An Incremental Genetic Algorithm Hybrid with Rough set Theory for Efficient Feature Subset Selection

N.Nandhini, K. Thangadurai

Abstract : Rough Set Theory (RST) is the most successful tool implemented in relevant feature selection or feature reduction domain. Conventionally, the feature subset selection methods make use of hill climbing strategy to find the reducts. The major limitation of such methods are that they are ineffective to find the optimal reduct since they can't give assurance for optimal feature subset. Hence the researchers moved to heuristic based feature subset selection methods. There are two types of data are subjected for feature subset selection: static and dynamic data. Static data has finite number of samples with finite set of attributes, whereas dynamic data are keep growing. This research work focuses on dynamic data, where new attributes might be added over a period of time. For such data, the feature subset algorithm doesn't require to be executed from the beginning whenever an attribute gets added. The class of incremental feature selection algorithms are efficient and prove their significance with these dynamic data. In this paper, an Incremental Genetic Algorithm (IGA) hybrid with RST is proposed for efficient feature selection with dynamic data, where GA is used to search relevant features heuristically while employing RST based fitness function. The proposed IGA-RST approach has three advantages like, (i) it starts with an effective initial population construction method which accelerates the convergence (ii) the fitness functions consists of feature weights estimated using Pseudo-Inverse matrix, which reduces the IGA-RST algorithm's computation cost, (iii) a novel incremental approach is proposed to construct the reduct for a group of top-weighted attributes first to find the partial reduct, further continued with the next set of top-weighted attributes while considering the partial reduct as elite chromosomes, which is to handle dynamic and higher dimension data. The proposed IGA-RST based feature reduction's performance is evaluated with benchmark datasets from UCI machine learning repository. Investigation results indicate that the IGA-RST improves the efficiency of feature subset selection significantly.

1. INTRODUCTION

Moreover, the dimension increases with respect to time. Such dynamic data requires efficient learning techniques to discover the knowledge. Hence, for the dynamic data, incremental learning algorithms are more suitable as they detect the changes and adopt accordingly. In general, the incremental algorithm learns the current data and update its hypothesis whenever the data gets changing, therefore it is not mandatory to execute the learning algorithm once again for the complete set of data. In this way, the incremental algorithms are more preferred as they minimizes both time and space complexity. In this paper, a novel GA-RST based incremental algorithm is proposed to discover the most relevant features from dynamic datasets.

In data mining, feature subset selection (Zhong&Skowron, 2001; Thangavel&Pethalakshmi, 2009)isone of the initial step performed to increase the classification performance typically. The main objective is to construct the optimal feature subset either by choosing the more relevant features or by removing the irrelevant/redundant attributes from the data set. The reduced feature set is called as optimal features, reduct or simply feature subset. Performing the classification or knowledge discovering phase with these relevant features will always improve the accuracy as well as reduces the computational cost too (Alpaydin, 2010). The feature reduction approaches can be classified into three categories: filter, wrapper and embedded (Chandrashekar&Sahin, 2014). The filter approach finds the reduct based on objective function without making use of any learning algorithm. Wrapper approach starts with an empty reduct set, further it adds the features based on some selection criterion and this partial reduct is evaluated with a classifier algorithm. Based on classification's performance, the latest attribute added to reduct set is either kept in the set or removed when it degrades the classification accuracy. The embedded approach integrates the feature reduction in the training procedure to decrease the time complexity. The filter approach has been followed in this research work.

Revised Manuscript Received on June 01, 2019.

N. Nandhini, Assistant Professor,Department of MCA,SNS College of Technology (Autonomous), Coimbatore 35.

Dr. K. Thangadurai, Professor and Head Department of Computer Science Government Arts College, Karur -639 005

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

The major objective of feature subset selection is to find the smaller feature subset from a higher dimensional data while maintaining better accuracy in representing the original attributes (Hu et al., 2007). Granular Computing (GC) is an evolving approach for data mining, which utilizes fuzzy and rough sets (Pedryez 2007; Pedryez et al., 2008). Pawlak (1991) introduced the concept of Rough Set Theory (RST), is most successful tool in GC, and it has proven its significance to handle uncertain, imprecise and vague data. Feature Selection is one of the major problem space where RST has its strong contribution (Xu et al., 2009; Jensen &Shen 2004; Shen et al., 2010). In general, RST tries to explore all possible feature subset and identify the optimal one. Exploring all possible combination of feature subset is computationally very expensive for higher dimension data. The concept of exhaustive search restrict the RST to handle only small size data. From the literature, it is shown that there are two different search strategies are followed in RST: greedy (hill-climbing) or stochastic (meta-heuristic). The greedy method starts with an empty reduct set and keep on accumulating the features based on some selection criterion, or else starts with universal set and eliminating the features which are redundant (Hu et al., 2003; Jing et al., 2013; Wang 2003). Greedy methods are robust to noise at certain level, though they doesn't guarantee for optimal feature subset. Due to this limitation, researchers shifted to apply stochastic methods such as Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC), etc. Among several meta-heuristic algorithm, GA has been extensively applied for feature subset selection. Holland (1992) proposed GA, the classical meta-heuristic algorithm which simulates natural evolution, and extensively used to elucidate combinatorial optimization problems. However the combination of GA and RST do not achieve competitive performance in feature subset selection when compared to other stochastic methods (Chen et al., 2010). In this paper, an Incremental Genetic Algorithm (IGA) is proposed to select the relevant features with Rough Set Theory (RST) based objective function, which improves GA's performance with feature subset selection problem.

The rest of the paper is organized as follows: the following section presents a brief review on feature subset selection. Section 3 introduces the preliminaries of RST and GA. Section 4 discusses the conventional GA & RST based feature subset selection. Section 5 explains the proposed Incremental GA & RST based optimal feature selection. Section 6 discusses the investigation results. Section 7 concludes the paper.

2. RELATED WORKS

Genetic Algorithm (GA) based attribute reduction are positively used in most of the applications. The following are some the evidence studied from literature. Zhong et al., (2001) suggested a feature reduction method where the rough set theory is hybrid with heuristics approach. The simulation results indicate that this method is able to discover the effective reduct faster for large datasets. Jian-hua et al., (2002) developed a heuristic GA for feature reduction, where a novel genetic operator is proposed to maintain the classification performance. Chen et al., (2003) presented GA for discretization, and the experimental results indicate the significant performance. Huang et al., (2005) presented a hybrid approach of GA and RST for a fault diagnosis inference system of motherboard electromagnetic interference application. This system approaches 80% diagnostic accuracy. Chien& Yang (2006) proposed a GA and RST based learning method for feature subset selection with numerical data. The investigation results shown that this method is effective in feature selection and classification. Lv& Liu (2007) presented a Quantum Genetic Algorithm (QGA) for feature reduction, which outperforms conventional GA in terms ofoptimized parameters and time efficient. Dai et al., (2008) presented a GA for finding minimal reduct with stego-images. The quantified results specify that this method is suitable for huge datasets. Guan & Yang (2008) proposed a hybrid GA and RST for briefest reduction and demonstrated that this method is effective in attribute reduction, and reported that this method is suitable for small datasets. Crossingham&Marwala (2008) presented a GA based rough set partition optimization for HIV dataset. The investigation results show the improvement in accuracy from 57.7% to 72.8%. Zhi et al., (2009) proposed an attribute reduction algorithm with RST and immune genetic algorithm. The simulation results indicate that the immune GA is able to avoid local minimum and converge faster. Liang & Huang (2009) proposed a hybrid GA and RST approach for a web service composition application, and reported that this hybrid approach archives promising results. Cheng et al., (2010) presented a hybrid GA and RST model stock market forecasting. The investigation results indicate that this hybrid model is superior to the existing approaches. Liu et al., (2010) presented an improved adaptive GA for attribute reduction, where the crossover and mutation probabilities are adjusted dynamically. The experimental results shown that this method improves the attribute reduction performance

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

significantly. Guo et al., (2010) proposed an attribute reduction with hybrid GA and RST model for intrusion detection application. Here, the core features are identified by RST, followed by GA to identify the optimal attributes from the rest of the dataset. The experimental results indicate that this method is able to discover optimal feature subset for large dataset. Chen & Zhang (2011) presented a clustering based hybrid model of GA and RST, and demonstrated that this method has the ability to adjust the outcomes and achieve the higher accuracy rate. Jaddi& Abdullah (2013) proposed a hybrid model of GA with great deluge algorithm for rough set attribute reduction. The simulation results show the effectiveness of the algorithm to resolve the attribute reduction problem. Jing (2014) developed a hybrid GA for incremental feature selection, where RST is used as a local search for GA operations. Das et al. (2018) proposed a GA based group incremental approach for attribute reduction.

The GA based feature subset selection approaches reported so far are investigated with offline or static data. However, recently data is more dynamic, hence the static data mining algorithms are inefficient to handle such data as they need to be executed repeatedly. Rough Set Theory (RST) has significant number of proposed methods to handle static or time invariant data (Zhong&Skowron, 2001; Thangavel & Pethalakshmi, 2009), which illustrates various approaches to find the reduct. These feature reduction methods are effective at some extent (Świniarski, 2001; Pawlak & Skowron, 2007), however there are other issues to be addressed in real-time, particularly the dynamically growing data, which is naturally irregular in time. The optimal feature selection algorithms for dynamic data are classified as online algorithms, where it is not necessary to run the algorithm for the old data, whereas it is enough to execute the incremental algorithms for the new data alone.

There are numerous incremental feature selection algorithms has been proposed in the literature (Wang et al., 2002; Yang, 2007; Liu et al., 2009; Qian et al., 2010). Guan (2009) presented an incremental approach to find the reduct set based on discernibility matrix, where the matrix is incrementally updated for every new objects and the reduct set is altered correspondingly. Hu et al., (2005) proposed an incremental feature selection procedure using simple set operations, which can estimate the reductattributes from a dynamic dataset. Wang et al., (2013) developed an incremental approach for feature reduction based on information entropy. Deng et al., (2010)suggested a parallel feature reduction method using positive region with an attributes significance measure. Bazan(1996) presented an evaluation measure, stability coefficient, to estimate the quality of the dynamic reducts. Xie et al., (2013) presented a relative positive region based incremental attribute reduction algorithms, which is capable of handling both incremental attributes and objects. Liang et al., (2012) developed a group incremental approach for attribute reduction based on information entropy. Lie et al., (2014) developed a matrix and tolerance relation based incremental attribute reduction method for a dynamic incomplete system. However this method requires higher computation for learning. Dey et al.,(2011) proposed a feature reduction method based on graph-cut approach, however the performance demonstrate with small data sets shows that this method may not be able to handle large dataset. Xu et al., (2011) developed an incremental feature reduction algorithm based on integer programming. However the performance of this method is not significant to other incremental methods. Shu&Shen (2014) proposed an incremental approach using positive region to handle the new data either as one by one or batch.

3. PRELIMINARIES OF RST AND GA

The hybrid model of GA and RST has been popularly used to resolve the problem of feature selection of a higher dimensional data with their superior ability to perform faster searching and guaranteed optimal feature subset selection. In most of the GA and RST combination methods, the GA utilizes RST based fitness function while searching for the minimal feature subset. This section introduces the fundamental concepts of both RST and GA.

3.1 Rough Set Theory (RST)

Rough Set Theory (RST) is an extension of conventional set theory that supports approximation in decision making. RST partitions the domain into two disjoint categories called lower and upper approximations. The objects belong to lower approximation are the certain objects belongs to the desired class, while the upper approximation objects are likely belong to the subset. The basic idea of rough set is to find the minimal number of attributes (reducts) to describe the domain, or to eliminate the redundant and irrelevant attributes from the domain. There are two basic methods for finding reducts: the first method is based on dependency between attributes and the second one is based on discernibility matrix. This section brief about both the methods. The banana data presented in Table 1 is

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

used for illustration.

Table 1. Banana Dataset

Object	Size	Smell	Color	Field	Quality
1	14	1	1	1	1
2	15	1	1	1	1
3	7	1	1	1	2
4	12	2	1	1	2
5	13	1	2	2	2
6	12	2	2	1	3
7	14	2	3	2	3
8	15	2	3	2	3

Most of the real-time dataset are presented in table format, called as information system, collection of objects (in rows) and attributes (in columns). For example, consider a medical information system, patients are the objects and their symptoms and measurements are the attributes. Sometimes the information system might be available with the decision attribute are known as decision systems. For example, the medical dataset, might be presented with the diagnosis of each patient. The terms attribute and feature are mean the same and interchangeably used in this paper.

Table 1 shows a sample decision system to identify the quality of banana fruit, it has eight objects with five attributes, where the fifth one is the decision attribute. The first four attributes {size, smell, color, field} are the conditional attributes and {quality} is the decision attribute. A decision system is said to be consistent, only when the set of objects having similar value for their conditional as well as decision attribute values.

In general, an information system (I) is a collection of finite objects $\mathbb{U}$ and a finite set of attributes $\mathbb{A}$, where both should be non-empty sets, represented asI = ($\mathbb{U}$, $\mathbb{A}$). For each attribute, their values are denoted as a: $\mathbb{U} \rightarrow V_a$ for every a $\in \mathbb{A}$. For a typical decision system, there will be two type of attributes such as conditional (C) and decision attributes (D), i.e., $\mathbb{A}$ = { C$\cup$ D }. Here, a decision d $\in$ D is itself a function d :$\rightarrow$ {0, 1} such that for a $\in \mathbb{U}$, d(a) = 1 if a has class d and d(a) = otherwise.

3.1.1 Indiscernibility Relation

For any P $\subseteq$ A there is an associated equivalence relation IND(P) defined as

$$IND(P) = \{(x, y) \in \mathbb{U}^2 | \forall a \in P, a(x) = a(y)\} \quad (1)$$

This equivalence relation relates two objects which are equivalent if any only if they have the same attribute values given by P. The partition of $\mathbb{U}$ determined by IND(P) is denoted $\mathbb{U}/IND(P)$ or $\mathbb{U}/P$, which is simple the set of equivalence classes generated by IND(P):

$$\mathbb{U}/IND(P) = \bigotimes \{\mathbb{U}/IND(\{a\}) | a \in P\}, \quad (2)$$

where

$$A \otimes B = \{X \cap Y | \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \quad (3)$$

Consider (x, y) $\in$IND(P), it represents that x and y are indiscernible with the attributes from P. further the equivalence class relation for P is defined as [x]p, x $\in \mathbb{U}$.For the illustrative example, if P = {smell, color}, then objects 1, 2, and 3 are indiscernible; as are objects 7 and 8. IND(P) creates the following partition of $\mathbb{U}$:

$$\mathbb{U}/IND(P) = \mathbb{U}/IND(smell) \otimes \mathbb{U}/IND(color)$$
$$= \{\{1,2,3,5\},\{4,6,7,8\}\} \otimes \{\{1,2,3,4\},\{5,6\},\{7,8\}\}$$
$$= \{\{1,2,3\},\{5\},\{4\},\{6\},\{7,8\}\} \quad (4)$$

3.1.2 Lower and Upper Approximations

Let X $\subseteq \mathbb{U}$ can be approximated using only the information contained within P by constructing the P-lower (denoted by $\underline{P}X$) and P-upper approximations (denoted by $\overline{P}X$) of the classical crisp set X, defined as

$$\underline{P}X = \{x | [x]_B \subseteq X\} \quad (5)$$

$$\overline{P}X = \{x | [x]_B \cap X \neq \emptyset\} \quad (6)$$

The lower approximation $(\underline{P}X)$, also known as the positive region, defines the objects that are certainly belongs to concept X with respect to information in attribute P. The upper approximation $(\overline{P}X)$ defines the object can be likely to be classified to concept X with respect to information in attribute P. Collectively, the tuple $\langle \underline{P}X, \overline{P}X \rangle$ is called a rough set. The set $BN_B(X) = \underline{P}X - \overline{P}X$ is called the boundary region and contains objects that cannot be classified into X with the knowledge of P.

3.1.3 Attribute Dependency

Let P and Q be equivalence relations over $\mathbb{U}$, then the positive regions is defined as

$$POS_P(Q) = \bigcup_{x \in \mathbb{U}/Q} \underline{P}X$$

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

The positive region combines all objects of $\mathbb{U}$ that can be classified to classes of $\mathbb{U}/Q$ using the information contained within attributes P. For example, P = {smell, color} and Q = {quality}, then

$$POS_P(Q) = \cup\{\emptyset, \{4,5\}, \{6,7,8\}\} = \{4, 5, 6, 7, 8\} \quad (8)$$

For the equivalence relation, {{1,2,3},{5},{4},{6},{7,8}} with three classes in Q={1,2,3}, no one subset can be classified to class-1, which results in {ϕ}. The subsets {4} and {5} are classified to class-2, which results {4, 5}, and the subsets {6} and {7, 8} are classified into class-3, which results in {6, 7, 8}. By combining all the results the positive region for P{smell, color} with Q{quality) is {4, 5, 6, 7, 8}.

The results illustrates that the objects 4, 5, 6, 7, and 8 can surely be categorized as fitting to a class in attribute *quality*, while considering the attributes *smell* and *color*. The remaining objects cannot be defined certainly as they hold different attribute values.

A major problem in data mining is to calculate the attribute dependencies between each other. In common, a set of decision attributes Q depends on a set of conditional attributes P, denoted by P ⇒Q,if all attribute values from Q are uniquely classified by values of attributes from P. If there exists a functional dependency between values of Q and P, then Q depends completely on P. In RST, dependency is defined as:

For P, Q ⊂𝔸, means that Q depends on P with a degree k(0 ≤ k ≤ 1), denoted as $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \quad (9)$$

where |F| denotes the cardinality of set F.

If k = 1, Q relies completely on P, if 0 < k < 1, Q reliespartly (in a degree k) on P, and if k=0 then Q does not rely on P. In the illustrative example, the degree of dependency of attribute {quality} from the attributes {smell, color} is:

$$\gamma_{\{smell,color\}}(\{quality\}) = \frac{|POS_{\{smell,color\}}(\{quality\})|}{|\mathbb{U}|} = \frac{|\{4,5,6,7,8\}|}{|\{1,2,3,4,5,6,7,8\}|} = \frac{5}{8}$$

Thus, the Q{quality} class has k=0.625 degree over P{smell, color} attributes. The degree of dependency could be estimated by choosing different possible set of attributes from 𝔸. The subset of attribute with higher degree in dependency is considered as most significant attribute set. If the significance is 0, then the feature is dispensible. More formally, given P, Q and a feature x ∈ P, the significance of feature x upon Q is defined by

$$\sigma_P(Q, x) = \gamma_P(Q) - \gamma_{P-\{x\}}(Q) \quad (11)$$

For example, if P = {size, smell, color} and Q = {quality} then

$$\gamma_{\{size,smell,color\}}(\{quality\}) = |\{1,2,3,4,5,6,7,8\}|/8 = 8/8$$
$$\gamma_{\{size,smell\}}(\{quality\}) = |\{1,2,3,5,7,8\}|/8 = 6/8$$
$$\gamma_{\{smell,color\}}(\{quality\}) = |\{4,5,6,7,8\}|/8 = 5/8$$
$$\gamma_{\{size,color\}}(\{quality\}) = |\{1,2,3,4,5,6,7,8\}|/8 = 8/8$$

Also, calculating the significance of the three attributes gives:

$$\sigma_P(Q, size) = \gamma_{\{size,smell,color\}}(\{quality\}) - \gamma_{\{smell,color\}}(\{quality\}) = 3/8$$
$$\sigma_P(Q, color) = \gamma_{\{size,smell,color\}}(\{quality\}) - \gamma_{\{size,smell\}}(\{quality\}) = 2/8$$
$$\sigma_P(Q, smell) = \gamma_{\{size,smell,color\}}(\{quality\}) - \gamma_{\{size,color\}}(\{quality\}) = 0$$

From this it shows that attributes {size} and {color} are indispensable, but attribute {smell} can be dispensed with then considering the dependency between the decision attribute and the given individual conditional attribute.

3.1.4 Reducts

Most of the application expects to maintain a summarized form of the information table. Dimensionality reduction is the procedure to discover a minimal subset of featureto represent the actual information table. In Rough Set Theory, the minimal feature subset is called as reduct (R), derived from the complete set of conditional attributes (C) and decision attribute (D), to satisfy that $\gamma_R(D) = \gamma_C(D)$. From the literature, R is a minimal subset if $\gamma_{R-\{a\}}(D) = \gamma_R(D)$ for all a ∈ R. This states that the deleting an attribute from the reduct would definitely affect the dependency degree. An information system might have several reduct sets, and the collection of them is given as

$$R_{all} = \{X \mid X \subseteq C, \gamma_X(D) = \gamma_C(D); \gamma_{X-\{a\}}(D) \neq \gamma_X(D), \forall a \in X\} \quad (10)$$

(12)

The intersection of all the sets in R_{all} is called the core reduct, eliminating the attributes presented in the core affects the dependency of the information system. For several problems, the main objective is to find the

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

reduct with lower number of attributes, without affecting the dependency degree, is denoted as $R_{min} \subseteq R_{all}$:

$$R_{min} = \{X \mid X \in R_{all}, \forall Y \in R_{all}, |X| \leq |Y|\} \quad (13)$$

3.2 Genetic Algorithm (GA)

Holland (1992) proposed an adaptive heuristic search procedure, Genetic Algorithm (GA) for resolving optimization problems. It mimics genetic process of natural evolution. The advantages of GA are, it explores the search space in all possible directions, for each iteration a group of solutions are evaluated for faster convergence, and various genetic operators improves the chance of achieving the global optimal result. Moreover, GA can utilize multi-objective fitness function while searching with a higher dimensional data. Conventional GA starts with a random population of chromosomes, where each chromosome represents a candidate solution. Fitness value of each chromosome is estimated and the population is subjected with the genetic operators like parent selection, crossover and mutation to generate another possible population. The selection operators are used to choose the chromosomes with higher fitness value, the crossover operator is used to exchange the data between two chromosomes (parents), and the mutation operator is used to avoid premature convergence. This process will continue to explore almost all possible chromosomes until the optimal solution is reached or a termination condition is satisfied. GA involves number of parameters to be well tuned for better performance. The following are the list of parameters used in GA

- Population size
- Termination condition
- Parent selection method
- Crossover and mutation type
- Crossover and mutation probability
- Single / Multi-objective fitness function

It is reported in the literature that, there is no single value for each parameter achieves optimal results for all the real-world problems. The parameter values are problem dependent, and they have to be tuned to find the suitable value for the corresponding problem. The following text describes the procedure of hybrid GA with RST for feature subset selection problem in detail.

3.2.1 Hybrid Genetic Algorithm with Rough Set Theory for Feature Subset Selection

A hybrid Genetic Algorithm (GA) with Rough Set Theory (RST) based fitness function, called HGA-RST, for attribute selection is described in this section. The first step of GA is to construct the population of chromosomes, the candidate solutions, based on either real or binary encoding. The chromosomes are either a sequence of real numbers or binary digits known as genes. Here, for the problem of feature selection, the binary encoding is used, where the length of the chromosome is equal to the number of attributes present in the dataset, each gene could be either in on (one) or off (zero) state to represent the corresponding attribute is selected for the feature subset or not. A 'n' number of chromosomes has to be built randomly for the initial population. For example, for the Banana data set as given in Table 1, the following are some of the possible chromosomes:

Chromosome-1	1	0	1	1
Chromosome-2	0	1	0	1
Chromosome-3	1	0	0	1
Chromosome-4	0	0	1	1
Chromosome-5	1	1	0	1

The length of a chromosome is four, as the dataset contains four conditional attributes {size, smell, color, field}. The first chromosome {1, 0, 1, 1} represents the feature subset of {size, color, filed} attributes, while ignoring {smell} attribute as it encodes zero for the second attribute. Similarly the second and third chromosomes considers {smell, field} and {size, field} feature subsets respectively. The following pseudocode illustrates the procedure of initial population generation.

Algorithm – initial_population (N)

Input: Number of conditional Attributes (N), also represents the chromosome's length

Output: Initial Population (P)

(i) P ← ∅;

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

(ii) for each chromosome (ch)
a. for each attribute

Generate a random number q, $0 \le q \le 1$

if (q < threshold) then
$ch_{i,j} = 1$
else
$ch_{i,j} = 0$
b. $P \leftarrow P \cup ch_i$

(iii) Return P

After the construction of initial population, each chromosome is evaluated with a fitness function to measure the quality of it. The evaluation could be either classifier based or classifier independent, the former one is the wrapper approach and the latter is the filter approach. This research work follows the filter approach, here the significance of the feature subset is evaluated with a quality measure, which is based on the attribute dependence measure $\gamma_S(D)$, where 'S' is the selected feature subset and D is the decision attribute (Wang et al., 2007), defined as

$$fitness(S) = \alpha \cdot \gamma_S(Q) + \beta\left(1 - \frac{|S|}{|C|}\right) \quad (14)$$

where $S \subseteq C, \gamma_S(D)$ is the classification significance of the feature subset S according to the decision attribute D, α and β are the two constant parameters used to represent the quality of the subset and their cardinality, $\alpha \in [0,1], \beta = 1 - \alpha$. Hence the power of RST is used in GA to make the feature subset selection effectively.

Genetic Operators

Once the fitness values are estimated for each chromosomes, then they are subjected to three genetic operators such as selection, crossover and mutation for generating next population of solution.

Selection

There are number of methods available for chromosome selection, here the roulette-wheel selection scheme is used for constructing next generation of chromosomes. Initially the chromosomes are sorted in ascending order based on their fitness values, and their fitness probability is estimated as given

$$p(i) = \frac{fitness(ch_i)}{\sum_{j=1}^{n} fitness(ch_j)} \quad (15)$$

Then the cumulative probability is estimated as

$$cp(i) = \sum_{j=1}^{i} p(j) \quad (16)$$

In the next step, a random number 'q' is chosen between 0 and 1, based on q, the first chromosome in the order has higher cumulative probability is selected for the next generation. This procedure is repeated for 'n' number of times to construct the next generation. Here, the chromosomes with higher fitness take greater chance to get selected, also there might be a chance of getting multiple copies of the same chromosome. This redundancy is avoided by applying the next two genetic operators. The following pseudocode explain the roulette wheel selection procedure.

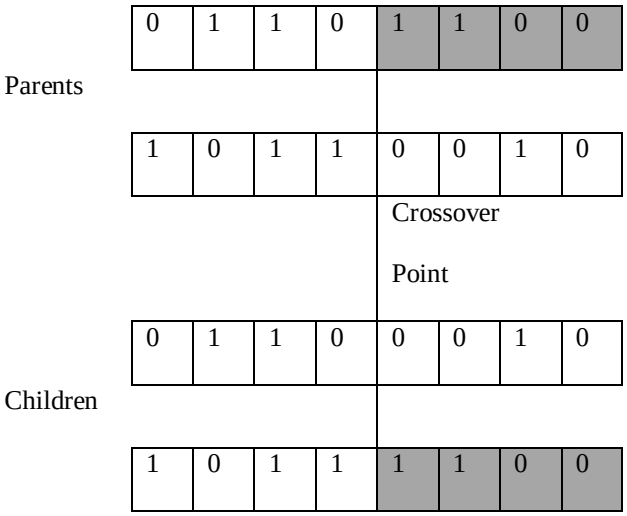
Algorithm – select_operation(P, *f*)

Input: Population of chromosomes, P; and their fitness values
Output: Population offspring, P'
(i) $P' \leftarrow \varnothing$
(ii) for i $\leftarrow$ 1 to n
a. Sort the chromosomes in ascending order based on their fitness values
b. Find their fitness probabilities and accumulate them, cp
c. Generate a random number q, $0 \le q \le 1$
d. Select the *i*th chromosome such that $P_{i-1} \le q \le P_i$
e. $P' \leftarrow P' \cup$ chi
(iii) Return P'

Crossover

With the newly selected chromosomes, every successive pairs (parents) are chosen two exchange their genes of bit, this operation is called as crossover. There are multiple crossover operators reported in the literature. An one-point crossover is applied here, where a random bit position is chosen for matting, with that index, the bits available at the right side of each parent will be swapped, thus a new chromosome will be constructed. This procedure will be repeated for all set of parents. Figure X illustrate the single-point crossover operation.

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication



A genetic parameter called crossover probability (p_m) is set to perform this operation, i.e., consider that the p_m = 0.6, then only 60% of the parents will be appear for crossover operation, the rest of the parents will be retained as it is. Sometimes, this is referred to retain the chromosome with higher fitness value as remain unchanged, those are known as elite parents and the concept is known as elitism. The following pseudocode explains the crossover operation.

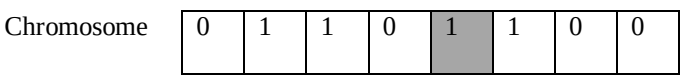
Algorithm – crossover_operation(P')

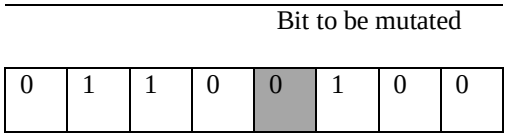
Input: Selected offspring (P')
Output: Population after matting (P'')

(i) Choose the parents for crossover operation based on p_m
(ii) For every pair of parents
  a. Choose a random crossover point
  b. Exchange the gene of information between ch_i & ch_{i+1}
(iii) Return P''

Mutation

Next to crossover operation, the mutation operator is performed with every single chromosome, where a random number is generated to choose a gene (bit), which is then complemented, i.e, if the bit is 1, then it will be changed to 0, vice versa. This operation is called as uniform mutation, and controlled by the mutation probability (p_m). Mutation operator is used to avoid premature convergence, and the probability will be set with very small value in general. Figure Y depicts a typical mutation operation.





The initial population will be regenerated after applying all three genetic operators. Then the best fitness value of the new population is estimated and compared with the best value of the old population, the global best chromosome is stored and the process continued with the best population. This procedure will be repeated for a finite number of iterations or till the algorithm converges. Suppose if the algorithm produces same results for about 'k' number of iterations, which is called as convergence, then the algorithm will be terminated.

The following pseudocode illustrates the hybrid GA with RST (HGA-RST) algorithm for feature subset selection.
Algorithm – HGA-RST based Feature Selection

Input: An information System I = (U, C ∪ D) and GA parameters
Output: An optimal feature subset

(i) Initialize the population P ←initial_population(N)
(ii) Repeat
  a. For i = 1 to n
  b. Select the parents, P' ←select_operation(P, *f*)
  c. Perform crossover, P'' ←crossover_operation(P')
  d. Perform mutation operation with P''
  e. Find the local best & update the global best chromosome
  f. Replace P with P''
(iii) Until termination condition
(iv) Return the feature reduct represented by the global best chromosome

4. HYBRID INCREMENTAL GENETIC ALGORITHM WITH ROUGH SET THEORY FOR FEATURE SUBSET SELECTION

Though the HGA-RST algorithm can discover optimal reduct for a dataset, it is not suitable for dynamic and higher dimensional data. The literature studied from Section 2 shown that there are several incremental approaches proposed for incremental feature selection. However, still there is a space for improving them with better time and space complexity. In this paper an Incremental GA hybrid with RST (IGA-RST) is proposed to

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

solve feature subset selection with dynamic and higher dimensional data. The basic idea of the proposed IGA-RST approach is to explore the feature space as local groups (bins) rather than treating them as a whole. The proposed method is similar to the method proposed in Das et al., (2018), however, they start with the global data set, and for further dynamic attribute addition, the previous reducts are carried out while estimating the reduct only for the recently added attributes alone. Though this method achieves significant reduct performance, starting with a complete set of conditional attributes might require greater computational cost for higher dimensional data. The proposed method partitions the dataset into smaller bins and the reducts are calculated for each bins and accumulating the subset sequentially, hence the proposed approach follows incremental approach efficiently. For partitioning the dataset, initially the feature weights are estimated using Pseudoinverse method as discussed below.

4.1 Moore-Penrose Pseudoinverse based Feature Weights

Ben-Israel &Greville (2003) introduced the concept of Moore-Penrose inverse which is also known as Pseudoinverse or Generalized Inverse, used to find the solutions for least square systems. The Pseudoinverse has the advantage that this could resolve the problem with rank deficient matrices, since each column vector of the result has lowest norm, which is the expected characteristics to solve least square systems. The Moore-Penrose inverseof a m×n matrix H is a distinct n×m matrix H^+, satisfies the given conditions

$$HH^+H = \mathrm{H}, H^+\mathrm{H}H^+ = H^+, (\mathrm{H}H^+)' = \mathrm{H}H^+, (H^+H)' = H^+H \quad (17)$$

Where H' represents the matrix transpose, and the preferred least square result could be derived as

$$W = H^+W \quad (18)$$

Each time when H has the rank (n), i.e., full rank, the Moore-Penrose inverse transforms to the usual Pseudoinverse:

$$H^+ = (H'H)^{-1}H' \quad (19)$$

Conversely, if the matrix H has lower rank than n (rank deficient), at that time the H+ computation is highly expensive. The cost could be reduced by equipping several methods form the literature. Singular Value Decomposition (SVD) is one such method which is used to estimate the Pseudoinverse matrix with less time complexity. Some of the readily available functions are "pinv" and "Pseudoinverse" from the tools like Matlab and Mathematica respectively. However, these methods are suitable for smaller matrices, if the matrix is large then these methods requires greater computation time. Other than SVD, Greville's algorithm and Gram-Schmidt ortho-normalization (GSO) are other two widely acceptable methods for estimating Pseudoinverse matrix, and they are iterative in nature (Ben-Israel &Greville, 2003). Rakha (2004) presented several extensions of Pseudoinverse matrix to design efficient procedures for parallel commutating (Wei & Wang, 2002). Still, the performance of these extensions are not satisfactory on serial processors.

Courrieu (2005) presented an approach to estimate the Pseudoinverse matrices with minimal time complexity. This basic idea is derived from reverse order law as suggested in Rakha (2004), and based on full rank Cholesky factorization of possible singular symmetric positive matrices (Courrieu, 2002). Courrieu (2005) presented the source code of the proposed generalized inverse method "geninv", written in Matlab tool. This function accept a matrix of any dimension and return its pseudoinverse matrix. The basic procedure follows the theory of full rank Cholesky factorization of H'H and its inversion of L'L. The time complexity requires for these computation are $O(n^3)$ and $O(n^2)$ respectively with serial processor. The complexity could be further reduced to O(n) for H'H and O(log r) for L'L computation with parallel architecture (Courrieu, 2004).

Here the pseudo inverse matrix is calculated for a matrix of conditional attributes given by {C}. The generated pseudo-inverse matrix G^+ of size k*n, holds the weights (w) of each attributes towards a common decision value. This procedure needs O(k log n) time complexity.

For an illustrative example, consider the 'Banana' dataset as given in Table 1, the corresponding pseudo inverse weights for each attribute is given in Table 2.

Table 2. Pseudoinverse Weights for 'Banana' Dataset

Object	Size	Smell	Color	Field	Quality
1	14	1	1	1	1
2	15	1	1	1	1
3	7	1	1	1	2

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

4	12	2	1	1	2
5	13	1	2	2	2
6	12	2	2	1	3
7	14	2	3	2	3
8	15	2	3	2	3
Weights	**-0.0606**	**1.1266**	**0.2867**	**0.4861**	

As shown in Table-2, the pseudoinverse weights (w) represents the significance of the attributes, the {smell} attributes has greater relevance among the others, whereas the {size} attribute takes lower weights represents that this attribute is no longer relevant to the dataset. However, most of the RST based reduction algorithms start the feature reduction process by exploring the conditional attributes set in sequence, this kind of exploration makes the reduction algorithm more expensive in terms of time complexity. Hence the features are sorted based on these weights to give higher priority to top weighted attributes and moreover the Incremental-GA is setup to start exploring these top weighted attributes first. For another illustration, even if two attributes has same set of attribute values the pseudoinverse matrix has the ability to return similar weights, hence the redundant attributes could be eliminated effortlessly. This property of pseudoinverse weights prove its consistence and robustness. For an illustration, with the 'Banana' dataset, the {filed} attribute values are copied with {color} attribute value as shown in Table-3. The pseudoinverse weights (w) for the modified 'Banana' dataset shown that the {color} and {filed} attributes are similar, hence any of the attribute could be removed from the dataset as it is redundant.

Table 3. Pseudoinverse Weights for 'Banana' Dataset with redundant attributes

Object	Size	Smell	Color	Field	Quality
1	14	1	1	1	1
2	15	1	1	1	1
3	7	1	1	1	2
4	12	2	1	1	2
5	13	1	2	2	2
6	12	2	2	2	3
7	14	2	3	3	3
8	15	2	3	3	3
Weights	**-0.0330**	**1.0006**	**0.2849**	**0.2849**	

Once the Pseudoinverse weights are calculated for the feature set {C}, then the features are sorted in descending order as the weight signifies that the greater weight represents the dataset more relevantly than the others. With the sorted conditional attribute set {C'}, the attributes are grouped into bins, the bin size is chosen in three different way such as uniform, random or bisecting bins.

- The uniform bin takes equally partitioned attribute subset from the sorted conditional attribute dataset C'. For example, consider a conditional attribute set with a cardinality of 50, then the bin size could be fixed as 10 to partition C' into 5 bins, each bin consists of 10 attributes.1
- The random bin divides the dataset into group of arbitrarily sized bins. Consider a dataset of 50 attributes, with the random bins, the feature subset could be chosen as {10}, {20}, {12} and {8}.
- In the third type, the dataset is sliced by using bisecting (binary search) method. Consider a dataset of 50 attributes, the bisecting method will partition the attributes with the bin size of {25}, {13}, {7}, and {5}.

Any of the three bin types are chosen to partition the dataset, and for each bin the IGA-RST is executed to discover the reduct and hold the result. While exploring the successive bins, the reducts derived from the previous bins are used as CORE attributes. For example, consider that C' has been divided into 5 bins with 10 attributes for each. For the first bin, C'_{b_1} if the estimated reduct is $R_{b_1} = \{4, 8\}$, this reduct will act as core attributes, while discovering the reduct for the second bin C'_{b_2}.Hence the final reduct$(\mathbb{R})$ is the union of successive bin reducts, as given.

$$\mathbb{R} = \bigcup_{i=1}^{k} reduct\left(C'_{b_i}\right) = \bigcup_{i=1}^{k} R_{b_i} \qquad (20)$$

where k is the number of bins. This procedure will be repeated to find the reduct for all the bins while accumulating the previous reducts. Hence, this approach reduces the space complexity as it requires minimal chromosome length, and also reduces the time complexity by exploring with bins of attributes rather than global. The Pseudoinverse weights are playing the major role for arranging the attributes better. At the same time, an extra bin can be reserved for the new attributes, they can buffer the dynamic data and execute IGA-RST while holding the previous reduct as core.

In addition to that, the fitness function is updated to make use of the attribute weights (w) from Pseudoinverse matrix for better convergence. Hence the fitness function given in (Wang et al., 2007) has been redefined as

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

$$fitness(S) = \frac{\alpha \cdot \gamma_S(Q)}{W} + \beta\left(1 - \frac{|S|}{|C|}\right) \quad (21)$$

$$where, W = \sum_i w_i, \ \{i \in S\}$$

The following algorithm illustrates the pseudocode of the proposed IGA-RST algorithm.

Algorithm – IGA-RST based Feature Selection

Input: An information System I = (U, C ∪ D) and GA parameters
Output: An optimal feature subset

(i) Estimate the feature weights (w) using PseudoInverse matrix
(ii) Sort the features and group them using either uniform, random or bisecting bins
(iii) Initialize the population P ←initial_population(N)
(iv) $R_{b_0} \leftarrow \emptyset$
(v) for each bin of k features
(vi) $S \leftarrow R_{b_i} \cup b_{i+1}$
(vii) Repeat
  a. For i = 1 to k
  b. Select the parents, P' ←select_operation(P, *f*)
  c. Perform crossover, P'' ←crossover_operation(P')
  d. Perform mutation operation with P''
  e. Find the local best & update the global best chromosome
  f. Replace P with P''
(viii) Until termination condition
(ix) Update the reduct, $R_{b_{i+1}} \leftarrow reduct\ (S)$
(x) Repeat from step (vi) with the next bin
(xi) Return the feature reduct represented by the global best chromosome

5. EXPERIMENTS & RESULTS

The performance of the proposed IGA-RST based feature subset selection method is studied in this section. Table 4 presents the ten datasets used for the experiments, were received from UCI machine learning repository. For each dataset, the experiments starts with partial number of the attributes first, that is according to the bin type, then the rest of attributes are incrementally added to find the final optimal feature subset.

Table 4. Datasets used for feature subset selection evaluation

Datasets	# Attributes	# Classes	# Objects
Breast Cancer	10	8	699
Dermatology	33	6	366
Ecoli	8	6	336
Iris	4	3	150
Letter	16	26	20000
Lung Cancer	56	4	32
Mushroom	22	2	8124
Parkinsons	23	3	197
Pima	8	2	768
Tic-Tac-Toe	9	2	958

Table 5 presents the list of parameters initialized for Genetic Algorithm (GA), these values are studied from the literature and also pruned according to the numerical experiments.

Table 5.Parameters settings for GA

Parameter	Description	Value
T	Maximum number of iterations	100
P_c	Crossover probability	0.8
p_m	Mutation probability	0.1
th	Threshold value used for population initialization	0.5
α	Fitness function constant	0.8
β	Fitness function constant	0.2
N	Chromosome length	|S|

Retrieval Number: A00390681S519/19©BEIESP
Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

n	Population size	10

The performance on the IGA-RST algorithm is studied with series of investigation and experiments. Initially the IGA-RST method is compared with recent static feature selection methods, followed by comparing with popular incremental approaches, and compared against with recent stochastic methods.

5.1 Performance analysis of IGA-RST algorithm with static feature subset selection methods

For the evaluation study of the feature selection performance, the static or non-incremental feature selection methods such as Correlated Feature Subset (CFS) Evaluator (Hall, 1999), Consistency Subset Evaluator (CON) (Dash & Liu, 2003), Shannon's Information Entropy (SIE) (Liang et al., 2012) and Relief-F (Kononenko, 1994) are used. The classification performance of the optimal feature subset derived from each algorithm is estimated with four classifiers reported in Wekatool are Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (kNN) and Multi-Layer Perceptron (MLP) (Alpaydin, 2010). The performance of IGA-RST is analyzed with three different bin type uniform, random and bisection are named as IGA-RST-U, IGA-RST-R, and IGA-RST-B respectively.Table 6 presents the comparison of attribute reduction methods, where the proposed IGA-RST algorithm is able to find optimum reducts than the exiting static feature reduction methods, as well as, the proposed method is able to achieve better classification accuracies. In particularly, the IGA-RST-U and IGA-RST-B approaches are outperforming the IGA-RST-R approach.

Table 6. Performance Comparison of IGA-RST with static feature reduction methods

Dataset (#Original Features)	Feature Selection Methods (#Reduced Features)	Classifiers Accuracy (%)			
		NB	SVM	kNN	MLP
Breast Cancer (9)	IGA-RST-U(3)	**96.06**	95.34	94.65	94.58
	IGA-RST-R(4)	**96.07**	95.34	94.63	94.59
	IGA-RST-B(3)	**96.02**	95.27	94.58	94.61
	CFS (4)	95.71	94.42	94.56	94.13
	CON (5)	95.99	95.27	94.56	94.56
	SIE (5)	95.99	94.34	94.56	94.21
	Relief-F (5)	95.93	94.34	94.56	94.27
Dermatology (33)	IGA-RST-U(9)	**98.98**	98.49	95.60	98.51
	IGA-RST-R(9)	98.76	**98.89**	95.61	98.48
	IGA-RST-B(9)	98.79	**98.92**	95.65	98.53
	CFS (9)	98.76	97.42	97.01	98.62
	CON (9)	98.52	98.25	95.56	98.67
	SIE (11)	98.73	98.30	97.42	98.07
	Relief-F (11)	98.72	98.45	95.56	98.46
Ecoli (8)	IGA-RST-U(4)	**99.55**	91.51	93.10	91.18
	IGA-RST-R(4)	90.24	**98.67**	94.54	95.61
	IGA-RST-B(4)	92.28	98.56	93.34	**98.89**
	CFS (8)	97.18	94.08	95.32	96.18
	CON (8)	96.45	97.16	94.05	91.60
	SIE (8)	98.34	92.50	95.97	92.90
	Relief-F (9)	91.96	95.03	89.51	98.43
Iris (4)	IGA-RST-U(3)	97.43	95.28	97.17	**98.64**
	IGA-RST-R(4)	92.73	**98.55**	96.68	97.40
	IGA-RST-B(3)	96.73	**98.84**	95.23	95.26
	CFS (8)	97.82	97.27	95.69	92.92
	CON (8)	94.31	96.41	94.69	92.65
	SIE (8)	91.82	98.44	95.00	97.74
	Relief-F (9)	97.17	97.91	93.33	94.68
Letter (16)	IGA-RST-U(10)	98.95	**99.77**	97.92	99.67
	IGA-RST-R(11)	98.97	**99.65**	96.95	99.48
	IGA-RST-B(11)	98.59	98.82	97.12	**99.09**
	CFS (13)	98.89	98.76	97.89	98.65
	CON (12)	98.91	98.56	96.89	97.43
	SIE (12)	98.56	98.76	97.05	97.03
	Relief-F (11)	98.50	98.46	97.01	98.93
Lung Cancer (56)	IGA-RST-U(5)	92.12	**98.69**	91.27	96.97
	IGA-RST-R(5)	95.09	**96.01**	87.81	97.21
	IGA-RST-B(5)	95.42	91.50	**96.02**	91.07
	CFS (8)	89.45	94.44	90.04	92.26
	CON (8)	94.87	94.75	94.75	94.43
	SIE (8)	95.37	92.55	95.35	93.63
	Relief-F (9)	90.75	97.80	88.77	95.56
Mushroom (22)	IGA-RST-U(3)	98.61	98.93	98.62	**99.89**
	IGA-RST-R(4)	98.10	98.34	**99.07**	99.05
	IGA-RST-B(3)	97.09	98.07	**99.10**	98.10
	CFS (4)	97.52	96.01	96.52	97.01
	CON (5)	98.52	98.85	98.52	96.86
	SIE (8)	98.02	98.32	99.02	96.01
	Relief-F (5)	97.04	98.03	98.03	98.10
Parkinsons (23)	IGA-RST-U(8)	96.32	96.76	96.48	**99.62**
	IGA-RST-R(7)	96.03	98.98	92.78	98.27
	IGA-RST-B(7)	93.85	**98.27**	95.36	95.01
	CFS (9)	95.54	**97.76**	96.70	95.99
	CON (9)	96.47	97.41	96.80	96.48
	SIE (10)	98.18	95.62	92.67	95.61
	Relief-F (9)	97.84	97.34	96.48	96.42
Pima (8)	IGA-RST-U(4)	98.02	98.38	**98.51**	98.32
	IGA-RST-R(4)	98.21	98.23	98.41	**98.68**
	IGA-RST-B(4)	97.30	**98.76**	97.51	98.05
	CFS (6)	98.02	98.34	98.43	98.23
	CON (6)	98.12	98.21	98.40	98.63
	SIE (6)	97.26	98.01	97.45	97.95
	Relief-F (5)	96.96	97.01	97.05	96.85
Tic-Tac-Toe (9)	IGA-RST-U(6)	96.20	**98.65**	97.25	97.27
	IGA-RST-R(7)	93.59	**97.96**	90.00	90.74
	IGA-RST-B(6)	**98.06**	93.69	94.22	93.22
	CFS (8)	95.58	95.67	96.07	94.12
	CON (9)	93.43	94.59	90.96	92.85
	SIE (8)	94.41	95.89	91.62	90.25
	Relief-F (9)	95.21	91.35	89.27	94.02

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

5.2 Performance analysis of IGA-RST algorithm with incremental feature subset selection methods

To investigate the IGA-RST feature subset selection performance, the following incremental algorithms such as IUAARI (Yang, 2007), IUAARS (Guan, 2009), DARIP(Xu et al., 2011), GIARC-L (Liang et al., 2012), and FSMV(Shu&Shen, 2014) are used. The performance is studied with the cardinality of the reduced feature set and the computation time. The experiments are carried out with a computer system with the following specifications: Intel i5-2400 processor @ 3.10 GHz, 4-GB memory, 32-bit Windows 8.1 Operating system, and the implementation is done with Matlab 2014. Table 7 illustrates the comparative results and indicate that the proposed IGA-RST based feature subset selection is able to reach minimal reduct with lowest computation time.

5.3 Performance analysis of IGA-RST algorithm with stochastic feature subset selection methods

For the next step of investigation, the attribute reduction performance is studied and compared with existing RST-based stochastic methods such as Tabu Search-RST (TS-RST) (Hedar et al., 2008), PSO-RST (Wang et al., 2007), ACO-RST (Chen et al., 2010), BGA-RST (Jing, 2014) and GA based Group-incremental feature selection (IFS) (Das et al., 2018) are used. Table 8 presents the investigation results and shown that the proposed IGA-RST outperforms the other stochastic approaches for feature reduction. And the same has been depicted in Figure 1.

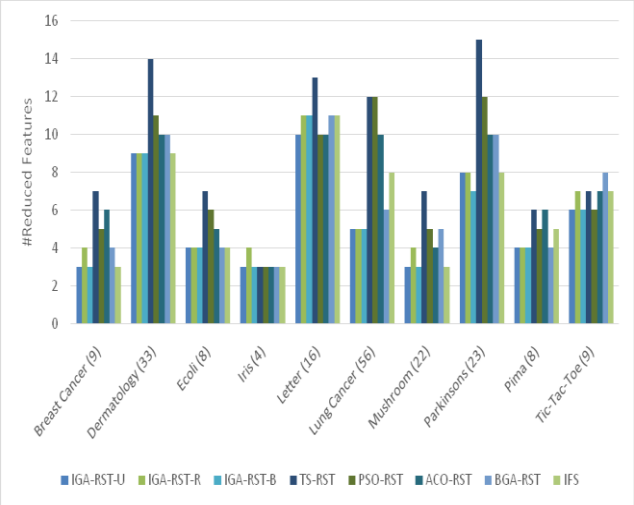


Figure 1. Feature Reduction Performance comparison between Stochastic Approaches

Datasets	IGA-RST-U		IGA-RST-R		IGA-RST-B		IUAARI		IUAARS		DARIP		GIARC-L		FSMV	
	R	T	R	T	R	T	R	T	R	T	R	T	R	T	R	T
Breast Cancer (9)	3	0.01	4	0.01	3	0.01	6	0.50	6	0.04	9	29.15	4	0.06	5	7.01
Dermatology (33)	9	0.15	9	0.20	9	0.12	11	0.50	9	0.25	11	7.34	10	0.20	11	9.03
Ecoli (8)	4	0.01	4	0.02	4	0.01	6	0.65	7	0.33	6	0.22	4	0.18	4	0.19
Iris (4)	3	0.01	4	0.02	3	0.01	3	0.72	3	0.14	3	0.08	3	0.06	3	0.07
Letter (16)	10	550.32	11	612.1	11	510.5	12	880.6	11	782.7	12	834.54	12	600.8	11	967.76
Lung Cancer (56)	5	9.67	5	11.12	5	8.25	12	22.71	11	12.77	12	9.97	9	5.28	8	25.28
Mushroom (22)	3	72.15	4	84.97	3	68.05	5	92.78	4	84.56	4	120.98	5	90.78	5	99.03
Parkinsons (23)	8	0.01	8	0.03	7	0.01	11	0.63	10	0.15	11	0.11	10	0.21	8	0.35
Pima (8)	4	0.03	4	0.03	4	0.02	6	0.44	5	0.21	6	0.17	4	0.11	5	0.71
Tic-Tac-Toe (9)	6	241.6	7	244.3	6	220.1	7	381.2	7	288.5	7	260.3	6	253.8	7	300.1

Table 7. Performance Comparison of IGA-RST with incremental feature reduction methods

Datasets	IGA-RST-U		IGA-RST-R		IGA-RST-B		TS-RST		PSO-RST		ACO-RST		BGA-RST		IFS	
	R	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A
Breast Cancer (9)	3	96.06	4	96.07	3	96.02	7	92.48	5	93.48	6	93.89	4	94.25	3	94.82
Dermatology (33)	9	98.98	9	98.89	9	98.92	14	95.59	11	95.96	10	96.91	10	97.09	9	97.76
Ecoli (8)	4	99.55	4	98.67	4	98.89	7	96.54	6	95.89	5	96.81	4	96.97	4	97.88
Iris (4)	3	98.64	4	98.55	3	98.84	3	95.29	3	95.55	3	96.76	3	96.69	3	97.59
Letter (16)	10	99.77	11	99.65	11	99.09	13	96.41	10	96.95	10	96.66	11	98.11	11	97.80
Lung Cancer (56)	5	98.69	5	96.01	5	96.02	12	95.09	12	93.19	10	93.85	6	94.48	8	94.83
Mushroom (22)	3	99.89	4	99.07	3	99.10	7	96.46	5	96.36	4	96.90	5	97.36	3	97.97
Parkinsons (23)	8	99.62	8	98.98	7	98.27	15	95.94	12	96.05	10	95.85	10	97.19	8	97.23
Pima (8)	4	98.51	4	98.68	4	98.76	6	94.96	5	96.09	6	96.68	4	96.83	5	97.51
Tic-Tac-Toe (9)	6	98.65	7	97.96	6	98.06	7	94.96	6	95.35	7	95.97	8	96.06	7	96.81

Table 8. Performance Comparison of IGA-RST with stochastic feature reduction methods

6. CONCLUSIONS

An Incremental Genetic Algorithm (IGA) hybrid with Rough Set Theory (RST) based fitness function (IGA-RST) is proposed here for efficient feature subset selection with dynamic data. The incremental approach, initially measures the feature relevance using Pseudoinverse matrix, and the features are sorted accordingly. Then the features are grouped into bins such as uniform, random or bisecting bins. Starting with the first bin, which contains the top weighted attributes, the Genetic Algorithm is used to estimate the reduct. This reduct is used as core while finding the reduct from the successive bins. The reducts are evaluated with the rough-set based dependency measure along with weight received from Pseudoinverse matrix. Accumulation of reducts from the first bin to the last, simulates the incremental approach effectively, hence the proposed IGA-RST reduces the encoding space for GA as well as the time complexity. The performance of the proposed IGA-RST based feature reduction method is compared against with static, incremental and stochastic feature reduction methods. The experimental results state the significance performance improvement with the proposed IGA-RST method as it achieves minimal reductin lowest computation time.

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

REFERENCES

1. Alpaydin. E, (2010) Introduction to Machine Learning, 2nd edition, PHI, New Delhi.
2. Bazan, J. G. (1996, July). Dynamic reducts and statistical inference. In Proc. 5-th Int. Conf. Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU (Vol. 96, pp. 1147-1151).
3. Ben-Israel, A., &Greville, T. N. (2003). Generalized inverses: theory and applications (Vol. 15). Springer Science & Business Media.
4. Chandrashekar, G., &Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.
5. Chen, Y., Miao, D., & Wang, R. (2010). A rough set approach to feature selection based on ant colony optimization. Pattern Recognition Letters, 31(3), 226-233.
6. Chen, C. Y., Li, Z. G., Qiao, S. Y., & Wen, S. P. (2003, November). Study on discretization in rough set based on genetic algorithm. In Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No. 03EX693) (Vol. 3, pp. 1430-1434). IEEE.
7. Chen, J., & Zhang, C. (2011). Efficient clustering method based on rough set and genetic algorithm. Procedia Engineering, 15, 1498-1503.
8. Cheng, C. H., Chen, T. L., & Wei, L. Y. (2010). A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. Information Sciences, 180(9), 1610-1629.
9. Chien, B. C., & Yang, J. H. (2006, October). Features selection based on rough membership and genetic programming. In 2006 IEEE International Conference on Systems, Man and Cybernetics (Vol. 5, pp. 4124-4129). IEEE.
10. Courrieu, P. (2002). Straight monotonic embedding of data sets in Euclidean spaces. Neural Networks, 15(10), 1185-1196.
11. Courrieu, P. (2004). Solving time of least square systems in Sigma-Pi unit networks. Neural Information Processing – Letters and Reviews, 4(3), 39-45.
12. Courrieu, P. (2005). Fast computation of Moore-Penrose inverse matrices. Neural Information Processing – Letters and Reviews, 8(2), 25-29.
13. Crossingham, B., &Marwala, T. (2008). Using genetic algorithms to optimise rough set partition sizes for HIV data analysis. In Advances in Intelligent and Distributed Computing (pp. 245-250). Springer, Berlin, Heidelberg.
14. Dai, M., Liu, Y., & Lin, J. (2008, June). Steganalysis based on feature reducts of rough set by using genetic algorithm. In 2008 7th World Congress on Intelligent Control and Automation (pp. 6764-6768). IEEE.
15. Das, A. K., Sengupta, S., & Bhattacharyya, S. (2018). A group incremental feature selection for classification using rough set theory based genetic algorithm. Applied Soft Computing, 65, 400-411.
16. Dash, M., & Liu, H. (2003). Consistency-based search in feature selection. Artificial intelligence, 151(1-2), 155-176.
17. Deng, D., Yan, D., & Wang, J. (2010, October). Parallel reducts based on attribute significance. In International Conference on Rough Sets and Knowledge Technology (pp. 336-343). Springer, Berlin, Heidelberg.
18. Dey, P., Dey, S., Datta, S., &Sil, J. (2011). Dynamic discreduction using rough sets. Applied Soft Computing, 11(5), 3887-3897.
19. Guan, H. B., & Yang, B. A. (2008, December). The briefest reduct of rough sets based on genetic algorithm. In 2008 IEEE International Symposium on IT in Medicine and Education (pp. 23-27). IEEE.
20. Guan, L. (2009, August). An incremental updating algorithm of attribute reduction set in decision tables. In 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery (Vol. 2, pp. 421-425). IEEE.
21. Guo, Y., Wang, B., Zhao, X., Xie, X., Lin, L., & Zhou, Q. (2010, August). Feature selection based on Rough set and modified genetic algorithm for intrusion detection. In 2010 5th International Conference on Computer Science & Education (pp. 1441-1446). IEEE.
22. Hall, M. A. (1999). Correlation-based feature selection for machine learning. Dept. of Computer Science, Univ. of Waikato, Hamilton, New Zealand, 1998 (PhD thesis).
23. Hedar, A. R., Wang, J., & Fukushima, M. (2008). Tabu search for attribute reduction in rough set theory. Soft Computing, 12(9), 909-918.
24. Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. MIT press.
25. Hu, Q., Xie, Z., & Yu, D. (2007). Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation. Pattern recognition, 40(12), 3509-3521.
26. Hu, K., Lu, Y., & Shi, C. (2003). Feature ranking in rough sets. AI communications, 16(1), 41-50.
27. Hu, F., Wang, G., Huang, H., & Wu, Y. (2005, August). Incremental attribute reduction based on elementary sets. In International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing (pp. 185-193). Springer, Berlin, Heidelberg.
28. Huang, C. L., Li, T. S., &Peng, T. K. (2005). A hybrid approach of rough set theory and genetic algorithm for fault diagnosis. The International Journal of Advanced Manufacturing Technology, 27(1-2), 119-127.
29. Jaddi, N. S., & Abdullah, S. (2013). Hybrid of genetic algorithm and great deluge algorithm for rough set attribute reduction. Turkish Journal of Electrical Engineering & Computer Sciences, 21(6), 1737-1750.
30. Jensen, R., &Shen, Q. (2004). Fuzzy–rough attribute reduction with application to web categorization. Fuzzy sets and systems, 141(3), 469-485.
31. Jian-hua, D., Yuan-xiang, L., &Qun, L. (2002). A hybrid genetic algorithm for reduct of attributes in decision system based on rough set theory. Wuhan University Journal of Natural Sciences, 7(3), 285.
32. Jing, S., She, K., & Ali, S. (2013). A Universal neighbourhood rough sets model for knowledge discovering from incomplete heterogeneous data. Expert Systems, 30(1), 89-96.
33. Jing, S. Y. (2014). A hybrid genetic algorithm for feature subset selection in rough set theory. Soft Computing, 18(7), 1373-1382.
34. Kononenko, I. (1994, April). Estimating attributes: analysis and extensions of RELIEF. In European conference on machine learning (pp. 171-182). Springer, Berlin, Heidelberg.
35. Liang, W. Y., & Huang, C. C. (2009). The generic genetic algorithm incorporates with rough set theory–An application of the web services composition. Expert Systems with Applications, 36(3), 5549-5556.
36. Liang, J., Wang, F., Dang, C., &Qian, Y. (2012). A group incremental approach to feature selection applying rough set technique. IEEE transactions on knowledge and data engineering, 26(2), 294-308.
37. Liu, B., Liu, F., & Cheng, X. (2010, October). An adaptive genetic algorithm based on rough set attribute reduction. In 2010 3rd International Conference on Biomedical Engineering and Informatics (Vol. 7, pp. 2880-2883). IEEE.
38. Liu, D., Li, T., Liu, G., & Hu, P. (2009, August). An approach for inducing interesting incremental knowledge based on the change of attribute values. In 2009 IEEE International Conference on Granular Computing (pp. 415-418). IEEE.
39. Liu, D., Li, T., & Zhang, J. (2014). A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems. International Journal of Approximate Reasoning, 55(8), 1764-1786.
40. Lv, Y. J., & Liu, N. X. (2007, November). Application of quantum genetic algorithm on finding minimal reduct.

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication

In 2007 IEEE International Conference on Granular Computing (GRC 2007) (pp. 728-728). IEEE.
41. Pawalk, Z. (1991). Rough sets: theoretical aspects of reasoning about data. Dordrecht: Kluwer Academic Publishers.
42. Pawlak, Z., &Skowron, A. (2007). Rudiments of rough sets. Information sciences, 177(1), 3-27.
43. Pedrycz, W. (2007). Granular computing-the emerging paradigm. Journal of uncertain systems, 1(1), 38-61.
44. Pedrycz, W., Skowron, A., &Kreinovich, V. (Eds.). (2008). Handbook of granular computing. John Wiley & Sons.
45. Qian, Y., Liang, J., Pedrycz, W., & Dang, C. (2010). Positive approximation: an accelerator for attribute reduction in rough set theory. Artificial Intelligence, 174(9-10), 597-618.
46. Rakha, M. A. (2004). On the Moore–Penrose generalized inverse matrix. Applied Mathematics and Computation, 158(1), 185-200.
47. Shen, Y., Li, T., Hermans, E., Ruan, D., Wets, G., Vanhoof, K., &Brijs, T. (2010). A hybrid system of neural networks and rough sets for road safety performance indicators. Soft Computing, 14(12), 1255-1263.
48. Shu, W., &Shen, H. (2014). Incremental feature selection based on rough set in dynamic incomplete data. Pattern Recognition, 47(12), 3890-3906.
49. Świniarski, R. W. (2001). Rough sets methods in feature reduction and classification. International Journal of Applied Mathematics and Computer Science, 11, 565-582.
50. Thangavel, K., &Pethalakshmi, A. (2009). Dimensionality reduction based on rough set theory: A review. Applied Soft Computing, 9(1), 1-12.
51. Wang, G. (2003). Rough reduction in algebra view and information view. International Journal of Intelligent Systems, 18(6), 679-688.
52. Wang, G. Y., Zheng, Z., & Zhang, Y. (2002). RIDAS-a rough set based intelligent data analysis system. In Proceedings. International Conference on Machine Learning and Cybernetics (Vol. 2, pp. 646-649). IEEE.
53. Wang, F., Liang, J., & Dang, C. (2013). Attribute reduction for dynamic data sets. Applied Soft Computing, 13(1), 676-689.
54. Wang, X., Yang, J., Teng, X., Xia, W., & Jensen, R. (2007). Feature selection based on rough sets and particle swarm optimization. Pattern recognition letters, 28(4), 459-471.
55. Wei, Y., & Wang, G. (2002). PCR algorithm for parallel computing minimum-norm (T) least-squares (S) solution of inconsistent linear equations. Applied mathematics and computation, 133(2-3), 547-557.
56. Xie, J., Shen, X. F., Liu, H. F., &Xu, X. (2013). Research on an incremental attribute reduction based on relative positive region. Journal of Computational Information Systems, 9(16), 6621-6628.
57. Xu, F. F., Miao, D. Q., & Wei, L. (2009). Fuzzy-rough attribute reduction via mutual information with an application to cancer classification. Computers & Mathematics with Applications, 57(6), 1010-1017.
58. Xu, Y., Wang, L., & Zhang, R. (2011). A dynamic attribute reduction algorithm based on 0-1 integer programming. Knowledge-Based Systems, 24(8), 1341-1347.
59. Yang, M. (2007). An incremental updating Algorithm for attribute reduction based on improved discernibility matrix. Chinese Journal of Computers, 30(5), 815–822.
60. Zhi, J., Liu, J. Y., & Wang, Z. (2009, August). Rough set attribute reduction algorithm based on immune genetic algorithm. In 2009 2nd IEEE International Conference on Computer Science and Information Technology (pp. 421-424). IEEE.
61. Zhong, N., &Skowron, A. (2001). A rough set-based knowledge discovery process. International Journal of Applied Mathematics and Computer Science, 11, 603-619.
62. Zhong, N., Dong, J., &Ohsuga, S. (2001). Using rough sets with heuristics for feature selection. Journal of intelligent information systems, 16(3), 199-214.

N.Nandhini was born in 1980 at Dharmapuri, Tamilnadu, India. She received Master of Computer Science in the year of 2002 under Bharathiar University, M.Phil(Computer Science)in 2005, from Bharathidasan University, Trichy, India and M.Sc Yoga Under Periyar University. She worked as an Assistant Professor, Department of Computer Science at Navarasam Arts and Science College for Women from 2002 to 2007 and Vysya College, Salem from 2007 to 2018. She published 6 papers in International Journals, paper presented in four International Conference and attended 12 seminar, FDP and workshops. She has completed NPTEL course and FDP in Data Science for Engineering. Currently she is working as Associate Professor, Department of MCA, SNS College of Technology (Autonomous), Coimbatore, Tamil Nadu, India. Her area of interest includes Data Mining, Rough Set, Genetic algorithm and Data science.

Dr. K. Thangadurai was born in 1974 at Karur, Tamilnadu, India. He received his Master of Science from the Department of Physics, Bharathidasan University in 1994, Master of Computer Applications from Jamal Mohamad college, Bharathidasan University in 1999, M.Phil from Manonmaniam Sundaranar University, India in 2002. He obtained his Ph.D. Degree from the Department of Computer Science, Vinayaka Missions University Salem in 2009. Currently he is working as Assistant Professor & Head, P.G and Research Dept. of Computer Science, Government Arts College (Autonomous), Karur. He published more than 65 papers in International Journals, 15 Papers in National Journals and attended more than 50 conferences, Seminar, FDP and workshop. He guided near 30 students under different University. His area of interests includes Software Engineering, Data Mining, OOAD and Rough Set.

Retrieval Number: A00390681S519/19©BEIESP

Published By:
Blue Eyes Intelligence Engineering &
Sciences Publication