Rough ACO: A Hybridized Model for Feature Selection in Gene Expression Data

Debahuti Mishra*

Institute of Technical Education & Research (ITER)
Siksha 'O' Anusandhan University,
Bhubaneswar, Orissa
Email: debahuti@iter.ac.in
*Corresponding author

Dr. Amiya Kumar Rath

College of Engineering Bhubaneswar (CEB)
Bhubaneswar, Orissa
Email: amiyaamiya@rediffmail.com

Dr. Milu Acharya

Institute of Technical Education & Research (ITER)
Siksha 'O' Anusandhan University,
Bhubaneswar, Orissa
Email: milu_acharya@yahoo.com

Tanushree Jena

College of Engineering Bhubaneswar (CEB)
Bhubaneswar, Orissa
Email: tanushreejenasmailbox@rediffmail.com

Abstract: Dimensionality reduction of a feature set is a common preprocessing step used for pattern recognition, classification applications and in compression schemes. Rough Set Theory is one of the popular methods used, and can be shown to be optimal using different optimality criteria. This paper proposes a novel method for dimensionality reduction of a feature set by choosing a subset of the original features that contains most of the essential information, using the same criteria as the ACO hybridized with Rough Set Theory. We call this method Rough ACO. The proposed method is successfully applied for choosing the best feature combinations and then applying the Upper and Lower Approximations to find the reduced set of features from a gene expression data.

Keywords: Feature Selection; Rough Sets; Lower Approximation; Upper Approximation; Data Preprocessing; Ant Colony Optimization.

Reference to this paper should be as follows: Mishra, D. ,Rath., A.K, Acharya., M and Jena., T.(2009) 'Rough ACO: A Hybridized Model for Feature Selection in Gene Expression Data', *Int. J.CCT*, Vol. 1, No. 1, 2009, pp.85-98.

Biographical notes: Debahuti Mishra is an Assistant Professor and research scholar in the department of Computer Sc. & Engg, Institute of Technical Education & Research (ITER) under Siksh `O` Anusandhan University, Bhubaneswar.She received her Masters degree from KIIT University, Bhubaneswar.Her research areas include Data mining, Bio-informatics Software Engineering,Soft computing . She is an author of a book Aotumta Theory and Computation by Sun India Publication (2008).

Dr.Amiya Kumar Rath obtained Ph.D in Computer Science in the year 2005 from Utkal University for the work in the field of Embedded system. Presently working with College of Engineering Bhubaneswar (CEB) as Professor of Computer Science & Engg. Cum Director (A&R) and is actively engaged in conducting Academic, Research and development programs in the field of Computer Science and IT Engg. Contributed more than 30 research level papers to many national and International journals. and conferences Besides this, published 4 books by reputed publishers. Having research interests include Embedded System, Adhoc Network,Sensor Network ,Power Minimization, Biclustering, Evolutionary Computation and Data Mining.

Dr. Milu Acharya obtained her Ph.D (Utkal University), she is Professor in Department of Computer Applications at Institute of Technical Education and Research (ITER), Bhubaneswar. She has contributed more than 20 research level papers to many national and International journals. and conferences Besides this, published 3 books by reputed publishers. Her research interests include Biclustering, Data Mining , Evaluation of Integrals of analytic Functions , Numerical Analysis , Complex Analysis , Simulation and Decision Theory.

Tanushree Jena obtained her Mtech from KIIT University, Bhubaneswar. She is presently working as Sr.Lecturer in the Dept. of Computer Sc. & Engg. College of Engineering Bhubaneswar, She has contributed many papers in national and International Conferences and Journals to her credit Her research area includes Data mining, Software Engineering.

1. Introduction

Data mining refers to extracting or "mining" knowledge from large amounts of data. There are many other terms carrying a similar or slightly different meaning to Data mining, such as knowledge mining from databases, knowledge extraction, data pattern analysis, data archaeology, and data dredging. Data mining treats as synonym for another popularly used term, Knowledge Discovery in Databases (KDD) KDD consists of the following steps to process it such as Data cleaning, Data integration, Data selection, Data transformation, Data mining, Pattern evaluation and Knowledge presentation.

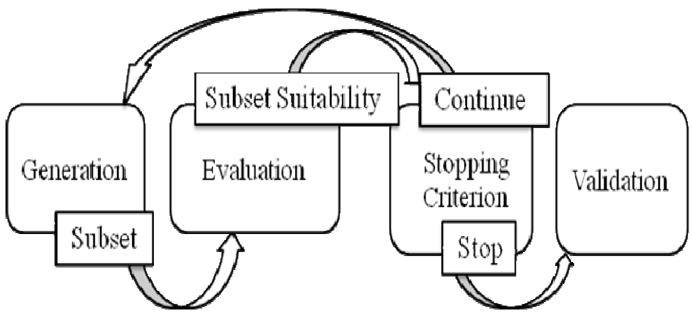
The main aim of feature selection (FS) [1] is to determine a minimal feature from a problem domain while retaining a suitably high accuracy in representing the original features. In many real world problems FS is a must due to the abundance of noisy, irrelevant or misleading features. For instance, by removing these factors, learning from data techniques can benefit greatly.

The usefulness of a feature or feature subset is determined by both its *relevancy* and *redundancy*. A feature is said to be relevant if it is predictive of the decision feature(s), otherwise it is irrelevant. A feature is considered to be redundant if it is highly correlated with other features. Hence, the search for a good feature subset involves finding those features that are highly correlated with the decision feature(s), but are uncorrelated with each other.

Feature selection algorithms may be classified into two categories based on their evaluation procedure. If an algorithm performs FS independently of any learning algorithm (i.e. it is a completely separate preprocessor), then it is a *filter* approach. In effect, irrelevant attributes are filtered out before induction. Filters tend to be applicable to most domains as they are not tied to any particular induction algorithm.

If the evaluation procedure is tied to the task (e.g. classification) of the learning algorithm, the FS algorithm employs the *wrapper* approach. This method searches through the feature subset space using the estimated accuracy from an induction algorithm as a measure of subset suitability. Although wrappers may produce better results, they are expensive to run and can break down with very large numbers of features. This is due to the use of learning algorithms in the evaluation of subsets, some of which can encounter problems when dealing with large datasets. The Figure 1 shows the steps required for feature selection.

Figure 1 Feature Selection



1.1 Goal of this Paper

In this paper, we are concerned on how to construct a heuristic function for feature selection. Our objective is to focus on how to improve the time efficiency of a heuristic feature subset selection algorithm. We employ a new rough set framework hybridized with ACO, which is called positive approximation. The main advantage of this approach stems from the fact that this framework is able to characterize the granulation structure of a rough set using a granulation order. Based on the positive approximation, we develop a common strategy for improving the time efficiency of a heuristic feature selection, which provides a vehicle of making algorithms of rough set based feature selection techniques faster.

1.2 Layout of the Paper

The layout of the paper is as follows: Section 1 deals with introductory part of data mining and feature selection for gene expression data with some related work on this approach followed by the goal of our proposed approach. The section 2 describes the preliminary concepts of Ant Colony Optimization (ACO) and Rough Set Theory (RST). Section 3 describes the proposed model along with the ACO framework and feature selection based on our approach. The Section 4 deals with proposed algorithms. The section 5 explains its experimental verification. The Section 6 gives the conclusion followed by future work.

2. Preliminaries

2.1 Ant Colony Optimization

Ant colony optimization (ACO) Meta heuristic, a novel population-based approach was recently proposed in to solve several discrete optimization problems [2]. The ACO mimics the way real ants find the shortest route between a food source and their nest. The ants communicate with one another by means of pheromone trails and exchange information about which path should be followed. The more the number of ants traces a given path, the more attractive this path (trail) becomes and is followed by other ants by depositing their own pheromone. This auto catalytic and collective behavior results in the establishment of the shortest route.

Ants find the shortest path from their nest to the food source with the help of pheromone trail. This characteristic of ants is adapted on ant colony optimization algorithms to solve real problems with using exactly some characteristics of ants and some new addition The method improved by modeling real ants use exactly the same specifications taken from real ants are below:

- The communication established with ants through pheromone trail
- Paths deposited more pheromone preferred previously
- Pheromone trail on short paths increase more rapidly.

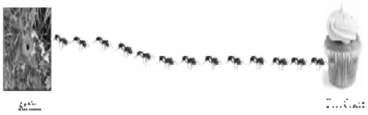
Addition of new specifications to this new technique is below:

- They live in an environment where time is discrete
- They will not be completely blind, they will reach the details about the problem
- They will keep information formed for the solution of the problem with has some memory.

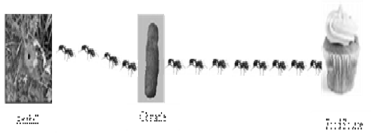
As shown in Figure 2-a, ants start from their nest and goes along a linear path through the food source.

a. Ants following a path between their nest and food source
b. Encountering a obstacle of ants
c. Selection of ants
d. Finding the shortest path of ants

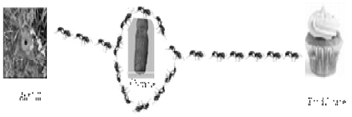
Figure 2 Behaviors of real ants between their nest and food source



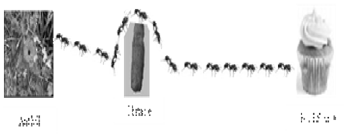
(a)



(b)



(c)



(d)

Actually, if there exists a difficulty on the path while going to the food source (Figure 2-b), ant lying in front of this difficulty can not continue and has to account a preference for the new outgoing path. In the present case, selection probability of the new direction alternatives of ants is equal. In other words, if ant can select anyone of the right and left directions, the selection chance of these directions is equal (Figure 2-c). Namely, two ants start from their nest in the search of food source at the same time to these two directions. One of them chooses the path that turns out to be shorter while the other takes the longer path. But it is observed that following ants mostly select the shorter path because of the pheromone concentration deposited mostly on the shorter one.

The ant moving in the shorter path returns to the nest earlier and the pheromone deposited in this path is obviously more than what is deposited in the longer path. Other ants in the nest thus have high probability of following the shorter route. These ants also deposit their own pheromone on this path. More and more ants are soon attracted to this path and hence the optimal route from the nest to the food source and back is very quickly established. Such a pheromone-meditated cooperative search process leads to the intelligent swarm behavior.

The instrument of ants uses to find the shortest path is pheromone. Pheromone is a chemical secretion used by some animals to affect their own species. Ant deposit some pheromone while moving, they deposit some amount of pheromone and they prefer the way deposited more pheromone than the other one with a method based on probability. Ants leave the pheromone on the selected path while going to the food source, so they help following ants on the selection of the path (Figure 2-d).

There are many algorithms derived from ant colony meta-heuristic and they are used on solution of many problems. These algorithms are derived from each other as formulation but all use the common specifications of ant colony meta-heuristic.

Generally, in ant colony optimization algorithms, operations described above are iterated in main loop until a certain number of iterations are completed or all ants begin to generate the same result. This situation is named as *stagnation behavior*, because after a point, algorithm finishes to generate alternative solutions. The reason of this situation is, after a certain number of iterations, ants generate continuously the same solutions because pheromone amount intensifies in some points and the difference between pheromone concentrations on paths become very huge.

Most ant colony optimization algorithms use this algorithmic diagram demonstrated below:

Initiation of the parameters which determines the pheromone trail
While *(until result conditions supplied)* ***do***
    Generate Solutions
    Apply Local Search
    Update Pheromone Trail
End

2 Rough Set Theory

Rough set theory [3][4] is a new mathematical approach to imprecision, vagueness and uncertainty. In an information system, every object of the universe is associated with some information. Objects characterized by the same information are indiscernible with respect to the available information about them. Any set of indiscernible objects is called an elementary set. Any union of elementary sets is referred to as a crisp set- otherwise a set is rough (imprecise, vague). Vague concepts cannot be characterized in terms of information about their elements. A rough set is the approximation of a vague concept by a pair of precise concepts, called lower and upper approximations. The lower approximation is a description of the domain objects, which are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects that possibly belong to the subset. Relative to a given set of attributes, a set is rough if its lower and upper approximations are not equal.

The main advantage of rough set analysis is that it requires no additional knowledge except for the supplied data. Rough sets perform feature selection using only the granularity structure of the data [5]. Let $I= (U, A)$ be an information system, where U is the universe, a non-empty finite set of objects. A is a non-empty finite set of attributes. For $\forall a \in A$ determines a function f_a: $U \rightarrow V_a$. If $P \subseteq A$, there is an associated equivalence relation:

$$IND(P) = \{(x, y) \in U \times U \mid \forall a \in P,\ f_a(x) = f_a(y)\} \quad (1)$$

The partition of U, generated by $IND(P)$ is denoted U/P. If $(x, y) \in IND(P)$, then x and y are indiscernible by attributes from P. The equivalence classes of the P-indiscernibility relation are denoted $[x]_p$. The indiscernibility relation is the mathematical basis of rough set theory.

Let $X \subseteq U$, the P-lower approximation $\underline{P}X$ and P-upper approximation PX of set X can be defined as:

$$\underline{P}X = \{ x \in U \mid [x]_p \subseteq X \} \quad (2)$$

$$PX = \{ x \in U \mid [x]_p \cap X \neq \varphi \} \quad (3)$$

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

$$POS_p(Q) = \bigcup_{x \in U|Q} \underline{P}X \quad (4)$$

$$NES_p(Q) = U - \overline{\bigcup}_{x \in U|Q} PX \quad (5)$$

$$BND_p(Q) = \bigcup_{x \in U|Q} PX - \overline{\bigcup}_{x \in U|Q} PX \quad (6)$$

The positive region of the partition U/Q with respect to P, $POS_p(Q)$, is the set of all objects of U that can be certainly classified to blocks of the partition U/Q by means of P. A set is rough (imprecise) if it has a non-empty boundary region.

An important issue in data analysis is discovering dependencies between attributes. Dependency can be defined in the following way. For P, $Q \subseteq A$, P depends totally on Q, if and only if $IND(P) \subseteq IND(Q)$. That means that the partition generated by P is finer than the partition generated by Q. We say that Q depends on P in a degree $0 \leq k \leq 1$ denoted $P \Rightarrow_k Q$, if

$$k = \gamma_p(Q) = \frac{|POS_p(Q)|}{|U|} \quad (7)$$

If $k=1$, Q depends totally on P, if $0 \leq k \leq 1$, Q depends partially on P, and if $k=0$ then Q does not depend on P. In other words, Q depends totally (partially) on P, if all (some) objects of the universe U can be certainly classified to blocks of the partition U/Q, employing P.

In a decision system the attribute set contains the condition attribute set C and decision attribute set D, i.e. $A = C \cup D$. The degree of dependency between condition and decision attributes, $\gamma_c(D)$, is called the quality of approximation of classification, induced by the set of decision attributes [3].

The goal of attribute reduction is to remove redundant attributes so that the reduced set provides the same quality of classification as the original. A reduct is defined as a subset R of the conditional attribute set C such that $\gamma_r(D) = \gamma_c(D)$. A given decision table may have many attribute reducts; the set of all reducts is defined as:

$$Red = \{ R \subseteq C \mid \gamma_r(D) = \gamma_c(D), \forall B \subset R, \ \gamma_B(D) \neq \gamma_c(D) \} \quad (8)$$

In rough set attribute reduction, a reduct with minimal cardinality is searched for. An attempt is made to locate a single element of the minimal reduct set $\mathrm{Re}d_{min} \subseteq \mathrm{Re}\ d$:

$$\mathrm{Re}d_{\min} = \{ R \in \mathrm{Re}d \mid \vee R' \in \mathrm{Re}d, \ |R| \leq |R'| \quad (9)$$

The intersection of all reducts is called the core, the elements of which are those attributes that cannot be eliminated. The core is defined as:

$$\mathrm{CORE}\ (C) = \cap\ \mathrm{Re}d \quad (10)$$

3 Proposed Model

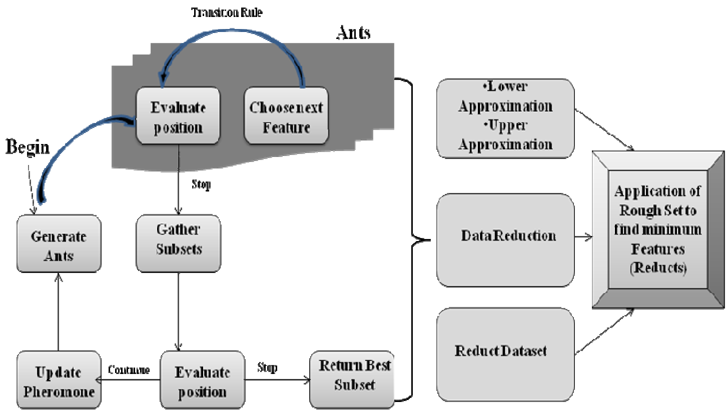
In this paper, ACO method can be combined with rough set (RST) theory to get a new algorithm. ACO is particularly attractive for feature selection as there seems to be no heuristic that can guide search to the optimal minimal subset every time. Additionally, it can be the case that ants discover the best feature combinations as they proceed throughout the search space. However Rough Set Theory is one of effective methods for dealing with incomplete information, which can reduce decision-making and classification rules so as to establish knowledge model through data analysis and knowledge reduction under the condition of maintaining the ability of classification unchangeable.

3.1 ACO Framework

An ACO algorithm can be applied to any combinatorial problem as far as it is possible to define:

1) *Appropriate problem representation*. The problem can be described as a graph with a set of nodes and edges between nodes.
2) *Heuristic desirability (*h*) of edges*. A suitable heuristic measure of the "goodness" of paths from one node to every other connected node in the graph.
3) *3) Construction of feasible solutions.* A mechanism must be in place whereby possible solutions are efficiently created. This requires the definition of a suitable traversal stopping criterion to stop path construction when a solution has been reached.
4) *4) Pheromone updating rule.* A suitable method of updating the pheromone levels on edges is required with a corresponding evaporation rule, typically involving the selection of the n best ants and updating the paths they chose.
5) *5) Probabilistic transition rule.* The rule that determines the probability of an ant traversing from one node in the graph to the next.

Figure 3 Proposed Hybridized Model using ACO and RST (Rough ACO)



Each ant in the artificial colony maintains a memory of its history - remembering the path it has chosen so far in constructing a solution. This history can be used in the evaluation of the resulting created solution and may also contribute to the decision process at each stage of solution construction.

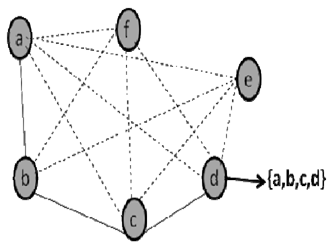
Two types of information are available to ants during their graph traversal, local and global, controlled by the parameters β and α respectively. Local information is obtained through a problem-specific heuristic measure. The extent to which the measure influences an ant's decision to traverse an edge is controlled by the parameter β. This will guide ants towards paths that are likely to result in good solutions. Global knowledge is also available to ants through the deposition of artificial pheromone on the graph edges by their predecessors over time. The impact of this knowledge on an ant's traversal decision is determined by the parameter α. Good paths discovered by past ants will have a higher amount of associated pheromone. How much pheromone is deposited, and when, is dependent on the characteristics of the problem. No other local or global knowledge is available to the ants in the standard ACO model, though the inclusion of such information by extending the ACO framework has been investigated [5].

3.2 Feature Selection

The feature selection task may be reformulated into an ACO-suitable problem [6][7]. ACO requires a problem to be represented as a graph here nodes represent features, with the edges between them denoting the choice of the next feature. The search for the optimal feature subset is then an ant traversal through the graph where a minimum number of nodes are visited that satisfies the traversal stopping criterion. Figure 4 illustrates this setup - the ant is currently at node *a* and has a choice of which feature to add next to its path (dotted lines). It chooses feature *b* next based on the transition rule, then *c* and then *d*. Upon arrival at *d*, the current subset {*a, b, c, d*} is determined to satisfy the traversal stopping criteria (e.g. suitably high classification accuracy has been achieved with this subset, assuming that the selected features are used to classify certain objects). The ant terminates its traversal and outputs this feature subset as a candidate for data reduction.

A suitable heuristic desirability of traversing between features could be any subset evaluation function - for example, an entropy-based measure or the fuzzy-rough set dependency measure. Depending on how optimality is defined for the particular application, the pheromone may be updated accordingly. For instance, subset minimality and ``goodness'' are two key factors so the pheromone update should be proportional to ``goodness'' and inversely proportional to size. How "goodness" is determined will also depend on the application. In some cases, this may be a heuristic evaluation of the subset, in others it may be based on the resulting classification accuracy of a classifier produced using the subset.

Figure 4 ACO problem representations for feature selection



The heuristic desirability and pheromone factors are combined to form the so-called probabilistic transition rule, denoting the probability of an ant *k* at feature *i* choosing to move to feature *j* at time *t*:

$$p_{i,j}^{k}(t) = \frac{[\tau_{i,j}(t)]^{\alpha}.[\eta_{i,j}]^{\beta}}{\sum_{l \in J_i^k}[\tau_{i,l}(t)]^{\alpha}.[\eta_{i,l}]^{\beta}} \tag{11}$$

Where J^k_i is the set of ant *k*'s unvisited features, $\eta_{i,j}$ is the heuristic desirability of choosing feature *j* when at feature *i* and $\tau_{i,j}(t)$ is the amount of virtual pheromone on edge (i,j). The choice of α and β is determined experimentally. Typically, several experiments are performed, varying each parameter and choosing the values that produce the best results.

3.3 Selection Process

The overall process of ACO feature selection can be seen in Figure 3. It begins by generating a number of ants, *k*, which are then placed randomly on the graph (i.e. each ant starts with one random feature). Alternatively, the number of ants to place on the graph may be set equal to the number of features within the data; each ant starts path construction at a different feature. From these initial positions, they traverse edges probabilistically until a traversal-stopping criterion is satisfied. The resulting subsets are gathered and then evaluated. If an optimal subset has been found or the algorithm has executed a certain number of times, then the process halts and outputs the best feature subset encountered. If neither condition holds, then the pheromone is updated, a new set of ants are created and the process iterates once more.

3.4 Complexity Analysis

The time complexity of the ant-based approach to feature selection is $O\ (IAk)$, where I is the number of iterations, A the number of original features, and k the number of ants. In the worst case, each ant selects all the features. As the heuristic is evaluated after each feature is added to the reduct candidate, this will result in A evaluations per ant. After one iteration in this scenario, Ak evaluations will have been performed. After I iterations, the heuristic will be evaluated IAk times.

3.5 Pheromone Update

Depending on how optimality is defined for the particular application, the pheromone may be updated accordingly. To tailor this mechanism to find rough set reducts, it is necessary to use the dependency measure as the stopping criterion. This means that an ant will stop building its feature subset when the dependency of the subset reaches the maximum for the dataset (the value 1 for consistent datasets). The dependency function may also be chosen as the heuristic desirability measure, but this is not necessary

The pheromone on each edge is updated according to the following formula:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{l=1}^{L} \Delta\tau_{ij}^{l}(t) \qquad (12)$$

Where,

$$\Delta\tau_{i,j}(t) = \sum_{k=1}^{n} \frac{\gamma_{S^k}(D)}{|S^k|} \qquad (13)$$

This is the case if the edge $(i,,j)$ has been traversed; $\Delta\ \tau_{ij}\ (t)$ is 0 otherwise. The value is decay constant used to simulate the evaporation of the pheromone, S^k is the feature subset found by ant k. The pheromone is updated according to both the rough set measure of the goodness of the ant's feature subset and the size of the subset itself.

By this definition, all ants update the pheromone. Alternative strategies may be used for this, such as allowing only the ants with the currently best feature subsets to proportionally increase the pheromone.

4 Rough ACO: Algorithmic Framework

Step1: *Take the input as a decision table S= (U, C, D)*
Step2: *Let Core=φ and Calculate the POS_c (D)*
Step3: *For $\forall$a C, calculate $POS_{(c-\{a\})}$ (D).If $POS_{(c-\{a\})}$ (D)$\neq$ POS_c(D), then CORE=CORE $\cup$ {a}; Else C=C-{a}*
Step4: *Execute iteratively step 2 until all attributes among C are calculated.*
Step5: *If POS_{core} (D) = POS_c (D), algorithm stops and return CORE as the result of feature selection; otherwise go to step 6*
Step6: *The pheromone of each arc (i,j) is assigned to an constant, i.e. $\tau_{ij}(0) = c$*
Step7: *Some ants (assumed the number of ants is m) are distributed to each core attribute node to conduct feature selection*
Step8: *Each ant selects next feature node*
Step9: *Calculate POS_{core} (D), a $\in$ C- CORE, if POS_{core} (D) = POS_c (D) algorithm stops and return FS= CORE $\cup$ a_i as the result of feature selection; else go to step 10*
Step10: *Update value of pheromone τ_{ij} for each path link and go to step 8*

5 Experimental Verification

The above algorithm was programmed in MATLAB and applied to two well known data sets [8] and three synthetic data sets. For each data set the Rough ACO MATLAB code was run for 10 times with different initial solutions. The exploring rate is 0.6 and initial pheromone is 0.001.Our algorithm could find all reducts and shortest reducts.

The general information of selected data set is shown in table 1.The performance of feature selection of Rough Set based algorithms is showed on the Table2.The experimental result shows that feature selection based on Rough ACO algorithm achieves better results than the traditional algorithms. The Rough ACO algorithm has higher speed convergence and has better capacity of optimization.

Table 1 General Information of Selected Dataset

Dataset	No. of features	Condition features	Decision features	No. of records
SPECTF Heart	23	22	1	188
Lung Cancer	57	56	1	33

Table 2 Feature Selection based on Rough ACO

Dataset	Average Iteration	Number of Reduction	Reduction Features
SPECTF Heart	>=30	8	0,1,2,3,4,5,6,7
Lung Cancer	>=80	4	0,3,6,11

6 Conclusion and Future work

Current method tends to concentrate on alternative evaluation functions; employing rough set concepts to gauge subset suitability. These methods can be categorized into two distinct approaches: those that incorporate the degree of dependency measure (or extensions), and those that apply heuristic methods to generated discernibility matrices. Methods based on traditional rough set theory do not have the ability to effectively manipulate continuous data. For these methods to operate, a discretization step must be carried out beforehand, which can often result in a loss of information.

Most of the effort in dependency degree-based feature selection has focused on the use of lower approximations and positive regions in gauging subset suitability.

However, there is still additional information that can be obtained through the use of upper approximations and the resulting boundary regions. The boundary region is of interest as this contains those objects whose concept membership is unknown objects in the positive region definitely belong to a concept, and objects in the negative region do not belong. It is thought that by incorporating this additional information into the search process, better subsets can be located.

References

[1] R. Jensen, "Performing Feature Selection with ACO. Swarm Intelligence and Data Mining," A. Abraham, C. Grosan and V. Ramos (eds.), Studies in Computational Intelligence, vol. 34, pp. 45-73. 2006.

[2] Debahuti Mishra, Dr.Amiya Kumar Rath, Dr. Milu Acharya ,” Finding Coherent Bi-clusters from Microarray Data by Imitating the Ecosystem: An Ant Colony Algorithmic Approach”, 4th Indian International Conference on Artificial Intelligence (IICAI-09,), Tumkur , Banglore

[3] Pawlak, Z. (1991) “Rough Sets: Theoretical Aspects of Reasoning about Data”.Kluwer Academic Publishing, Dordrecht

[4] Polkowski, L., Lin, T. Y., & Tsumoto, S. (Eds). (2000),”Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems”, Vol.56. Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg,

[5] Jensen, R., Shen, Q., 2003 “Finding Rough Set Reducts with Ant Colony Optimization”, Proceedings of the 2003 UK Workshop on Computational Intelligence, pp 15-22.

[6] Bonabeau, E., Dorigo, M., & Theraulez, G. (1999) “Swarm Intelligence: From Natural to Artificial Systems”, Oxford University Press Inc., New York, NY, USA.

[7] Yan Yan, YANG Huizhong.," Rough set attribute reduction algorithm based on GA". Computer Engineering and Applications. 2007, 43(31): 156-158.
[8] UCI Repository for Machine Learning Databases retrieved from the World Wide Web: http://www.ics.uci.edu
[9] Pawlak, Z.: "Rough sets", International Journal of Information and Computer Sciences, 11 (1982) 341-356.
[10] Shelokar, V.K., Jayaraman, V.K., Kulkarni, B.D., "An Ant Colony Approach for Clustering", Analytica Chimica Acta 509, 187–195, 2004.