Attribute Reduction on Continuous Data in Rough Set Theory using Ant Colony Optimization Metaheuristic

Srilatha Chebrolu
Dept. of Computer Science and Engineering,
National Institute of Technology Warangal,
India 506004.
srilatha.chebrolu@nitw.ac.in

Sriram G. Sanjeevi
Dept. of Computer Science and Engineering,
National Institute of Technology Warangal,
India 506004.
sgs@nitw.ac.in

ABSTRACT

Attribute reduction techniques based on Pawlak rough set theory work only on data sets with discrete attributes. In real-world applications, the domain of a few or all attributes of the data set may be continuous. These continuous attributes need to be discretized as a pre-processing step to attribute reduction. In this paper, we have proposed an algorithm to the problem of attribute reduction on continuous data in rough set theory. The proposed algorithm does not need any extra information or expert domain knowledge apart from the continuous data set. The proposed algorithm is based on the concepts of rough set theory. These include principle of indiscernibility, basic cuts and discernibility matrix. It adapts the search techniques provided by the ant colony optimization meta-heuristic. As ant colony optimization is a graph based meta-heuristic algorithm, we have introduced a fully connected graph whose nodes are the basic cuts. We have evaluated the proposed algorithm on various data sets found in University of California, machine learning repository. For each data set, a reduced data set is obtained by retaining the attributes in the reduct determined by the proposed algorithm and removing the attributes not in the reduct. The obtained reduced data set is found to give better classification accuracies when tested using i) C4.5 classifier and ii) Naive Bayes classifier in comparison with those obtained on the data set before attribute reduction.

CCS Concepts

•**Computing methodologies** → **Feature selection;** *Optimization algorithms; Representation of Boolean functions;* Vagueness and fuzzy logic;

Keywords

Discretization; Attribute reduction; Rough set theory; Indiscernibility; Boolean reasoning; Ant colony optimization

1. INTRODUCTION

Rough set theory (RST) was introduced by Pawlak in 1982 for data analysis [23, 24, 25]. RST provides a mathematical methodology to deal with imprecision and uncertainty in information systems. Based on the principle of indiscernibility of objects and approximation of concepts, RST identifies data dependencies of decision attributes on the conditional attributes in decision systems. Semantics preserving attribute reduction and decision rule acquisition are the important applications of RST. As RST does not require any additional information apart from the data set, it is successfully applied to many domains [13, 29]. RST is used to solve the problem of attribute reduction and is widely studied in the literature [1, 4, 5, 8, 16, 18, 35, 38].

Pawlak RST is designed to work on data sets consisting of only discrete-valued attributes. But the domain of a few or all attributes of the data set may be continuous-valued. These continuous-valued attributes need to be discretized as a pre-processing step to attribute reduction. Various algorithms are proposed in the literature based on RST to discretize the continuous-valued attributes [28, 32, 36]. The main limitation of discretizing the data set as a pre-processing step to attribute reduction is loss of information. Fuzzy rough set theory provides a framework to perform attribute reduction on continuous-valued attributes by taking the concept of fuzziness from fuzzy set theory and the concept of indiscernibility from RST. Various approaches have been discussed in the literature based on fuzzy rough sets, for attribute reduction on decision tables consisting of continuous-valued attributes [6, 10, 11, 12, 13, 14, 15, 21, 31, 33, 34, 37]. The limitation with these approaches is the requirement of computation of fuzzy partitions using fuzzy similarity relations.

In this paper, we have proposed an algorithm to the problem of attribute reduction in RST, on data sets consisting of continuous-valued attributes. Initially, we construct discernibility matrix and basic cut graph based on RST principles of indiscernibility. We adapt Ant colony optimization (ACO) meta-heuristic by providing the stopping criteria for an ant traversal and the evaluation criteria for the traversed path of an ant. The proposed algorithm returns the near optimal set of irreducible cuts. Next, we extract attributes from the obtained set of irreducible cuts.

The rest of the paper is organized as follows. Section 2 introduces RST concepts. Section 3 introduces the concepts of discretization of data set consisting of continuous-valued attributes using RST. Section 4 describes our proposed algorithm for attribute reduction on continuous data. Section

WCI '15, August 10 - 13, 2015, Kochi, India

DOI: http://dx.doi.org/10.1145/2791405.2791438

5 describes experimental results obtained and comparisons performed. Section 6 describes the conclusion and future work of the paper.

2. ROUGH SET THEORY

In RST [22], a decision system S is defined as $S = (U, C, d, f)$, where $U = \{o_1, o_2, o_3, ..., o_n\}$ is the universe of n objects, $C = \{c_1, c_2, c_3, ..., c_m\}$ is the set of m conditional attributes, d is the decision attribute, f is a mapping function, $f(o_i, a) \in V_a$ where $a \in (C \cup d)$ and V_a is the value set of attribute a. The principal idea of RST is the indiscernibility relation I_A. In a decision system S, if objects o_i and o_j have same values with respect to the attributes in $A \subseteq C$, then the object o_i is said to be indiscernible from the object o_j with respect to A. The indiscernibility relation I_A is defined as follows

$$I_A = \{(o_i, o_j) \in U \times U \mid \forall a \in A,\ f(o_i, a) = f(o_j, a)\}$$

I_A is said to be an equivalence relation, because it satisfies reflexive, symmetric and transitive properties. In a decision system S, an equivalence class $[o_i]_A$ of an object o_i is defined as the set of all objects in U that are indiscernible from o_i with respect to the attributes in A.

$$[o_i]_A = \{o_j \in U \mid o_i I_A o_j\}$$

Let $V_d = \{d_1, d_2, d_3, ..., d_k\}$ be the set of all decision classes of S. In a decision system S, a concept X_{d_i} is defined as the set of all objects in U that are having d_i as their decision attribute value.

$$X_{d_i} = \{o_j \in U \mid f(o_j, d) = d_i, d_i \in V_d\}$$

In a decision system S, the set of all concepts D partitions U into k disjoint subsets, where $k = |V_d|$.

$$D = \{X_{d_1}, X_{d_2}, X_{d_3}, ..., X_{d_k}\}, \quad U = \bigcup_{i=1}^{k} X_{d_i}$$

Rough set of a concept X_{d_i} with respect to $A \subseteq C$ is defined in terms of lower approximation and upper approximation. Lower approximation $LA_A(X_{d_i})$ of the concept X_{d_i} with respect to $A \subseteq C$ is defined as the set of all objects that certainly belong to the concept X_{d_i}.

$$LA_A(X_{d_i}) = \bigcup_{o_j \in U} \{[o_j]_A \mid [o_j]_A \subseteq X_{d_i}\}$$

Upper approximation $UA_A(X_{d_i})$ of the concept X_{d_i} with respect to $A \subseteq C$ is defined as the set of all objects that possibly belong to the concept X_{d_i}.

$$UA_A(X_{d_i}) = \bigcup_{o_j \in U} \{[o_j]_A \mid [o_j]_A \cap X_{d_i} \neq \phi\}$$

Positive region $POS_A(X_{d_i})$ of the concept X_{d_i} with respect to $A \subseteq C$ is defined same as the lower approximation $LA_A(X_{d_i})$.

$$POS_A(X_{d_i}) = LA_A(X_{d_i})$$

Boundary region $BND_A(X_{d_i})$ of the concept X_{d_i} with respect to $A \subseteq C$ is defined as the difference of upper approximation $UA_A(X_{d_i})$ and the lower approximation $LA_A(X_{d_i})$.

$$BND_A(X_{d_i}) = UA_A(X_{d_i}) - LA_A(X_{d_i})$$

Further,

$$POS_A(D) = \bigcup_{X_{d_i} \in D} POS_A(X_{d_i})$$

$$BND_A(D) = \bigcup_{X_{d_i} \in D} BND_A(X_{d_i})$$

Consistency of a decision table is defined in terms of a generalized decision function. A generalized decision function $\partial_A(o_i)$ of an object o_i with respect to $A \subseteq C$, is defined as the set of decision classes of all the objects in the equivalence class of o_i.

$$\partial_A(o_i) = \{f(o_j, d) \mid o_j \in [o_i]_A\}$$

Decision table is said to be consistent, if the cardinality of $\partial_A(o_i)$ is 1 for all the objects in the universe. Otherwise, the decision table is said to be inconsistent.

$$S = \begin{cases} consistent & \text{if } \forall o_i \in U\ |\partial_C(o_i)| = 1 \\ inconsistent & \text{otherwise} \end{cases}$$

where $|\partial_C(o_i)|$ is the cardinality of $\partial_C(o_i)$.

3. DISCRETIZATION USING RST

Discretization of continuous data using RST is based on defining a set of cuts on all the continuous-valued attributes of the decision system [23]. Let V_c be the value set of attribute $c \in C$. Let l_c be the left bound and r_c be the right bound of V_c such that $l_c < r_c$. Let $\Re$ be the set of real numbers. $V_c = [l_c, r_c) \subset \Re$. Let p_i be a real number such that $l_c \leq p_i < r_c$. p_i partitions the universe of objects U into two disjoint sets U_l and U_r where $U_l = \{o_j \in U \mid f(o_j, c) \leq p_i\}$ and $U_r = \{o_j \in U \mid f(o_j, c) > p_i\}$ respectively. Both U_l and U_r are non-empty sets. p_i is defined as the cut on attribute c. Let P_c be the set of cuts on attribute c defined as $P_c = \{p_1, p_2, ..., p_k\}$ such that $l_c \leq p_1 < p_2 < ... < p_k < r_c$.

Let the value set of the attribute c be defined as $V_c = \{v_1, v_2, v_3, ..., v_{|V_c|}\}$ such that $v_1 < v_2 < ... < v_{|V_c|}$. The set of basic cuts B_c on attribute c is defined as follows

$$B_c = \left\{ \frac{(v_1 + v_2)}{2}, \frac{(v_2 + v_3)}{2}, ..., \frac{(v_{|V_c|-1} + v_{|V_c|})}{2} \right\}$$

The set of basic cuts B on the set of attributes C is defined as $B = \bigcup_{c \in C} B_c$.

The discretized version of a consistent decision system S is a new decision system called $P-$*discretization* of S and is defined as a 5 tuple $S^P = (U, C, d, P, f^P)$, where P is the set of cuts on C, defined as $P = \bigcup_{c \in C} P_c$ and the function f^P is defined as follows

$$f^P(o_j, c) = \begin{cases} 0, & \text{if } f(o_j, c) < p_1 \\ i, & \text{if } f(o_j, c) \in [p_i, p_{i+1}), 1 \leq i \leq k-1) \\ k, & \text{if } f(o_j, c) \geq p_k \end{cases}$$

If the generalized definition functions $\partial_C(S)$ and $\partial_C(S^P)$ of the decision systems S and S^P respectively are same, then the set of cuts P is called *S-consistent*. If P is *S-consistent* and $\forall P' \subset P \mid P'$ is not *S-consistent*, then the set of cuts P is called *S-irreducible*. If P is *S-consistent* and $\forall$ *S-consistent* set of cuts P' such that $|P| \leq |P'|$, then the set of cuts P is called *S-optimal* with respect to the cardinality of the set of cuts.

The problem of finding the set of *S-optimal* cuts is an optimization problem. Boolean reasoning approach is used to the problem of finding the set of *S-optimal* cuts in RST by defining discernibility matrix and a discernibility function [30]. Skowron et al. [30] have introduced discernibility matrix and discernibility function. The discernibility matrix M is defined as follows

$$M = \begin{cases} e_{ij} & \text{if } (1 \leq i < j \leq n) \ \wedge \ (f(o_i, d) \neq f(o_j, d)) \\ \phi & \text{otherwise} \end{cases}$$

where $e_{ij} = \bigcup_{c \in C} \bigcup_{p_k \in B_c} \{(c, p_k) \mid p_k = \frac{(v_k + v_{k+1})}{2} \wedge$

$[v_k, v_{k+1}) \subseteq [\min(f(o_i, c), f(o_j, c)), \max(f(o_i, c), f(o_j, c)))\}$

The discernibility function F is defined as follows

$$F = \bigwedge_{i \geq 1} \bigvee_{i < j \leq n} e^*_{ij}$$

where $e^*_{ij} = \bigcup_{(c,p_k) \in e_{ij}} \{p^*_k \mid p^*_k$ is the boolean variable corresponding to $(c, p_k)\}$. Discernibility function is the conjunct of the disjuncts that are defined on each element of the discernibility matrix by considering the cuts as boolean variables. Each implicant of F is defined as the set of *S-consistent* cuts of the decision system S. Each prime implicant of F is defined as the set of $S-$*irreducible* cuts of the decision system S. The set of $S-$*irreducible* cuts with minimum number of cuts is defined as the set of $S-$*optimal* cuts of the decision system S.

4. PROPOSED ALGORITHM TO THE PROBLEM OF ATTRIBUTE REDUCTION ON CONTINUOUS DATA IN RST USING ACO

ACO is a meta-heuristic introduced by Dorigo et al. in 1996 [7]. ACO meta-heuristic is applied to a wide range of combinatorial problems [2, 3, 20, 27]. ACO is inspired from the collaborative behavior of real ants in finding the shortest path to food sources from their colonies. Real ants find the shortest path by following the principle of stigmergy. Initially, all ants move randomly from their colonies in search of food. Once, an ant finds food source, it starts its return journey to the colonies. These ants deposit pheromone trails in their return journey. Ants that are present with in the influence of the deposited pheromone follow the pheromone trails to food sources. These ants also deposit pheromone in their return journey. This results in an increase in the concentration of the pheromone in this path. Eventually, all ants follow this path. Pheromone also gets slowly evaporated which helps ants not to get stuck in the sub-optimal paths.

The data flow diagram of the proposed algorithm is as shown in Figure 1. The proposed algorithm gets input as the decision table consisting of continuous-valued attributes. In this work, we construct basic cut graph G by applying RST principles of indiscernibility. Each vertex of the basic cut graph corresponds to a unique basic cut in the set of all basic cuts B. So, there will be a one to one mapping between the vertices of the basic cut graph and the set of basic cuts B. Finally, the number of vertices in G is $|B|$, where $|B|$ is the cardinality of B. We also construct a discernibility matrix M whose elements are the set of basic cuts that are discerning a row object and a column object. With inputs as the basic cut graph G, discernibility matrix M and ACO parameters, the proposed algorithm determines an approximately optimal set of irreducible cuts. Here, optimality of the set of irreducible cuts is defined in terms of the cardinality of the set of irreducible cuts. The set of irreducible cuts of minimal length, is said to be optimal. Reduct is obtained from the determined set of irreducible cuts by extracting the attributes corresponding to the cuts in the obtained set of irreducible cuts.

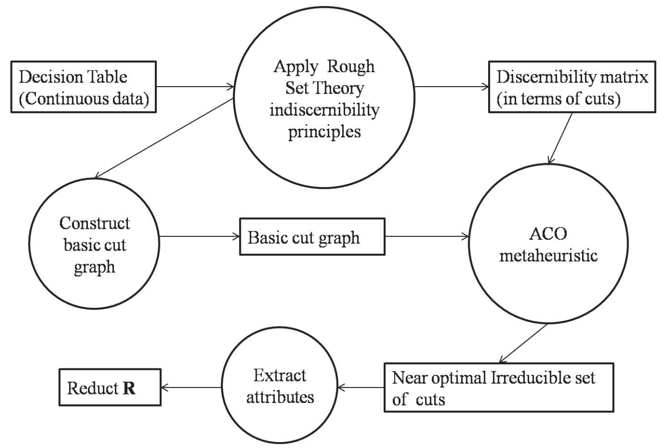

Figure 1: Data flow diagram of the proposed algorithm for attribute reduction on continuous data in RST using ACO meta-heuristic

4.1 Representation of problem

In ACO meta-heuristic, the combinatorial problem is described as a graph $G(V, E)$, where V is the set of vertices and E is the set of edges. The solution to the combinatorial problem is to find the shortest path from source node to the destination node. In this paper, the problem of attribute reduction on continuous data sets is represented as a problem of finding the shortest path of the basic cut graph, where each node of the basic cut graph corresponds to the basic cut $p_k \in B$. All nodes of the graph are fully connected.

4.2 Stopping criteria and evaluation criteria

The solution is to find the shortest path $P = \{p_1, p_2, ..., p_k\}$ of the basic cut graph satisfying the following conditions

1. Stopping criteria: $\forall e \in M, \exists p_k \in P \mid p_k \in e$
2. Evaluation criteria: $\forall p_k \in P, \exists e \in M \mid (p_k \in e) \ \& \ ((P - \{p_k\}) \cap e = \phi)$

The first condition is defined as the stopping criteria, because an ant has to stop its traversal through the graph G if this condition is met. The second condition is defined as the evaluation criteria, because once an ant stops its traversal, then the path traversed by the ant is evaluated using this condition.

4.3 Probability of selection of a node

Probability of selecting the node p_j when the ant is at node p_i is given as

$$pr(p_i, p_j) = \begin{cases} \frac{\tau_{p_i p_j}}{\sum_{p_k} \tau_{p_i p_k}} & \text{if } p_k \in N_{p_i} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where N_{p_i} is the set of nodes that are not yet traversed by the ant.

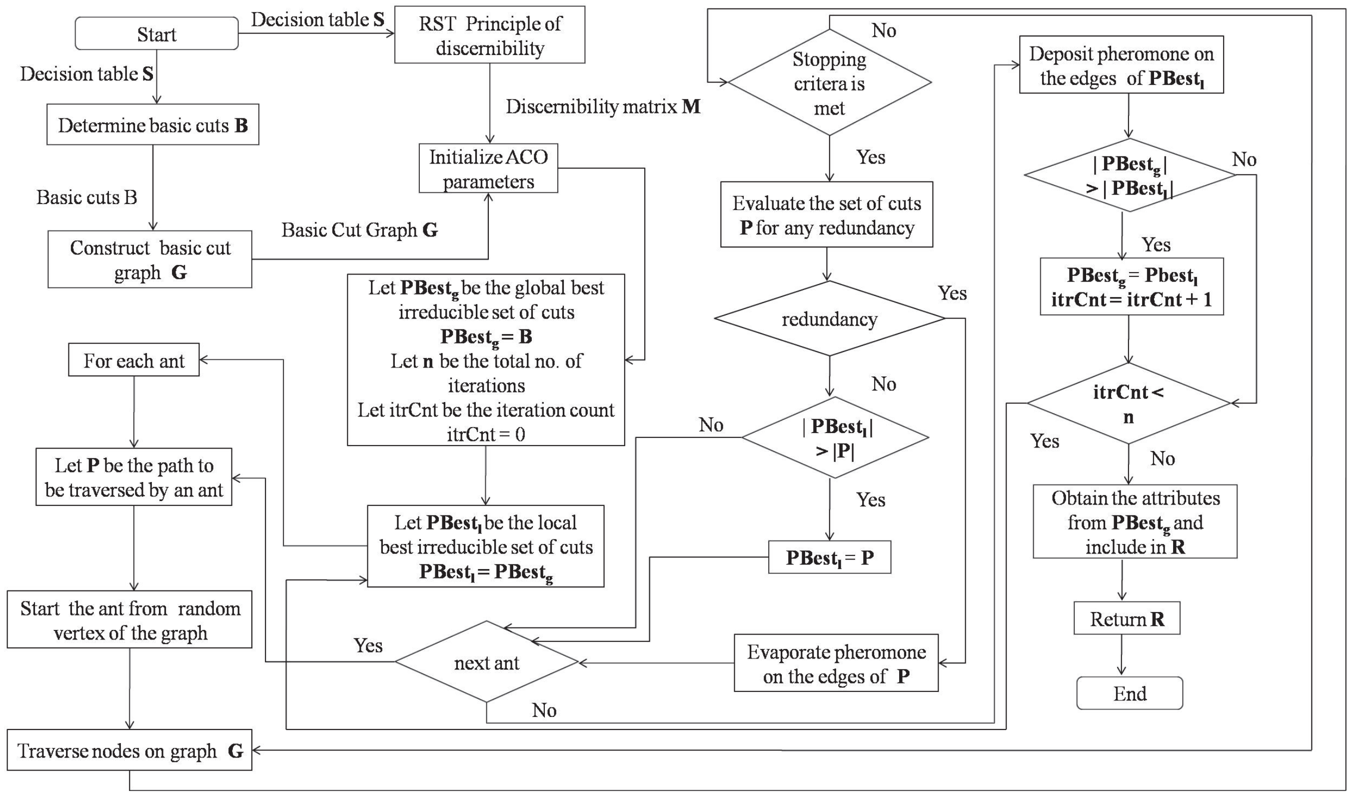

Figure 2: Flow chart of the proposed algorithm for attribute reduction on continuous data in RST using ACO meta-heuristic

4.4 Deposit and evaporation of pheromone

An ant traversal through the graph is stopped if the stopping criteria defined in Section 4.2 is satisfied. The solution obtained by an ant is the set of all nodes traversed by the ant. Let P be the set of all nodes traversed by the ant. The obtained solution is then evaluated based on the evaluation criteria defined in Section 4.2. The obtained solution is discarded if it fails to satisfy the evaluation criteria. This path is retraced backward and the pheromone trails of all the edges in this path are evaporated as follows

$$\tau_{p_i p_j} = (1-\rho)\tau_{p_i p_j} \tag{2}$$

where p_i and p_j are two consecutive nodes in the path P and $\rho \in (0, 1]$ is a parameter. Pheromone trail evaporation decreases the probability of selecting the nodes present in this path in future traversals of ants. If the obtained solution satisfies the evaluation criteria, then the ant deposits an amount of $\Delta\tau$ pheromone trails on all the edges in this path as follows

$$\tau_{p_i p_j} = \tau_{p_i p_j} + \Delta\tau \tag{3}$$

Pheromone trail deposits increase the probability of selecting the nodes present in this path in future traversals of ant through the graph G.

4.5 The proposed algorithm

The overview of the proposed algorithm is as shown in Figure 2. Decision table consisting of continuous-valued attributes is given as input to this algorithm. The output of this algorithm is the reduct R. The detailed description of the proposed algorithm is as shown in Algorithm 1. Initially, the algorithm determines the set of basic cuts B and constructs the discernibility matrix M. Fully connected graph G is constructed with basic cuts of B as vertices. Next, the proposed algorithm adapts ACO meta-heuristic to determine the approximately optimal reduct. The proposed algorithm starts with initializing all the ACO parameters. Each run of the algorithm is limited to the maximum number of iterations. Let $PBest_g$ be the global best set of irreducible cuts to be determined by the proposed algorithm. Let $PBest_l$ be the local best set of irreducible cuts to be determined by the ants in the current iteration.

In each iteration, the algorithm creates n ants and does the following for each ant i) let P be the path to be traversed by an ant ii) an ant starts from a random vertex and depending upon the probability of traversal, it selects the next vertex to traverse iii) an ant stops its traversal through graph G if the stopping criteria $(\forall e \in M, \exists p_k \in P \mid p_k \in e)$ is met iv) evaluates the solution P obtained by an ant using the evaluation criteria $(\forall p_k \in P, \exists e \in M \mid (p_k \in e)\ \&\ ((P - \{p_k\}) \cap e = \phi))$ v) if the solution P fails to satisfy the evaluation criteria, then the pheromone trails of the path P are evaporated as given in Equation 2 vi) otherwise, pheromone trails are deposited on all the edges of P as given in Equation 3 and if the cardinality of $PBest_l$ is greater than the cardinality of P, then the set of irreducible cuts $PBest_l$ is updated with the obtained solution P.

The algorithm deposits an amount of $\Delta\tau$ pheromone trails

Algorithm 1: Proposed algorithm to determine reduct on continuous data in RST using ACO meta-heuristic

```
Input  : Decision table S(U, C, d, f)
Output: Reduct R
Let B be the set of basic cuts on the set of attributes C
// Determine the set of basic cuts B
for (each c ∈ C) do
    Let V_c be the value set of attribute c
    Sort V_c as V_c = {v_1, v_2, v_3, ..., v_|V_c|} such that v_1 < v_2 < ... < v_|V_c|
    Let B_c be the set of cuts on attribute c ∈ C
    Determine B_c as B_c = { (v_1+v_2)/2, (v_2+v_3)/2, ..., (v_|V_c|-1 + v_|V_c|)/2 }
    B = B ⋃ B_c
Let PBest_g be the global best set of irreducible cuts
Initialize PBest_g with the complete set of basic cuts, PBest_g = B
Let G be the graph with basic cuts in B as vertices
Construct a fully connected graph G
Let M be the discernibility matrix
// Construct discernibility matrix M
for (each o_i in U) do
    for (each o_j in U) do
        Let e be an element of the matrix M
        if (f(o_i, d) ≠ f(o_j, d)) then
            e = ⋃_{c∈C} ⋃_{p_k∈B_c} {(c, p_k) | p_k = (v_k + v_{k+1})/2 ∧ [v_k, v_{k+1}) ⊆ [min(f(o_i, c), f(o_j, c)), max(f(o_i, c), f(o_j, c)))}
// ACO meta-heuristic parameters
Let t be the termination condition
Let n be the number of ants
Let τ_{p_i p_j} be the pheromone trail on edge connecting any two vertices p_i and p_j of the graph G, initially 0
Let Δτ be the amount of pheromone trail to be deposited on an edge of G
Let ρ ∈ (0, 1] be a parameter
Initialize all ACO meta-heuristic parameters
while (!t) do
    Let PBest_l be the local best set of irreducible cuts, initialize with PBest_g
    Create n ants
    for (each ant) do
        Let P be the set of vertices to be traversed by the ant, initially empty
        Start the ant from a random vertex p_i of the graph G, P = {p_i}
        // Stopping criteria for ant traversal
        while (!(∀e ∈ M, ∃p_k ∈ P | p_k ∈ e)) do
            Select the next vertex p_j to be traversed with probability pr(p_i, p_j)

                pr(p_i, p_j) = { τ_{p_i p_j} / Σ_{p_k} τ_{p_i p_k}    if p_k ∈ N_{p_i}
                               { 0                                   otherwise

            where N_{p_i} is the set of vertices that are not yet traversed by the ant
            P = P ⋃ {p_j}
        // Evaluate P
        if (∀p_k ∈ P, ∃e ∈ M | (p_k ∈ e) & ((P − {p_k}) ∩ e = φ)) then
            // Set of irreducible cuts PBest_l is updated if the newly obtained set of irreducible cuts is
               of minimal length
            if (|P| < |PBest_l|) then
                PBest_l = P
        else
            // Evaporate pheromone trails on all the edges in the path P
            for (each edge connecting two vertices p_i and p_j in the path P) do
                τ_{p_i p_j} = (1 − ρ) τ_{p_i p_j}
    // Deposit pheromone trails on all the edges in the path PBest_l
    for (each edge connecting two vertices p_i and p_j in the path PBest_l) do
        τ_{p_i p_j} = τ_{p_i p_j} + Δτ
    if (|PBest_l| < |PBest_g|) then
        PBest_g = PBest_l
// Extract conditional attributes from the set of irreducible cuts PBest_g
for (each p_i ∈ PBest_g) do
    Let c_j be the attribute corresponding to the cut p_i
    Include c_j in R, R = R ⋃ {c_j}
return R
```

Table 1: ACO parameters

Parameter	Value
Termination condition (t)	No. of iterations = 1000
No. of ants (n)	100
Initial pheromone trail ($\tau_{p_i p_j}$)	0
Pheromone evaporation factor (ρ)	0.01

on all the edges of the path $PBest_l$. If the cardinality of $PBest_g$ is greater than the cardinality of $PBest_l$, then the set of irreducible cuts $PBest_g$ is updated with the obtained solution $PBest_l$. Once, the algorithm completes the maximum number of iterations, it returns the set of unique attributes corresponding to the cuts in $PBest_g$ as the final reduct R.

5. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we analyze the effectiveness of our proposed algorithm for attribute reduction on continuous data in RST using ACO meta-heuristic by conducting a series of experiments on 7 data sets of different sizes from machine learning repository at University of California, Irvine [19]. The experiments have been conducted with ACO parameters as shown in Table 1. While conducting experiments, each execution of the proposed algorithm is limited to 1000 iterations. The details of all the 7 data sets selected for our experimentation are shown in Table 2. All the chosen data sets are of multi-category, with the number of classes ranging from 2 to 8, the number of objects ranging from 150 to 5473 and the number of conditional attributes ranging from 3 to 13. Except haberman data set, for all the remaining 6 data sets all of the conditional attributes are continuous-valued. The haberman data set has only one discrete-valued conditional attribute and rest all are continuous-valued conditional attributes. The experiments were conducted on a 2.27 GHz PC running Windows 7 with Intel Core i3 processor with 4 GB RAM.

We conducted a series of experiments using the proposed algorithm to determine a reduct on all the 7 data sets shown in Table 2. We have also performed comparison analysis with respect to the classification performance for each of the data set before and after attribute reduction. Criteria for comparison chosen are i) classification accuracies obtained, when tested using the C4.5 classifier [26] and ii) classification accuracies obtained, when tested using the Naive Bayes classifier [17]. C4.5 classification algorithm was introduced by J. Ross Quinlan. C4.5 algorithm constructs a decision tree from training data based on the concept of information entropy. Classification of an object with unknown decision is done using the constructed decision trees. Naive Bayes classifier is based on Bayes theorem. Classification of an object with unknown decision is done using the conditional probability of the object to belong to a class. Conditional probability is calculated by assuming zero dependency between the conditional attributes.

Table 3 shows the set of irreducible cuts and reduct obtained using the proposed algorithm. The set of irreducible cuts obtained is an approximate to optimal solution. Here, optimality is defined in terms of the cardinality of the set

Table 2: Data sets description

Data set	No. of objects	No. of conditional attributes	No. of classes
breast-cancer	699	9	2
ecoli	336	7	8
haberman	306	3	2
iris	150	4	3
liver-disorders	345	6	2
page-blocks	5473	10	5
wine	178	13	3

of irreducible cuts. The reduct is obtained from the set of irreducible cuts by selecting the attributes corresponding to the cuts without including duplicates.

Table 4 shows the cardinality of the set of irreducible cuts, cardinality of the reduct and percentage of reduction obtained using the proposed algorithm. On all the data sets the proposed algorithm has achieved more than 10 percentage of reduction. The wine data set has achieved a highest percentage of reduction of 69.23. Reduced data sets are obtained by retaining the attributes of reduct in the original data sets and removing other attributes not in the reduct. The performance of the proposed algorithm is evaluated by training C4.5 classifier and Naive Bayes classifier on the original data sets and the reduced data sets. We have used the implementation of C4.5 classifier and Naive Bayes classifier provided by WEKA software tool [9]. Classification accuracy was obtained by using C4.5 classifier and Naive Bayes classifier with 10 fold cross-validation approach for validation.

Now, we present the comparison analysis of the performance of our proposed algorithm. Table 5 shows the comparisons on the classification performance of the data sets before and after attribute reduction. For the different data sets, columns 2 and 3 of Table 5 shows the comparisons of classification accuracies obtained by training C4.5 classifier on the original data set and reduced data set obtained using the proposed algorithm. These results show that the proposed algorithm is achieving a reduced data set that is obtaining higher classification accuracies on majority i.e., 4 of the 7 data sets. The original data set is achieving higher classification accuracies on 3 of the 7 data sets. For the different data sets, columns 4 and 5 of Table 5 shows the comparisons of classification accuracies obtained by training Naive Bayes classifier on the original data set and reduced data set obtained using the proposed algorithm. These results show that the proposed algorithm is achieving a reduced data set that is obtaining higher classification accuracies on majority i.e., 4 of the 7 data sets. The original data set is achieving higher classification accuracies on 3 of the 7 data sets. The bold values in this table indicate the higher classification accuracies.

From the experimental results obtained, we conclude that the proposed algorithm is giving a reduced data set in terms of i) cardinality of the conditional attributes found for different data sets ii) classification accuracies obtained for different data sets, when tested using the C4.5 classifier and iii) classification accuracies obtained for different data sets, when tested using the Naive Bayes classifier.

Table 3: Set of irreducible cuts and reduct obtained using the proposed algorithm

Data set	Set of irreducible cuts	Reduct
breast-cancer	$\{(c_1, 6.5), (c_2, 3.5), (c_6, 3.77), (c_8, 2.5)\}$	$\{c_1, c_2, c_6, c_8\}$
ecoli	$\{(c_1, 0.56), (c_2, 0.5), (c_5, 0.52), (c_6, 0.52)\}$	$\{c_1, c_2, c_5, c_6\}$
haberman	$\{(c_1, 30.5), (c_1, 80.5), (c_3, 49)\}$	$\{c_1, c_3\}$
iris	$\{(c_2, 2.95), (c_3, 4.45), (c_4, 0.8), (c_4, 1.75)\}$	$\{c_2, c_3, c_4\}$
liver-disorders	$\{(c_1, 90.5), (c_2, 65.5), (c_5, 22.5), (c_6, 3.5)\}$	$\{c_1, c_2, c_5, c_6\}$
page-blocks	$\{(c_1, 7.5), (c_1, 14.5), (c_1, 28.5), (c_1, 38.5), (c_2, 40.5), (c_4, 1.32), (c_4, 1.9), (c_4, 2.87),$ $(c_4, 17.9), (c_4, 34.69), (c_5, 0.21), (c_5, 0.34), (c_5, 0.44), (c_5, 0.6), (c_5, 0.9), (c_6, 0.52),$ $(c_6, 0.74), (c_7, 1.83), (c_7, 3.5), (c_7, 5.4), (c_7, 7.28), (c_8, 9.5), (c_8, 15.5), (c_9, 98.5), (c_{10}, 5.5)\}$	$\{c_1, c_2, c_4, c_5, c_6,$ $c_7, c_8, c_9, c_{10}\}$
wine	$\{(c_1, 13.04), (c_7, 1.4), (c_{10}, 3.82), (c_{13}, 719)\}$	$\{c_1, c_7, c_{10}, c_{13}\}$

Table 4: Cardinality of the set of irreducible cuts, cardinality of the reduct and percentage of reduction obtained using the proposed algorithm

Data set	Cardinality of the set of irreducible cuts	Cardinality of the reduct	Percentage of reduction
breast-cancer	4	4	55.56
ecoli	4	4	42.86
haberman	3	2	33.33
iris	4	3	25
liver-disorders	4	4	33.33
page-blocks	25	9	10
wine	4	4	69.23

6. CONCLUSION AND FUTURE WORK

The contribution of this paper is the proposed algorithm to determine reduct on continuous data; determines a set of irreducible cuts with no redundancy. We have analyzed the performance of our proposed algorithm by conducting a series of experiments on 7 data sets from machine learning repository at University of California, Irvine. Initially, we conducted a series of experiments using the proposed algorithm to determine reducts for each of the 7 data sets. Testing for classification accuracy was done on all 7 reduced data sets using C4.5 classifier and Naive Bayes classifier respectively. We have also performed comparison analysis with respect to the classification performance for each of the data set before and after attribute reduction. We conclude that our algorithm is giving better results compared to the results that are obtained before attribute reduction as shown by the experimental results obtained for classification accuracies for different data sets, when tested using i) C4.5 classifier and ii) Naive Bayes classifier. The problem of attribute reduction in RST on imbalanced data is to be studied in the future as most learning algorithms expect equal distribution of classes.

7. ACKNOWLEDGMENTS

This research work was supported by TATA Consultancy Services under the TCS Research Scholar Program Scheme.

8. REFERENCES

[1] R. B. Bhatt and M. Gopal. On fuzzy-rough sets approach to feature selection. *Pattern Recognition Letters*, 26(7):965 – 975, 2005.

[2] C. Blum. Beam-ACO-hybridizing ant colony optimization with beam search: an application to open shop scheduling. *Computers & Operations Research*, 32(6):1565 – 1591, 2005.

[3] C. Blum and M. Sampels. An ant colony optimization algorithm for shop scheduling problems. *Journal of Mathematical Modelling and Algorithms*, 3(3):285–308, 2004.

[4] C. Cornelis, R. Jensen, G. Hurtado, and D. Ślęzak. Attribute selection with fuzzy decision reducts. *Information Sciences*, 180(2):209 – 224, 2010.

[5] J. Dai and Q. Xu. Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification. *Applied Soft Computing*, 13(1):211 – 221, 2013.

[6] C. Degang and Z. Suyun. Local reduction of decision system with fuzzy rough sets. *Fuzzy Sets and Systems*, 161(13):1871 – 1883, 2010.

[7] M. Dorigo, V. Maniezzo, and A. Colorni. Ant system: Optimization by a colony of cooperating agents. *Trans. Sys. Man Cyber. Part B*, 26(1):29–41, 1996.

[8] Y.-Y. Guan, H.-K. Wang, Y. Wang, and F. Yang. Attribute reduction and optimal decision rules acquisition for continuous valued information systems. *Information Sciences*, 179(17):2974 – 2984, 2009.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: An update. *SIGKDD Explorations*, 11:10–18, 2009.

[10] Q. Hu, D. Zhang, S. An, D. Zhang, and D. Yu. On robust fuzzy rough set models. *IEEE Transactions on Fuzzy Systems*, 20(4):636–651, Aug 2012.

[11] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, and D. Yu. Gaussian kernel based fuzzy rough sets: Model, uncertainty measures and applications. *International Journal of Approximate Reasoning*, 51(4):453 – 471, 2010.

[12] R. Jensen. Combining rough and fuzzy sets for feature selection, 2005.

[13] R. Jensen and Q. Shen. Fuzzy-rough attribute reduction with application to web categorization. *Fuzzy Sets and Systems*, 141(3):469 – 485, 2004.

[14] R. Jensen and Q. Shen. Fuzzy-rough sets assisted attribute selection. *IEEE Transactions on Fuzzy Systems*, 15(1):73–89, Feb 2007.

[15] R. Jensen and Q. Shen. New approaches to fuzzy-rough feature selection. *IEEE Transactions on Fuzzy Systems*, 17(4):824–838, Aug 2009.

[16] X. Jia, W. Liao, Z. Tang, and L. Shang. Minimum cost attribute reduction in decision-theoretic rough set models. *Information Sciences*, 219(0):151 – 167, 2013.

[17] G. H. John and P. Langley. Estimating continuous

Table 5: Comparisons of classification accuracy before and after reduction obtained by training C4.5 classifier and Naive Bayes classifier

Data set	C4.5		Naive Bayes	
	Before Reduction	After Reduction	Before Reduction	After Reduction
breast-cancer	94.89	**95.35**	**96.12**	95.67
ecoli	**82.83**	78.31	**85.50**	77.91
haberman	72.16	**73.53**	**75.36**	73.53
iris	94.73	**96.00**	95.53	**96.00**
liver-disorders	**65.84**	64.06	54.89	**66.27**
page-blocks	**96.99**	96.90	90.01	**92.00**
wine	93.2	**97.42**	97.46	**97.76**

distributions in bayesian classifiers. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, 1995.

[18] M. Li, C. Shang, S. Feng, and J. Fan. Quick attribute reduction in inconsistent decision tables. *Information Sciences*, 254(0):155 – 180, 2014.

[19] M. Lichman. UCI machine learning repository [http://archive.ics.uci.edu/ml]. University of California, Irvine, School of Information and Computer Sciences, 2013.

[20] D. Merkle, M. Middendorf, and H. Schmeck. Ant colony optimization for resource-constrained project scheduling. *IEEE Transactions on Evolutionary Computation*, 6(4):333–346, 2002.

[21] N. M. Parthaláin and R. Jensen. Unsupervised fuzzy-rough set-based dimensionality reduction. *Information Sciences*, 229(0):106 – 121, 2013.

[22] Z. Pawlak. Rough sets. *International Journal of Computer and Information Sciences*, 11:341–356, 1982.

[23] Z. Pawlak and A. Skowron. Rough sets and boolean reasoning. *Information Sciences*, 177(1):41 – 73, 2007.

[24] Z. Pawlak and A. Skowron. Rough sets: Some extensions. *Information Sciences*, 177(1):28 – 40, 2007.

[25] Z. Pawlak and A. Skowron. Rudiments of rough sets. *Information Sciences*, 177(1):3 – 27, 2007.

[26] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[27] M. Reimann, K. Doerner, and R. F. Hartl. D-ants: Savings based ants divide and conquer the vehicle routing problem. *Computers & Operations Research*, 31(4):563 – 591, 2004.

[28] A. Roy and S. K. Pal. Fuzzy discretization of feature space for a rough set classifier. *Pattern Recognition Letters*, 24(6):895 – 902, 2003.

[29] Q. Shen and R. Jensen. Selecting informative features with fuzzy-rough sets and its application for complex systems monitoring. *Pattern Recognition*, 37(7):1351 – 1363, 2004.

[30] A. Skowron and C. Rauszer. The discernibility matrices and functions in information systems. In R. Słowiński, editor, *Intelligent Decision Support: Handbook of Applications and Advance of the Rough Sets Theory*, volume 11 of *Theory and Decision Library*, pages 331–362. Springer Netherlands, 1992.

[31] B. Sun, W. Ma, and D. Chen. Rough approximation of a fuzzy concept on a hybrid attribute information system and its uncertainty measure. *Information Sciences*, 284(0):60 – 80, 2014.

[32] D. Tian, X. jun Zeng, and J. Keane. Core-generating approximate minimum entropy discretization for rough set feature selection in pattern classification. *International Journal of Approximate Reasoning*, 52(6):863 – 880, 2011.

[33] E. Tsang, D. Chen, D. Yeung, X.-Z. Wang, and J. Lee. Attributes reduction using fuzzy rough sets. *IEEE Transactions on Fuzzy Systems*, 16(5):1130–1141, Oct 2008.

[34] Y.-Q. Yao, J.-S. Mi, and Z.-J. Li. Attribute reduction based on generalized fuzzy evidence theory in fuzzy decision systems. *Fuzzy Sets and Systems*, 170(1):64 – 75, 2011.

[35] L. Yong, H. Wenliang, J. Yunliang, and Z. Zhiyong. Quick attribute reduct algorithm for neighborhood rough set model. *Information Sciences*, 271(0):65 – 81, 2014.

[36] J. ZHAO and Y. hua ZHOU. New heuristic method for data discretization based on rough set theory. *The Journal of China Universities of Posts and Telecommunications*, 16(6):113 – 120, 2009.

[37] S. Zhao and E. C. Tsang. On fuzzy approximation operators in attribute reduction with fuzzy rough sets. *Information Sciences*, 178(16):3163 – 3176, 2008.

[38] K. Zheng, J. Hu, Z. Zhan, J. Ma, and J. Qi. An enhancement for heuristic attribute reduction algorithm in rough set. *Expert Systems with Applications*, 41(15):6748 – 6754, 2014.