding to the importance of classification quality and subset length, $\alpha \in [0, 1], \beta = 1 - \alpha$. Based on the fitness value evaluated of each individual, enhancement process is applied on individuals to increase their fitness value.

(4) *Enhancement Process*
To enhance the result of an individual on the basis of fitness calculated, a combination of GA operators and PSO is applied to each chromosome. The fitter individuals from the population are selected to produce new offspring using tournament selection method.
After selection of individuals, genetic operators are applied to generate new population. To enhance the fitness of individuals, PSO is applied on each individual. The steps involved in the process are elaborated as follows:

—*Crossover:* After the selection process, multipoint crossover operation is applied to the selected individuals. Suppose two parents are selected, parent 1 is 101110110 and parent 2 is 011001101, applying multipoint crossover on these two parents will generate one new offspring.

—*Particle Swarm Optimization:*The population generated by crossover is passed as the new population to PSO in order to enhance the fitness of generated offsprings (particles). Random velocity and positions are initialized to each offspring. By determining global best among swarm, other particles are updated.
For example, to update velocity and position of the chromosome, the global best is computed from the population. Suppose the global best particle (P_{gbest}) is [1, 0, 1, 0, 0, 1, 1, 0, 1] and there are other particles with their local best (L_{best}) position, like X_i is one of the particle with

Table 1. Datasets used for Experimentation

Datasets	Number of Features	Number of Classes	Number of Instances
Monk1	6	2	124
Monk2	6	2	169
Monk3	6	2	122
Breast Cancer	9	2	699
Tic-Tac-Toe	9	2	958
Zoo	17	7	101
Mushroom	22	2	8124
Lung-Cancer	56	3	32
Soybean (Small)	35	4	47

its L_{best} as [0, 1, 1, 0, 0, 0, 1, 1, 1]. Then to update the velocity of particle $(P_{gbest}X_i)$ is applied to X_i, , which give values [1, 1, 0, 0, 0, 1, 0, -1, 0]. Where, 1 signifies that the feature must be selected, -1 represent that the feature s not necessary for the classification as compared with global best and 0 signifies that the bit does not require any changes. The enhanced offsprings generated by PSO may give constant results. Hence to overcome this drawback of PSO, mutation is applied to the obtained offsprings.

—*Mutation:* The population generated after PSO can be more explored by using mutation operator. Suppose the offspring generated after PSO is 110100111, randomly the 2^{nd} and 8^{th} bit are flipped. If the bit is 1 it will be flipped to 0 and if the bit is '0' it will be flipped to '1'. Hence the new generated offspring will be 100100101.
Further the core reducts computed in the subsection (i) are updated in the offsprings generated after mutation. Fitness of the updated offsprings are again evaluated using fitness function as described in subsection (iii). These newly generated offsprings provide better fitness values than the previous generation. And hence with each iteration the results get closer to the optimal solution.

(5) Termination of Process: To terminate the process, either the individuals in the population must attain 100 % fitness value or predefined number of generations are completed.

4. EXPERIMENTAL RESULT

The proposed approach is implemented in R language using R studio framework. The proposed approach is tested and experimented on the nine benchmark datasets taken from UCI Machine Learning Repository[14]. A brief description of the datasets used is summarized in Table 1.
Best results are achieved on following parameter settings:

(1) Maximum Number of Generations: 100
(2) Population Size: 10
(3) Mutation Probability: 0.1
(4) Crossover Probability: 0.8
(5) Minimum Reduct Subset Size: 4

The results shows that by applying RGAP on the dataset, lesser number of features are selected with good fitness value as given in Table 2.
The core reducts obtained by the RSAR method are represented by "$*$". The complexity of local search operation is dependent on the computation of its three functions.

Table 2. Feature subset Computed and its Fitness Value

Datasets	Selected Feature Subset	Fitness Value
Monk1	1*,2*,5*	0.95
Monk2	1*,2*,3,4,5,6	0.90
Monk3	1*,2*,4*,5*	0.933
Breast Cancer (Original)	1,4,5,6,7	0.92
Tic-Tac-Toe	1,3,5,6,7,9	0.915
Zoo	2,3,4,6*,9,,13*	0.963
Mushroom	1,7,14,20	0.982
Soybean (small)	21,22	0.994
Lung-cancer	4,6,20,23,26,39,40,50	0.986

Table 3. Comparison of Existing Method and Proposed Method

Datasets	Existing Method Accuracy (in %)	Proposed Method Accuracy (in %)
Monk1	93.5	**96.8**
Monk2	65.4	**68.9**
Monk3	99.2	**100**
Breast Cancer (original)	99.3	97.9
Tic Tac Toe	94.3	**96.8**
Zoo	**100**	**100**
Mushroom	**100**	**100**
Lung Cancer	70.0	**75.0**
Soybean (small)	**100**	**100**

The functions are as follows:

(1) Computing fitness value of the feature subset. As given by Wroblewski in [11] the worst time complexity of fitness function used is $\mathcal{O}(nm^2)$.
(2) Evaluating global best particle takes $\mathcal{O}(m+t)$.
(3) Updating velocity of swarm takes $\mathcal{O}(mt)$.

Where n represent number of instances, m represent number of features and t represent number of particles. The worst time complexity calculated for PSO based local search operation of the proposed approach comes out to be $\mathcal{O}(nm^2)$.
Tenfold cross validation method applied to estimate the classification accuracy of the feature subset obtained. These results are compared with results of the existing [1] approach as shown in Table 3.
The worst time complexity of the RST based local search operation used in the existing approach (HGARSTAR) is $O(knm^3)$, where k is the number of iteration, n is number of instances and m is number of features. The worst time complexity computed for PSO based local search operation used in the proposed approach is $(O(nm^2))$. Hence the computational cost of the proposed approach is lesser $O(nm^2)$ as compared to the existing approach $(O(knm^3))$. Also as given in [1] the RST based local search operation gives higher computational cost, but the PSO based local search operation as used in proposed approach gives less complexity with lower computational cost. By applying RGAP the classification accuracy of each dataset is also improved.
The graphical representation of the results obtained is shown in Figure 2. Horizontal axis represents the nine benchmark datasets used in this project and vertical axis represents the accuracy of a classifier (in%).

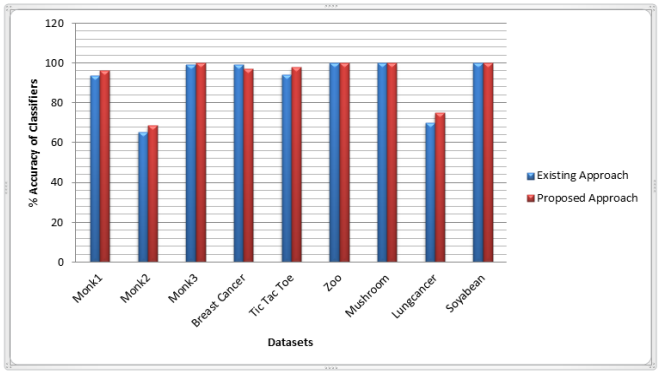

Fig. 2. Comparison between Existing Approach and Proposed Approach

5. CONCLUSION AND FUTURE ENHANCEMENT

In this paper, a soft computing based hybrid approach (RGAP) combining rough set theory, genetic algorithm and particle swarm optimization to perform feature selection is proposed. Rough set theory based, positive region method is used to compute core features of the dataset. Particle swarm optimization based local search operation is embedded with genetic algorithm operators, to explore the search space more efficiently. This local search operation improves the complexity and computational cost of the system, along with the improvement in classification performance. The worst time complexity of local search operation in proposed approach is computed as $\mathcal{O}(nm^2)$ which is better as compared to $\mathcal{O}(knm^3)$ of the existing approach. The classification accuracy is also improved by 0.8-5% for the datasets taken. Hence, as compared to the existing approach, the proposed approach gives better results.

In future more research can be done to deal with real valued datasets, very high dimensional datasets and to work on big datasets also.

6. REFERENCES

[1] Si-Yuan Jing, A Hybrid Genetic Algorithm for Feature Subset Selection in Rough Set Theory, Springer Transaction on Soft Computing: Methodologies and Application, vol. 18, pp.1373-1382, October 2013.

[2] Pedram Ghamisi, Feature Selection Based on Hybridization of Genetic Algorithm and Particle swarm Optimization, IEEE Transaction on Geoscience and Remote Sensing, vol.12, issue 2, pp. 309-313, February 2012.

[3] Yuanning Liu, Gang Wang , Huiling Chen, Hao Dong, Xiaodong Zhu, Sujing Wang, An Improved Particle Swarm Optimization for Feature Selection, Science Direct Transaction on Bionic Engineering, vol. 8, issue 2, pp. 191-200, June 2011.

[4] Xiangyang wang, Jie yang, Xiaolong teng, Weijun Xia, Richard Jensen, Feature Selection based on rough sets and particle swarm optimization, Science Direct Transaction on Pattern Recognition Letter, vol. 28, issue 4, pp. 459-471, March 2007.

[5] K. Jaganath, Mr. P. Sasikumar, Graph Clustering and Feature Selection for High Dimensional Data, International Conference On Global Innovations In Computing Technology, vol. 2, issue 1, pp. 3786-3791, March 2014.

[6] Pradipta Maji, Partha Garai, FuzzyRough Simultaneous Attribute Selection and Feature Extraction Algorithm, IEEE Transaction on Cybernetics, vol. 43, issue 4, pp. 16-1177, August 2013.

[7] Mattia Pedergnana, A Novel Technique for Optimal Feature Selection in Attribute Profiles Based on Genetic Algorithms, IEEE Transaction on Geoscience and Remote Sensing, vol. 51, issue 6, pp. 3514-3528, June 2013.

[8] Roman W. Swiniarski, Rough Sets Methods in Feature Reduction and Classification, International Journal on Applied Mathematics and Computer Science, Vol.11, issue 3, pp. 565-582, 2001.

[9] Z. Pawlak, Rough Sets, International Journal of Computer and Information Sciences, Vol.11, pp. 341-356, 1982.

[10] J. Kennedy and R. C. Ebherhart, "Particle swarm Optimization,"IEEE International Conference on Neural Network, bol. 4, pp. 1942-1948, 1995.

[11] Wroblewski J," Finding minimal reducts using genetic algorithms,"Proceedings of second annual join conference on information sciences, Wrightsville Beach, NC, pp 186189, 1995.

[12] Mark A. Hall, Lloyd A. Smith, "Feature Subset Selection: A Correlation Based Filter Approach," Journal of Machine Learning in University of Waikato, Hamilton, New Zealand, 1997.

[13] Binita kumari, Tripti Swarnakar, Filter Versus Wrapper Feature Subset Selection in Large Dimensionality Microarray: A Review, International Journal of Computer Science and Information Technologies, Vol.2 (3), pp.1048-1053, 2011.

[14] http://archive.ics.uci.edu/ml/

[15] J. Kennedy and R. C. Ebherhart, "Particle swarm Optimization,"IEEE International Conference on Neural Network, vol. 4, pp. 1942-1948, 1995.

[16] Wroblewski J," Finding minimal reducts using genetic algorithms,"Proceedings of second annual join conference on information sciences, Wrightsville Beach, NC, pp. 186189, 1995.