Comparative Analysis of PSO and ACO Based Feature Selection Techniques for Medical Data Preservation

Dhanalakshmi Selvarajan[1], Abdul Samath Abdul Jabar[2], and Irfan Ahmed[3]
[1]Department of Computer Applications and Software Systems, Sri Krishna Arts and Science College, India
[2]Department of Computer Science, Government Arts College, India
[3]Department of Computer Applications, Nehru Institute of Engineering and Technology, India

Abstract: *Sensitive medical dataset consist of large number of disease attributes or features, not all these features are used for diagnosis. In order to preserve the medical dataset it is not essential to perturb all the features before it is shared for mining purpose. To reduce the computational cost and to increase the efficiency, in this work tried to use Ant Colony Optimization (ACO) for feature subset selection which is used to reduce the dimension and also compared with feature subset selection using Particle Swarm Optimization (PSO) which is also used to reduce the dimension. Both the techniques are explored to reduce the dimension before applying preservation technique. By using randomization method a known distribution is added to the reduced sensitive data before the data is sent to the miner. The approach is analyzed using standard UCI medical datasets. The result is analyzed based on classification accuracy using machine learning algorithms (Naïve Bayes, Decision Tree) build on the randomized dataset. The experimental results show that the accuracy is maintained in the reduced perturbed datasets. The results also show that ACO search based feature selection has more accuracy than PSO search based selection.*

Keywords*: Randomization, particle swarm optimization, ant colony optimization, feature selection.*

Received October 9, 2015; accepted November 9, 2016

1. Introduction

Many medical dataset consist of large number of attributes making the privacy preservation process difficult. Huge number of features in dataset makes the preservation process slower and time consuming. To overcome this problem the work focused on finding a promising feature selection mechanism before applying the perturbation technique. In this work tried to analyze the feature selection strategy based on rough sets and optimization techniques Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) and then a simple randomization technique is applied to the reduced data set.

Feature selection is the process of selecting only relevant features from the dataset for model construction. A feature selection process can be used to remove features in the training dataset that are not correlated to the class labels and thus reduce the number of terms used in classification, thus improving both efficiency and accuracy [8]. In this paper used a fuzzy rough set attribute reduction algorithm that includes Particle Swarm Optimization and Ant Colony Optimization based search method. Both the methods are compared based on classification performance.

The concept of rough set deals with uncertain and imprecise data [21]. The rough set based feature selection removes features and selects only the appropriate feature subset that has the same as the original set of features. Various works has been carried out using rough set and fuzzy for feature selection [12, 14, 24].

In this paper the fuzzy rough set attribute reduction algorithm is included with a search method based on ACO and is compared with PSO based fuzzy rough set reduction. PSO is a population-based method for solving optimization problems. In PSO particles moves in the search space of an optimization problem. The particle position represents a candidate solution to the optimization problem [6]. It is an extremely simple algorithm that seems to be effective for optimizing a wide range of functions. It seems to lie somewhere between genetic algorithm and evolutionary programming [9].

The work integrate fuzzy rough set attribute reduction algorithm with a search method based on ACO. ACO is a meta heuristic in which a colony of artificial ants cooperates in finding good solutions to difficult discrete optimization problems. Static and dynamic combinatorial optimization problems can be solved using ACO algorithms. An artificial ant in ACO procedure tries to find a solution by incrementally adding defined solution components to the partial solution under construction. The main idea is to incorporate the principles of coordinated behavior of real ants to solve the computational problems by exploiting the ants behavior to coordinate the populations of artificial agents [4, 5].

Various preservation methods have been adopted for protecting the sensitive data [17]. In this paper tried to calculate attributes that should appear in every valid reduced dataset using fuzzy rough feature selection based on Ant search and PSO search. The reduced dataset is randomized and checked for accuracy before and after transformation. The paper is organized as follows. In section 2 addressed the basic concepts and in section 3 discussed the proposed work and the solution and evaluation results are addressed in section 4. Finally conclusion is addressed in section 5.

2. Basic Concepts

2.1. Fuzzy Rough Feature Selection

Rough set theory only by using the data without any additional information can be used as a tool to discover data dependencies and also can be used to reduce the number of attributes contained in the dataset [24]. Attribute reduction is achieved by comparing equivalence relations generated by attributes sets. The main limitation of rough set theory is that the method can be applied only for discrete values.

Fuzzy-rough sets provide a way by which real-valued or discrete data can also be reduced. Fuzzy based reduction requires additional information in the form of fuzzy partitions for each attribute which can be derived from the data automatically. It is a highly useful technique in reducing dimensionality of data.

Fuzzy equivalence classes are most important to the fuzzy rough set approach. There are two main approaches in the hybridization of rough sets and fuzzy. The approaches are constructive approach and axiomatic approach. In the constructive approach based on fuzzy relations, a generalized lower and upper approximations are defined. In axiomatic approach using different set of axioms various classes of fuzzy rough approximation operators are characterized. From the literature the fuzzy p-lower and p-upper approximations are defined as [13].

$$\mu\_P\ X(Fi) = inf\_x\ max\{1\text{-}\mu Fi(x),\mu X(x)\}\ \ \forall i \quad (1)$$

$$\mu^{-}P\ X(Fi) = sup\_x\ min\{\mu Fi(x),\mu X(x)\}\ \ \forall i \quad (2)$$

Where *Fi* is a fuzzy equivalence class, and *X* is the concept to be approximated.

In traditional roughset attribute reduction the reduct dataset is consistent that is the reduced subset of attributes have same information as that of full attribute set. But in fuzzy-rough approach the total dependency is reduced when objects belong to many fuzzy equivalence classes. Fuzzy rough hill climbing search algorithm has been developed to solve this issue. In this algorithm the fuzzy rough dependency function choose the attributes that add to the reduct in a manner similar to Quick Reduct. When addition of remaining attributes does not increase the dependency of the algorithm terminates. The Quick Reduct (QR) algorithm [2] attempts to calculate a reduct without exhaustively generating all possible subsets [11]. The process starts with an empty set and adds one attribute at a time. The attributes which satisfy greatest increase in rough set dependency metric are added. The process continues until maximum dataset value is generated. The QR Algorithm [2] is mentioned below.

Algorithm 1: Quickreduct(C,D)

Input:C, the set of all conditional features,
D the set of decision features.
Output:R,the attribute reduct,R ⊆ C
1) R ← ∅
(2)do
(3) T ← R
(4)for each x ∈(C-R)
(5)if γR∪{x}(D)> γT (D)
(6) T ←R ∪ { x }
(7) R ←T
(8) until γR(D)= γC (D)
(9)return R

2.2. Particle Swarm Optimization

PSO is motivated by the behavior of flocks of birds or swarms of fishes to search for a good food place. The coordination of movements of the individuals in the swarm is the central aspect that inspires PSO [20].

Eberhart and Kennedy [7] developed an evolutionary computation technique called particle swarm optimization [6, 15]. Later Shi and Eberhart [22] introduced interia weight into particle swarm optimizer to PSO. Algorithm is initialized with a population of particles having a random position. The particles are associated with velocity. According to the behavior of particles its velocity is adjusted. The velocity is calculated as mentioned below [9]:

$$v[\ ]=c0*v[\ ]+c1*rand(\ )*(pbest[\ ]-present[\ ])+ c2*rand(\ )*(gbest[\ ]-present[\ ]) \quad (3)$$

$$Present[\ ]=present[\ ] + v[\ ] \quad (4)$$

ci:weighting factor

v[ ]:velocity of agent at iteration

rand:uniformly distributed random number between 0 and 1

pbest:pbest of agent

gbest:gbest of the group

present[ ]:current position of agent

A PSO algorithm maintains particle swarm population where each particle denotes a location in multidimensional problem space. The particles start from random locations and search for maximum or minimum objective function by moving through the problem space. The function is measured by the quality or the amount of food available at each location. The particle swarm tries to find the location with the best food or with more food available. The movements of a particle swarm are determined by its velocity and the locations where good solutions have already been found by the particle itself or other (neighboring) particles in the swarm [20]. Pseudo code for PSO is mentioned below [9].

```
For each particle
    Particle is initialized
End
Do
For each particle
    Fitness value is calculated
If  fitness value is better than  personal best
        current value is set as the new pBest
 End
Choose the particle with the best fitness value of all
as gBest
For each particle
   particle velocity is calculated using Equation (3)
   particle position is updated using Equation (4)
 End
Uuntil maximum iterations or minimum error
criteria is not attained
```

In PSO algorithm the coordinates in the search space that are associated with the best solution is tracked by each particle. The corresponding fitness value of the objective function is stored. The best value obtained by each particle is tracked with its topological neighborhood. When a particle considers the whole population as its neighbours, the best values is known as global best. PSO algorithm at each iteration changes the velocity of each particle towards the neighborhood best or global best locations. A random component is also included into the velocity update [20]. Various works has been carried out using PSO for feature selection [16, 23, 25].

2.3. Ant Colony Optimization

ACO concept is inspired by how the real ants find the shortest paths from their nest to the food source location. A chemical substance is released by the ants into the environment when they are moving towards the food location. This helps in indirect communication to other ants about the species and also the chemical substance released indicates their paths towards the food sources. The chemical substances released can be smelled by other ants and helps them to reach the food source.

Thus the ant colony optimization algorithms are derived from the observation of ants behavior. This become the inspiration to derive new algorithms for discovering solution to optimization and distributed control problems [4].

The self-organizing principles which makes the coordinated behavior of real ants possible is explored to co-ordinate the populations of artificial agents that collaborate to solve computational problems. Many aspects of ant colonies behavior has lead to different kinds of ant's algorithm [4].

The requirement for designing an ACO algorithm is to have a constructive method such as used by an ant to create different solutions through a sequence of decisions. Typically the solution to an ACO is to construct a sequence of probabilistic decisions by extending a partial solution by adding a new solution component until the complete solution is reached.

The aim is to allow the artificial ants to find paths through the decision graph that yields good solutions. This is done in an iterative process. Thus ants have to find a good solution and are allowed to mark the edges of the specific path in the decision graph. Intermediate path guides the following ants in the next iteration to search the nearest path that leads to a good solution.

ACO procedure follows an iterative process in which the information is transferred from one iteration to the next one. The process continues until some stopping criterion is met: e.g., until a certain number of iterations have been done or a solution of a given quality has been found. ACO algorithm is given below [19]. Various works has been carried out using ACO for feature selection [1, 5, 7, 10].

Algorithm 2: ACO Algorithm

```
Initialize pheromone values
repeat
    for ant A: e ( 1 , . . . ,m}
          construct a solution
    end for
    for all pheromone values do
        {evaporation} the value is decreased by
         certain percentage
    end for
    for all pheromone values corresponding to good
   solutions
   do
     {intensification} the value is increased
   end for
until stopping criterion is met
```

3. Proposed Work

Feature selection is the most relevant features for classification problems and it is one of the main data mining tasks. For all search methods need a performance evaluation to decide how well a feature subset will probably perform on the given data set. In this paper used PSO and ACO incorporated with fuzzy rough set for feature subset selection.

In the work compared the dimension reduction using fuzzy rough search based with optimization techniques PSO [3] and ACO for data preservation

The first step is to reduce the feature set by applying two techniques. In the first approach a feature extraction is made using fuzzy rough set attribute reduction algorithm that incorporates a search based on PSO [3]. A number of particles are initialized at random locations (which correspond to feature subsets) and then swarm move towards promising areas via the global best solution so far and each particle's local best. The smallest subset found overall with maximum quality is returned.

In the second approach the feature extraction is explored by using fuzzy rough set along with ACO. For each generation, ants start off at a random feature and move probabilistically until there is no improvement in their constructed subset quality. The smallest subset found overall with maximum quality is returned.

Next step is to randomize the retrieved most relevant attributes of the data set. Let $m_1,m_2,m_3,\ldots.m_n$ be the attribute values of a transformed data set. Let d_1, d_2, ….,d_n be the random distribution values used to distort the transformed data set. The random distribution values r_1, r_2, r_3, …,r_n are generated using uniform distribution method. The random variable has a uniform distribution over an interval $-\alpha$ to $+\alpha$. The mean of the random variable is 0. The perturbed data set contains p_1, p_2, p_3,…,p_n. Where p_1, p_2….., p_n is generated by adding the random distribution values with the optimized transformed data set. The data mining models are constructed on the optimized perturbed data set and compared the results with the original data set results. The transformed dataset are evaluated based on the classification accuracy. The proposed transformation method is predicted in Figure1.

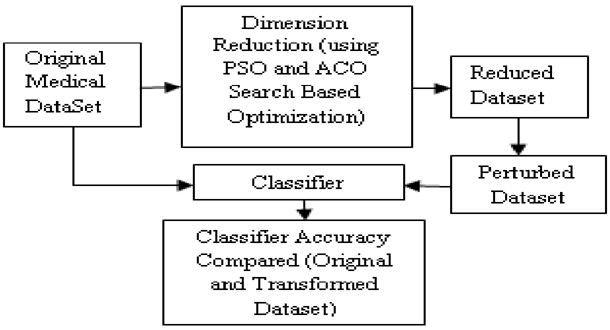

Figure 1. Proposed transformation method.

In the proposed solution the original UCI medical data set which contain n attributes are transformed in to reduced dataset which contains m attributes where m<n. Two reduced transformed datasets are generated by using rough set and PSO based search and also by using ACO based search feature selection. The transformed data set contains m attributes where m < n. The next step is random distribution values r_1, r_2,...,r_n are generated using uniform distribution method. Perturbed data set contains attributes p_1,p_2…..,p_n is generated by adding the random distribution values with the optimized transformed data set.

The data mining models like Naïve Bayes and decision tree are constructed from the perturbed data set directly. The effectiveness of both the techniques on different real world datasets are compared and the evaluation results show that ACO based optimized and perturbed dataset maintains more classification accuracy than PSO search based feature reduction

4. Implementation Results

The goal is to protect the individual's privacy in the medical dataset and also to obtain accurate data mining results. To analyze the result of privacy preserving data mining based on classification accuracy, decision trees and Naive Bayes classifiers are build on real-world medical data sets. In the proposed work for privacy preserving transformation the following UCI medical data sets [18] is used in the experiments.

- *Dermatology data set*: Dermatology data set is used to determine the type of Eryhemato-Squamous Disease. This database contains 34 attributes and 366 instances. They all share the clinical features of erythema and scaling, with very little differences.
- *Breast data set*: This breast cancer databases was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg. The original data set has 699 instances which include 10 attributes and a class label.

The reduced attributes using the concept of fuzzy rough set with PSO and ACO based search are listed in Table 1.

Table 1. Reduced attributes.

UCI Dataset	Total No of attributes	Feature Selection Algorithms	Reduced Attributes
Dermatology	34	PSO	12
Breast	10	PSO	7
Dermatology	34	ACO	22
Breast	10	ACO	7

The randomization technique is applied to the reduced attributes data set and perturbed the original features. Rapid Miner data mining software is used to run decision tree and Naive Bayes classifiers on original and perturbed reduced data set of UCI medical data sets.

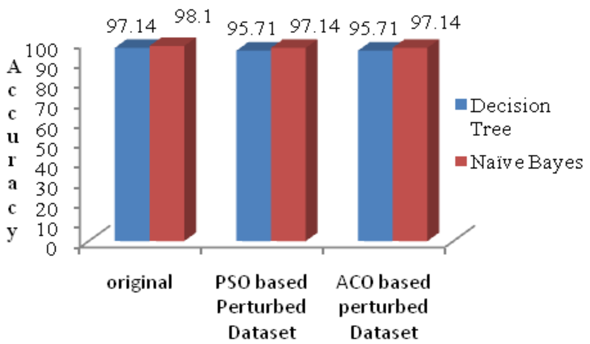

Figure 2. Classifier accuracy (%) on original and optimized-perturbed breast dataset.

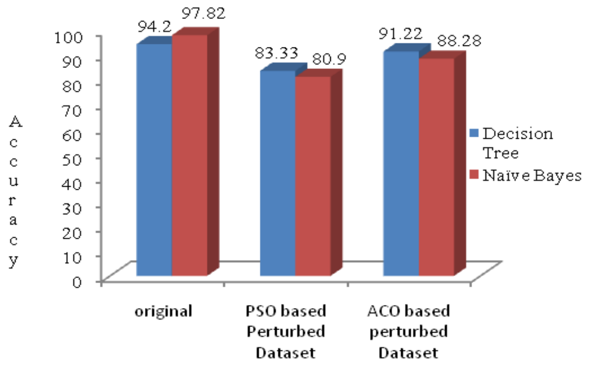

Figure 3. Classifier accuracy (%) on original and optimized-perturbed dermatology database.

The experiment results are shown in Figures 2 and 3. It clearly shows that when the total number of attributes is less the PSO and ACO reduction based perturbation gives more accurate results than the data set with more number of attributes. For datasets with more features ACO based perturbation gives more classification accuracy than PSO based perturbation.

5. Conclusions

In this paper compared a feature selection strategy based on rough sets and optimization techniques. PSO and ACO search based methods are applied to the UCI Medical datasets to reduce the features of the dataset. Simple randomization technique is applied to the reduced data set. To analyze the accuracy of the results decision trees and Naive Bayes classifiers are build on the reduced perturbed dataset. The comparative results show that the accuracy was almost equal to that of original data set. When compared the feature selection strategy based on both PSO and ACO, it clearly shows that ACO Based feature selection method gives more accurate results than PSO based feature selection method.

References

[1] Abd-Alsabour N. and Randall M., “Feature Selection for Classification Using an Ant Colony System,” *in Proceedings of 6th IEEE International Conference on E-Science Workshops*, Brisbane, pp. 86-91, 2010.

[2] Chouchoulas A. and Shen Q., “Rough Set-Aided Keyword Reduction for Text Categorization,” *Applied Artificial Intelligence*, vol. 15, no. 9, pp. 843-873, 2001.

[3] Dhanalaksmi S., Abdul Samath J., and Prashanth R ., “Fuzzy Rough Set and PSO Search Based Feature Selection for Data Preservation,” *International Journal of Applied Engineering Research*, vol. 10, no. 20, pp. 17767-17770, 2015

[4] Dorigo M., *Ant Colony Optimization*, Scholarpedia, 2007.

[5] Dorigo M. and Thomas S., “The Ant Colony Optimization Metaheuristic: Algorithms, Applications, and Advances,” *Handbook of Metaheuristics*, Boston, pp. 250-285, 2003.

[6] Dorigo M., Marco A., Montes O., and Engelbrecht A., *Particle Swarm Optimization*, Scholarpedia, 2008.

[7] Eberhart R. and Kennedy J., “A New Optimizer Using Particle Swarm Theory,” *in Proceedings of the 6th International Symposium on Micro Machine and Human Science*, Nagoya, pp. 39-43, 1995.

[8] Han J. and Kamber M., *Data Mining Concepts and Techniques*, Morgan Kaufmann, 2006.

[9] HU X., “Particle Swarm Optimization: Tutorial [online]. 2006 [cit. 2011-05-09].” PSO Tutorial.WWW:<http://www. swarmintelligence. org/tutorials. php, Last Visited, 2006.

[10] Jaganathan P., Thangavel K., Pethalashmi A., and Karnan M., “Classification Rule Discovery with Ant Colony Optimization and Improved Quick Reduct Algorithm,” *International Journal of Computer Science*, vol. 33, no. 1, pp. 50-55, 2007.

[11] Jensen R. and Shen Q., “Fuzzy-Rough Attribute Reduction with Application to Web Categorization,” *Fuzzy Sets and Systems*, vol. 141, no. 3, pp. 469-485, 2004.

[12] Jensen R. and Shen Q., “New Approaches to Fuzzy-Rough Feature Selection,” *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824-838, 2009.

[13] Jensen R. and Shen Q., *Computational Intelligence and Feature Selection: Rough and Fuzzy Approaches*, John Wiley and Sons, 2008.

[14] Jensen R. and Shen Q., “Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 12, pp. 1457-1471, 2004.

[15] Kennedy J. and Eberhart R., “Particle Swarm Optimization,” *in Proceedings of IEEE International Conference on Neural Networks*, Perth, pp. 1942-1948, 1995.

[16] Lai C., Wu C., and Tsai M., “Feature Selection Using Particle Swarm Optimization with Application in Spam Filtering,” *International Journal of Innovative Computing, Information and Control*, vol. 5, no. 2, pp. 423-432, 2009.

[17] Li G. and Wang Y., “A Privacy-Preserving Classification Method Based on Singular Value Decomposition,” *The International Arab Journal of Information Technology*, vol. 9, no. 6, pp. 529-534, 2012.

[18] Machine Learning Repository, archive.ics.uci.edu/ml/datasets.html, Last Visited, 2014.

[19] Martens D., Backer M., Haesen R., Member S., Vanthienen J., Snoeck M., and Baesens B.,

"Classification with Ant Colony Optimization," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 5, pp. 651-665, 2007.

[20] Merkle D. and Middendorf M., "Swarm intelligence," *in Proceedings of Search Methodologies*, Boston, pp. 401-435, 2005.

[21] Pawlak Z., *Rough Sets: Theoretical Aspects of Reasoning About Data*, Springer, 1991.

[22] Shi Y. and Eberhart R., "A Modified Particle Swarm Optimizer," *in Proceedings of IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence*, Anchorage, pp. 69-73, 1998.

[23] Wang X., Yang J., Teng X., Xia W., and Jensen R., "Feature Selection Based on Rough Sets and Particle Swarm Optimization," *Pattern Recognition Letters*, vol. 28, no. 4, pp. 459-471, 2007.

[24] Wang X., Yang T., Jensen R., and Liu X., "Rough Set Feature Selection and Rule Induction for Prediction of Malignancy Degree in Brain Glioma," *Computer Methods Programs Biomedicine*, vol. 83, no. 2, pp. 147-156, 2006.

[25] Xue B., Zhang M., and Browne W., "Multi-Objective Particle Swarm Optimisation (PSO) For Feature Selection," *in Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation*, Philadelphia, pp. 81-88, 2012.

Dhanalakshmi Selvarajan had received her Ph.D Degree in computer applications from Anna University. Her research interest includes privacy preserving data mining, security issues in databases.

Abdul Samath Abdul Jabar had received his Ph.D Degree from Gandhigram Rural University. He is a member in several professional bodies likes CSI, ISCA, LIA, INNS, IAENG and ISTE. His research interest includes neural networks, simulation, image processing, data mining.

Irfan Ahamed had received his Ph.D Degree from Alagappa University. He is also a member in several professional bodies. His research interests include networks, image processing and data mining.