Dimensionality Reduction Using Genetic Algorithm And Fuzzy-Rough Concepts

Moumita Saha , Jaya Sil
Computer Science and Engineering Department
Bengal Engineering and Science University
Shibpur, India
Email: moumita.saha2012@gmail.com , js@cs.becs.ac.in

Abstract

Real-world datasets are often vague and redundant, creating problem to take decision accurately. Very recently, Rough-set theory has been used successfully for dimensionality reduction but is applicable only on discrete dataset. Discretisation of data leads to information loss and may add inconsistency in the datasets. The paper aims at developing an algorithm using fuzzy-rough concept to overcome this situation. By this approach, dimensionality of the dataset has been reduced and using genetic algorithm, an optimal subset of attributes is obtained, sufficient to classify the objects. The proposed algorithm reduces dimensionality to a great extent without degrading the accuracy of classification and avoid of being trapped at local minima. Results are compared with the existing algorithms demonstrate compatible outcome.

1. INTRODUCTION

$Dimensionality\ Reduction$ compromises of selection of most relevant features that are predictive of the class-outcome and rejection of irrelevant features with minimal information loss [6]. The computational efficiency of a classification problem depends on selection of number of attributes used to build the classifier. Dimensionality reduction can be performed by applying the concept of Rough-Set Theory where the attribute values must be discrete. However, most often the values of attributes are continuous. In addition, after discretisation it is not possible to judge the extent to which the attribute value belongs to the corresponding discrete levels. This is the source of information loss, and it affects the classification accuracy negatively. Therefore it is very essential to work with real-valued data for combating the information loss which can be achieved by Fuzzy and Rough set theory. $Fuzzy-Rough\ Quick\ Reduct(FRQR)$ [10] is an efficient method of attribute reduction, overcoming the need of discretised attribute values. But the search of most informative attributes may terminates at local optimum whereas the global optimum may lie elsewhere in the search space.

In order to handle vaugeness in data and to remove local minima, the paper proposes an algorithm using fuzzy-rough set concept and genetic algorithm that searches most informative attribute set. In addition, classification accuracy using the proposed algorithm demonstrate compatible result with the other existing methods.

2. THEORETICAL FOUNDATIONS

2.1. Fuzzy Sets

Fuzzy sets were introduced by Lotfi A. Zadeh in 1965 as an extension of the classical notion of set.
A fuzzy set A in the universe of discourse U can be defined as a set of ordered pairs,

$$A = \{(x, \mu_A(x)) | x \in U\} \quad (1)$$

Where $\mu_A(x)$ is the degree of membership [0,1] of x in A. The membership function $\mu_A(.)$ maps U to the membership space M, that is $\mu_A : U \to M$.

2.2. Rough Sets

Another important parallel concept along with $Fuzzy\ Sets$ is $Rough\ Sets$. Rough Sets theory was introduced by Z. Pawlak (1982) as a mathematical

978-1-4673-0126-8/11/$26.00

approach to handle vagueness.Rough Set is a formal approximation of a crisp set (i.e., conventional set) in terms of a pair of sets which give the lower and the upper approximation of the original set [3].

2.2.1. Information and Desision System. An information system is a data table represented by $S = (U, R)$. S consisting of data universe U and set of attributes, R known as $condition\ attribute$. The attribute a (in R) characterizes each of the data elements x (in U) [5]. $Decision\ System$ is the information system represented by $S = (U, R \cup \{D\})$ where $D \notin R$ known as $decision\ attribute$.

2.2.2. Indiscernibility Relation. $Indiscernibility\ Relation\ IND(P)$ [5] is an equivalence relation defined below.

$$IND(P) = \{(e, f) \in U \times U, \quad \forall a \in R, \quad a(e) = a(f)\} \tag{2}$$

where e and f are indiscernible objects.

2.2.3. Lower Approximation. In U/R, the objects which are positively classified, called $lower\ approximation$ [8] of the set X and written as:

$$\underline{R}(X) = \{ x \in U, \quad [x]_R \subseteq X \} \tag{3}$$

2.2.4. Upper Approximation. The $R - upper\ approximation$ [8] is the union of all equivalence classes in $[\,X\,]_R$ which have non-empty intersection with the target set X. Mathematically, it is written as :

$$\overline{R}(X) = \{ x \in U, [x]_R \cap X \neq \emptyset \} \tag{4}$$

2.2.5. Positive Region. The $positive\ region$ [8] of a target set X is defined below .

$$POS_R(Q) = \cup_{X \in U/Q} \underline{R}(X) \tag{5}$$

where Q is the decision attribute.

2.2.6. Dependency. An important issue in data analysis is discovering dependencies between the attributes. Intuitively, a set of attributes D depends totally on a set of condition attributes R, denoted by $C \Longrightarrow D$. D depends on R to a degree $k\,(0 \leq k \leq 1)$ as given below.

$k = \gamma(R, D) = |(POS_R D)|/|U|$

The higher the dependency the more significant the attribute is.

2.3. Fuzzy-Rough Sets :

$Fuzzy\ set\ theory$ and $Rough\ set\ theory$ are useful computational intelligence tools in many real-world applications for dealing with vague information and to take important decision in uncertain domain [2]. Both of them works in different aspects in dealing with huge data information and have their own merits and demerits. The selection of $appropriate\ membership\ function$ is the main bottleneck of fuzzy set. On the other hand, $Rough\ set\ theory$ is useful for decision making in situation where $indisernibility$ is present. As opposed to fuzzy sets, rough sets do not require experienced knowledge engineers to provide additional information about the membership functions of the data before it can be processed. But the main bottleneck of applying rough sets theory is that it only deals with discrete data values. The merits of $rough\ sets$ and $fuzzy\ sets$ are integrated to develop a much more powerful and efficient tool known as $Fuzzy - Rough\ Set$ emerged as a new research area.

$Fuzzy - rough\ set$ is a derivation of rough set theory in which the concept of crisp equivalence class is extended using fuzzy set theory to form fuzzy equivalence classes [7], [12]. Thus, every objects have degree of membership values to lower and upper approximation fuzzy sets. In fuzzy-rough sets the $equivalence\ class\ is\ fuzzy$. In addition, $fuzziness$ is introduced in the Output classes too.

Let, the equivalence classes are in the form of fuzzy clusters $F_1, F_2 ... F_H$, which are generated by the fuzzy partitioning of the input set X into H number of clusters [1]. Each fuzzy cluster represents an equivalence class consisting of patterns of different output classes. The definite and possible number of output classes are identified using lower and upper approximations of the fuzzy equivalence classes.

2.3.1. Lower and Upper Approximation. The description of a fuzzy set X (output class) by means of the fuzzy partitions under the form of lower and upper approximations $\underline{R}X$ and $\overline{R}X$ is as follows [6],

$$\mu_{\underline{R}X}(F_j) = inf_x\{max(1 - \mu_{F_j}(x), \mu_X(x))\}, \forall j \tag{6}$$

$$\mu_{\overline{R}X}(F_j) = sup_x\{min(\mu_{F_j}(x), \mu_X(x))\}, \forall j \tag{7}$$

R is an attribute subset, $\mu_{F_j}(x)$ and $\mu_X(x)$ are the fuzzy membership values of the object x in the fuzzy equivalence class F_j and output class X respectively. Fuzzy-Rough lower and upper approximation can be defined more explicitly as:

$$\mu_{\underline{R}X}(x) = sup_{F \in U/R} min$$
$$(\mu_F(x), inf_{y \in U} max\{1 - \mu_F(y), \mu_X(y)\}) \quad (8)$$
$$\mu_{\overline{R}X}(x) = sup_{F \in U/R} min$$
$$(\mu_F(x), sup_{y \in U} min\{\mu_F(y), \mu_X(y)\}) \quad (9)$$

The tuple $< \underline{R}X, \overline{R}X >$ is defined as $fuzzy - roughset$.
The lower and upper approximation of Fuzzy-Rough Set are fuzzy unlike the crisp value of Rough-Set, represented in figure 1.

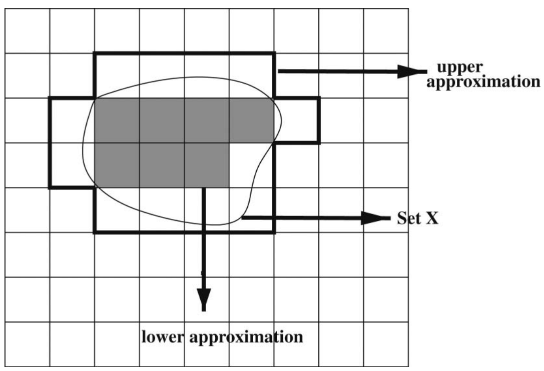

Figure 1. Lower and Upper Approximation of Fuzzy-Rough Set

2.3.2. Positive Region. As an extension to crisp positive region in traditional rough set theory, the membership of an object $x \in U$, belonging to fuzzy positive region [6] is defined as:

$$\mu_{POS_R(Q)}(x) = sup_{X \in U/Q} \mu_{\underline{R}X}(x). \quad (10)$$

2.3.3. Fuzzy-Rough Dependency. Fuzzy-Rough Dependency [6] can be defined with the aid of positive region as :
$\gamma'_R(Q) = |\mu_{POS_R(Q)}(x)|/|U|$
or,
$$\gamma'_R(Q) = \Sigma_{x \in U} \mu_{POS_R(Q)}(x)/|U| \quad (11)$$

2.4. Genetic Algorithm

Genetic algorithm (GA) is a search heuristic, used to generate useful solutions to optimization and search problems [11]. They generate solutions to optimization problems using techniques inspired by natural evolution, such as inheritance, mutation, selection, and crossover. In the genetic algorithm, a population of strings (called chromosomes), which encode candidate solution to an optimization problem, evolves toward better solutions.

3. Dimensionality Reduction

The proposed $Dimensionality\ Reduction$ [1], [12] method has been presented below.

3.1. Data Preparation

Three different kinds of dataset are considered here: (1)Hypothyroidism dataset, (2)Pulmonary-Embolism dataset and (3)Wine dataset [14]. First the dataset are partitioned using fuzzy-c-means clustering algorithm [15] which provides the degree of membership values of each object belonging to c number of clusters. The assignment of membership value of each object to different class labels are obtained by designing a $fuzzy\ inference\ system(FIS)$, as narrated below.

- Fuzzification of input data is performed based on the minimum and maximum value of each attribute that fixes the spread of membership value of respective attribute.
- The objects are grouped based on the class labels of decision attribute.
- For each conditional attribute and for every class label $linguistic\ variables$ are assigned.
- The membership function for each attribute is identified and corresponding curves are drawn.
- With the help of the decision system and linguistic variables representing the attributes, the rule-set has been developed.
- The rule set along with the membership curves are used to build the Mamdani model of FIS, which finally produce the membership values of each objects in different classes.

A dataset with two attributes and two output class-label is considered in table 1 to illustrate the above procedure.

Table 1. Illustration

$Objects$	$Attr1$	$Attr2$	$Class$
$O1$	2	10	1
$O2$	7	5	2
$O3$	5	15	1
$O4$	6	8	2
$O5$	12	16	2
$O6$	8	20	1

step 1: The minimum and maximum ranges of the attributes for spread of corresponding membership curve are:
Attr1 = 2 - 12, Attr2 = 5 - 20.
step 2: Now reconstruct the decision table by grouping the class-label, shown in table 2.

step 3: Fuzzy Values are assigned for the range of attribute values for each attributes corresponding to

Table 2. Illustration-step 2

Objects	*Attr1*	*Attr2*	*Class*
O1	2	10	1
O3	5	15	1
O6	8	20	1
O2	7	5	2
O4	6	8	2
O5	12	16	2

individual class-levels.
Attr1 : LOW (2 - 8)   HIGH (7 - 12).
Attr2 : LITTLE (5 - 16)   MORE (10 - 20).

step 4: The membership curves for the attributes are ploted in figure 2.

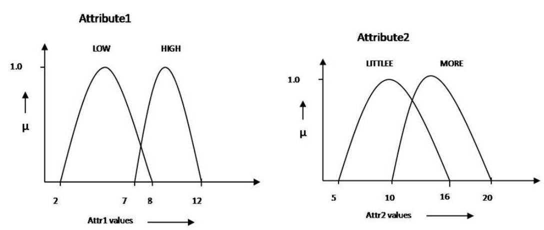

Figure 2. Membership Curve

step 5: The rule-set corresponding to the decision system is :
IF Atr1 is LOW and Atr2 is LITTLE THEN class is 1
IF Atr1 is LOW and Atr2 is MORE THEN class is 1
IF Atr1 is HIGH and Atr2 is MORE THEN class is 1
IF Atr1 is LOW and Atr2 is LITTLE THEN class is 2
IF Atr1 is HIGH and Atr2 is LITTLE THEN class is 2
IF Atr1 is HIGH and Atr2 is MORE THEN class is 2

Finally, Mamdani model has been applied for evaluating degree of membership value of each object in different classes.

3.2. DIM-RED-GA Algorithm

In the proposed method, the dependency of the decision attribute on different set of conditional attribute is calculated and attributes with highest dependency value is selected as optimum reduct by applying genetic algorithm.
In order to apply the DIM-RED-GA, population size is considered as the total dataset of the decision table. The $fuzzy - rough\ dependency\ factor$ is considered as the fitness function. The chromosomes are built up by taking the attribute values of the data objects of the decision table. Two parents are chosen and $crossing\ over$ between them is performed by randomly choosing a crossing over point with a probability of 0.10. $Mutation$ of a data object is performed with a probability of 0.02. In each generation, combination of different attributes are formed. Termination condition is kept as combination of two conditions: (1) Number of generation is greater than $MAX - NUMBER - OF - GENERATION$ or, (2) Number of times same Dependency factor appears greater than $MAX - NUMBER - OF - ITERATION$.
Following functions are applied on data objects represented by x.

function crossing-over (x, no-of-objects, no-of-attributes)
(1) Choose a cross-over point randomly within the attribute set.

cp ⟵ rand() % no-of-attributes.

(2) Choose two parents within the data objects of the decision system.

P1 ⟵ rand() % no-of-objects.
P2 ⟵ rand() % no-of-objects.

(3) Develop new generation children by cross-over.

for (i=0 ; i < cp ; i++)
temp [i] ⟵ x [P1] [i] .
x [P1] [i] ⟵ x [P2] [i] .
x [P2] [i] ⟵ temp [i] .

function mutation-flip (x, no-of-objects, no-of-attributes)
(1) Choose the attribute column to be mutated.

mut-col ⟵ rand() % no-of-attributes.

(2) Calculate the value to be added to the mutated attribute column.

(a) for (i=0 ; i < no-of-objects ; i++)
mut-val += x [i] [mut-col].

(b) Find the maximum of the mutated attribute values in max.
(c) Final mutation value is calculated as:

mut-val /= (no-of-objects * max).

(3) Update the mutated attribute column .

for (i=0 ; i < no-of-objects ; i++)
x [i] [mut-col] += mut-val.

function variation (column, no-of-attributes)
(1) Compute the attributes absent in column set and store them in left set.
(2) Generate a random attribute to be replaced from

column set and a random attribute from left set that will be utilized for replacement.

replace $\longleftarrow$ rand() % no-of-attributes in column set.
select $\longleftarrow$ rand() % no-of-attributes in left set.

(3) Replace the corresponding attribute from column set by selected attribute of the left set.

for (j=0 ; j $<$ no-of-attributes-of-column-set ; j++)
if (column [j] == column [replace])
column [j] = column [select].

DIM-RED-GA (x)
BEGIN
(1) Take input the information system along with membership values of each data objects in every classes.
(2) Initialize γ_{prev} = 0.0, γ_{best} = 0.0, flag = 0, count-of-generation = 0.
(3) *do repeat until (flag == 1) :*

(a) Count-of-generation += 1.
(b) Select randomly number of attributes to be taken.

number-attr $\longleftarrow$ rand() % NO-OF-ATTRIBUTES
number-attr += 1.

(c) Generate the combination set containing all combinations of $number-attr$ number of attributes.
(d) Select a combination from the combination set.

comb-num $\longleftarrow$ rand() % total-number-of-elements in combination set.

(e) Take the reduced information system $number-attr$ number of attributes for $comb-num^{th}$ combination of the combination set.
(f) Find out the crossing over probability.

if (cross-over probability is 10 %)
Call function *cross-over (x , no-of-objects , no-of-attributes)*

(g) Modify the information system (x) as required after crossing over.
(h) Find out the mutation probability.

if (mutation probability is 2 %)
Call function *mutation (column , no-of-attributes)*

(i) Modify the information system (x) as required after mutation.
(j) Call function *variation (x , no-of-objects , no-of-attributes)* for generating different combination of attributes.

(k) Evaluate the membership values of each objects in every clusters by $fuzzy - c - means\ clustering\ algorithm.$
(l) Calculate the $fuzzy - rough \quad lower - approximation$ of each data objects in each class.

$$\mu_{\underline{R}X}(x) = sup_{F\in U/R} min$$
$$(\mu_F(x), inf_{y\in U} max\{1-\mu_F(y), \mu_X(y)\})$$

(m) Then evaluate $fuzzy - rough\ positive\ region$ of each data objects.

$$\mu_{POS_R(Q)}(x) = sup_{X\in U/Q} \mu_{\underline{R}X}\,(x)\,.$$

(n) Finaly, evaluate the $fuzzy - rough\ dependency$ factor for the information system with specified number of attributes.

$$\gamma_R^{'}(Q) = \Sigma_{x\in U} \mu_{POS_R(Q)}(x)/|U|$$

(o) check if ($\gamma^{'} > \gamma_{prev}$)

Update : reduct $\longleftarrow$ present set of attributes.
$\gamma_{best} = \gamma^{'}$.

(p) check if (γ_{best} == γ_{prev})
iteration += 1 .
(q) Termination condition .
check if ((iteration == MAX-ITERATION-TERMINATION) || (count-of-generation == MAX-GENERATION))
flag = 1.
end do while

(4) Display the final reduced set of attributes in $reduct$ and the dependency degree of the reduced set as it is the best dependency achieved .

END

4. Results And Analysis

It is very likely that all the attributes of a decision system are not required to determine the class-label. Different attributes have different weight and evaluating the most informative attributes among them is the main aim of $Dimentionality\ reduction$ method. In order to evaluate the efficiency of the proposed algorithm, two key factors must be observed:

- The extent of dimensionality reduction, i.e observing the number of attributes present in the $reduct.$
- The accuracy of classification for the reduct set .

The proposed algorithm is applied to three datasets and comparisons between DIM-RED-GA() and other rough-set and fuzzy-rough set based methods [6],[9] are summarized in table 3.

Table 3. Dimensionality Reduction

Datasets	*Actualno−of−atr*	*FRDR−BE*	*DIM−RED−GA*
Hypothyroidism	3	3	3
PulmonaryEmbolism	4	4	4
Wine	13	8	3

The accuracy of classification is judged using three classifiers [16], shown in table 4.

Table 4. Accuracy

Accuracy									
Classifier	Hypothyroidism			Pulmonary-Embolism			Wine		
	Or−Atr	*FRQR−BE*	*DIM−RED−GA*	*Or−Atr*	*FRQR−BE*	*DIM−RED−GA*	*Or−Atr*	*FRQR−BE*	*DIM−RED−GA*
DTree	92	92	92	75	75	75	99.43	99.43	99.43
kStar	88.40	88.40	88.40	86	86	86	90.44	89.88	97.19
NBayes	88	88	88	76.5	76.5	76.5	92.69	93.82	93.82

The accuracy and coverage of reduct formation by DIM-RED-GA is compared with standard Rough-Set-Exploration-System(RSES), given in table 5.

Table 5. Comparison of Accuracy and Coverage

DataSet	Accuracy		Coverage	
	RSES	*DIM−RED−GA*	*RSES*	*DIM−RED−GA*
Hypothyroidism	75.80	92.00	48.00	100
PulmonaryEmbolism	71.40	75.00	15.20	100
Wine	88.10	99.43	69.50	100

It has been observed that the accuracy of classification and the coverage of **DIM-RED-GA** is even better than standard **RSES**. The algorithm uses fuzzy-rough concepts so we need not to discretise the values of the information system, hence overcoming the problem of information loss as in case of Rough-Set approach. In addition, we have used genetic algorithm in optimising the number of attributes of the reduct set, it is very efficient in two aspects – (1) We need not to search exhaustively all combination of attributes for evaluating the reduct set, thus improving the run-time efficiency (2) In fuzzy-rough quick-reduct, as the fuzzy-rough dependency factor is non-monotonic, it is possible that the search terminates by reaching a local optimum whereas the global optimum may lie elsewhere in the search space, this is overwhelmed by random search and updation process of optimisation of genetic algorithm.

5. Conclusions

In this paper, the shortcomings of traditional rough set attribute reduction has been highlighted and a new algorithm **DIM-RED-GA** based on fuzzy-rough sets has been proposed. The new approach incorporates the information usually lost in crisp discretization by utilizing fuzzy-rough sets to provide a more informed technique. $DIM-RED-GA$ also utilizes the concept of genetic algorithm to obtain the optimal reduct set by incorporating randomness in search process. This helps in surmounting the problem of getting struct into the local optimum as in case of fuzzy-rough QuickReduct(FRQR). It is ascertained that in $DIM-RED-GA$ the average length of reduct are less or equal than those found in other traditional methods and the accuracy of classification is also compatible.

References

[1] Rajen B. Bhatt, M.Gopal, *On Fuzzy Rough Set Approach To Feature selection*.

[2] CHEN Haishan, WANG Meihong, QIAN Feng, JIANG Qingshan, *Research on Conbined Rough Sets with Fuzzy Sets*.

[3] (Pawlak,1982) Zdzislaw Pawlak, *Rough sets, International Journal of Computer and Information Sciences*, 11, 341-356,1982.

[4] Pramod Kumar P, Prahalad Vadakkepat, Loh ai Poh, *Fuzzy-rough discriminative feature selection and classification algorithm,with application to microarray and image datasets*.

[5] Sankar K. Pal, Pabitra Mitra, *Case Generation Using Rough Sets with Fuzzy Representation*.

[6] Richard Jensen, Qiang Shen, *Fuzzy-Rough Sets Assisted Attribute Selection*.

[7] Richard Jensen, Qiang Shen, *New Approaches to Fuzzy-Rough Feature Selection*.

[8] C.P.Chandran, *Feature Selection from Protein Primary Sequence Database using Enhanced QuickReduct Fuzzy-Rough Set*.

[9] Hu Guohua, Shi Yuemei, *An Attribute Reduction Method based on Fuzzy-Rough Sets Theories*.

[10] Richard Jensen, Qiang Shen, *Fuzzy-Rough Sets for Descriptive Dimensionality Reduction*.

[11] Pengfei Guo, Xuezhi Wang, Yingshi Han *The Enhanced Genetic Algorithms for the Optimization Design*.

[12] Richard Jensen, Qiang Shen, *Semantics-Preserving Dimensionality Reduction: Rough and Fuzzy-Rough-Based Approaches*.

[13] Richard Jensen, Qiang Shen, *Are More Features Better? A Response to Attributes Reduction Using Fuzzy Rough Sets*.

[14] UCI : Machine Learning Repository, *http://archive.ics.uci.edu/ml/*.

[15] Weina Wang, Yunjie Zhang, Yi Li and Xiaona Zhang, *The Global Fuzzy C-Means Clustering Algorithm*.

[16] WEKA software classifier rules, *http://weka.sourceforge.net/doc/weka/classifiers/rules/package-summary.html* .