IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-6, June 2018 E-ISSN: 2347-2693

Divergence Based Generalized Fuzzy Rough Sets

Sheeja T.K. [1*], Sunny Kuriakose A.[2]

[1]Department of Mathematics, T. M. Jacob Memorial Government College, Manimalakunnu, Kerala, India
[2]Federal Institute of Science and Technology, Angamaly, Kerala, India

[*]*Corresponding Author: sheejakannolil@gmail.com*

Available online at: www.ijcseonline.org

Accepted: 15/Jun/2018, Published: 30/Jun/2018

Abstract— Fuzzy set theory and rough set theory are two formal mathematical tools to handle vagueness, imperfection or incompleteness in data. Fuzzy rough set theory is an embodiment of the prime features of both the theories. This hybrid theory has been proved to be an effective tool for data mining, particularly for feature selection. In this paper, generalized fuzzy rough approximations based on divergence measure of fuzzy sets in an information system is defined using a fuzzy implicator and a fuzzy t-norm. Also, the properties of the fuzzy rough approximations are investigated. Further, an algorithm for feature selection using the fuzzy boundary region of the proposed approximations is presented and experimented with twelve real data sets.

Keywords—Information System, Approximations, Divergence Measure, Fuzzy rough set, Feature selection.

I. INTRODUCTION

In this digital era, the extraction of useful knowledge from a huge amount of raw data is a great challenge[1,2,3]. Both fuzzy set theory[4] and rough set theory[5] address the problem of vagueness, imperfection or incompleteness in data. The successful applications of these theories in various fields have lead to a hybrid theory known as fuzzy rough set theory.

Fuzzy rough set theory has been successfully applied in feature selection[6]. Many different approaches to fuzzy rough sets are available in the literature. Most of the definitions are based on fuzzy relations[7,8,9,10,11]. The concept of divergence based fuzzy rough sets is introduced by T. K. Sheeja and A. Sunny Kuriakose[12]. In this paper, the generalized divergence based fuzzy rough lower and upper approximations of a fuzzy set in a fuzzy information system are defined using the a fuzzy implicator and a fuzzy t-norm. The properties of the proposed approximations are investigated. Further, an algorithm for feature selection using the fuzzy boundary region is presented. The proposed algorithm is implemented using an OCTAVE program and experimented with twelve real data sets.

The rest of the paper is structured as follows: Section II recalls some basic concepts related to fuzzy set theory and fuzzy rough set theory. The generalized fuzzy rough approximations based on divergence measure in a fuzzy information system is defined in section III and their properties are studied. Section IV describes an algorithm for feature selection using the fuzzy boundary region in the proposed. The experimental results of the application of the proposed algorithm to twelve real data sets is presented in Section V. The conclusion and future work are given in section VI.

II. RELATED WORK

In this section, some basic concepts related to fuzzy rough set theory are recalled. Further details of fuzzy set theory and rough set theory can be found in [13,14].

A. Fuzzy implicators

A *fuzzy implicator[10]* is a function $\mathcal{I}: [0,1 \times [0,1] \rightarrow [0,1]$ such that $\mathcal{I}(1,0) = 0$, $\mathcal{I}(1,1) = \mathcal{I}(0,1) = \mathcal{I}(0,0) = 1$. The implicator $\mathcal{I}$ is called a *border implicator* iff $\mathcal{I}(1,x) = x$, $\forall x \in [0,1]$. The implicator $\mathcal{I}$ is said to be *left monotonic* if it is decreasing on its first argument and *right monotonic* if it is increasing on its second argument.

Let $\mathcal{T}$, $\mathcal{S}$ and $\mathcal{N}$ be a fuzzy t-norm, t-conorm and a negator respectively. Then, the implicator $\mathcal{J}(a,b) = \mathcal{T}[\mathcal{N}(a), b]$ is called an *S-implicator[13]*. If $\mathcal{T}$ is continuous, then the implicator $\mathcal{J}(a,b) = \sup\{\lambda \in [0,1] : \mathcal{T}(a,\lambda) \leq b\}$ is called an *R - implicator[13]*. If $\mathcal{T}$ and $\mathcal{S}$ are dual with respect to $\mathcal{N}$, then $\mathcal{J}(a,b) = \mathcal{S}[\mathcal{N}(a), \mathcal{T}(a,b)]$ is an implicator known as a *QL implicator[13]*.

B. Divergence measure

Let $\mathcal{F}(U)$ be the family of all fuzzy sets on U. Then a function $\delta: \mathcal{F}(U) \times \mathcal{F}(U) \longrightarrow [0,1]$ is called a *divergence measure[15]* if and only if $\forall A, B, C \in \mathcal{F}(U)$,

i. $\delta(A,B) = \delta(B,A)$
ii. $\delta(A,A) = 0$
iii. $\max\{\delta(A \cup C, B \cup C), \delta(A \cap C, B \cap C)\} \leq \delta(A,B)$

C. Fuzzy rough sets

Let U be a nonempty set of objects and R be a fuzzy equivalence relation on U. then, the pair (U,R) is called a *fuzzy approximation space*. The *fuzzy rough lower and upper approximations* of a fuzzy subset $A \subseteq U$ defined by Dubois and Prade[7] are given by

$$\mu_{\underline{R}(A)}(x) = inf_{y\in U}\{\max[1-R(x,y),\mu_A(y)]\} \quad (1)$$

$$\mu_{\overline{R}(A)}(x) = sup_{y\in U}\{\min[R(x,y),\mu_A(y)]\} \quad (2)$$

Radzikowska and Kerre[10] generalized this definition by replacing the max and min operators by a border implicator $\mathcal{J}$ and a t-norm $\mathcal{T}$ respectively.

$$\mu_{\underline{R}(A)}(x) = inf_{y\in U}\ \mathcal{J}[R(x,y),\mu_A(y)] \quad (3)$$

$$\mu_{\overline{R}(A)}(x) = sup_{y\in U}\ \mathcal{T}(R(x,y),\mu_A(y)] \quad (4)$$

Afterwards, many extensions and generalizations of fuzzy rough approximations have been proposed by many authors [8,11,16,17,18]. A review of the different approaches to fuzzy rough set is presented in [19].

D. Feature selection using fuzzy rough sets

Fuzzy rough set theory has been successfully applied to feature selection. A number of papers were authored by Jensen and Shen[20,21,22] in which they develop a fuzzy rough quick reduct algorithm. Another approach to fuzzy-rough feature selection is to use fuzzy entropy as a criteria for feature selection[22]. Algorithms based on discernibility matrix to compute the attribute reducts are also proposed by many authors[9,23]. Fuzzy boundary region based feature selection methods are also there in the literature[23,24,25].

III. DIVERGENCE BASED GENERALIZED FUZZY ROUGH APPROXIMATIONS

Let (U,C,D) be a fuzzy information system, where U is a nonempty set of objects, C is the set of fuzzy conditional attributes and D is the set of decision attributes. If $P \subseteq C$, then each object $x \in U$ can be associated with a fuzzy set P_x on C, with membership function

$$\mu_{P_x}(a) = \begin{cases} a(x), & if\ a \in P \\ 0, & otherwise \end{cases} \quad (5)$$

Definition 3.1: Consider (U,C,D), with $U=\{x_1,x_2,\dots,x_n\}$, $C=\{a_1,a_2,\dots,a_m\}$ and $D=\{d_1,d_2,\dots,d_k\}$. Let $\delta(A,B)$ be a divergence measure of fuzzy sets. Then, the *divergence matrix* of U with respect to $P \subseteq C$ is defined as

$$\Delta_P = [\delta_{ij}]_{n\times n}, \text{ where } \delta_{ij} = \delta\left(P_{x_i},P_{x_j}\right), \forall\ i,j = 1,2,\dots,n. \quad (6)$$

Remark 3.2: If the range of δ is not a subset of [0,1], then the *normalized divergence matrix* may be considered where the matrix entries are given by $\delta_{ij}{}^* = \frac{\delta_{ij}}{\max_{i,j}(\delta_{ij})}$. Therefore, without any loss of generality, it may be assumed that $0 \le \delta_{ij} \le 1,\ \forall\ i,j = 1,2,\dots,n$.

Definition 3.3: Let (U,C,D) be a fuzzy information system. Consider a border implicator $\mathcal{J}$ and a t-norm $\mathcal{T}$. Then, the *divergence based generalized fuzzy rough lower and upper approximations* of a fuzzy set A on U with respect to the divergence measure δ are defined $\forall x_i \in U$ as

$$\mu_{\underline{\delta}(A)}(x_i) = inf_{x_j\in U}\ \mathcal{J}[1-\delta_{ij},\mu_A(x_j)] \quad (7)$$

$$\mu_{\overline{\delta}(A)}(x_i) = sup_{x_j\in U}\ \mathcal{T}[1-\delta_{ij},\mu_A(x_j)] \quad (8)$$

respectively.

Next, we will show that the proposed approximations are fuzzy sets on U.

Proposition3.4: The divergence based generalized fuzzy rough lower and upper approximations of a fuzzy set A on U are fuzzy sets on U.

Proof: Clearly, $\mu_A(x_j) \in [0,1]$, $\forall x_j \in U$ and $\delta_{ij} \in [0,1]$, $\forall\ i,j = 1,2,\dots,n$. Again, the range of the implicator $\mathcal{J}$ is a subset of [0,1]. Therefore, $\mu_{\underline{\delta}(A)}(x_i) \in [0,1], \forall i$. Similarly, $\mu_{\overline{\delta}(A)}(x_i) \in [0,1],\ \forall x_i \in U$.

The properties of the proposed approximations are given in the following theorems.

Theorem3.5: The general properties of the fuzzy rough lower and upper approximations with respect to δ are as follows:

i. $\underline{\delta}(A) \subseteq A \subseteq \overline{\delta}(A),\ \forall A \in \mathcal{F}(C)$
ii. $\underline{\delta}(\phi) = \phi = \overline{\delta}(\phi)$
iii.
  a. $\underline{\delta}(U) = U\ \ if\ \mathcal{J}\ is\ left\ monotonic$
  b. $\overline{\delta}(U) = U$
iv.
  a. $A \subseteq B \Rightarrow \underline{\delta}(A) \subseteq \underline{\delta}(B)\ if\ \mathcal{J}\ is\ right\ monotonic.$
  b. $A \subseteq B \Rightarrow \overline{\delta}(A) \subseteq \overline{\delta}(B)$
v.
  a. $\underline{\delta}(\hat{\alpha}) = \hat{\alpha}, \forall \alpha \in [0,1]\ \ if\ \mathcal{J}\ is\ left\ monotonic$
  b. $\overline{\delta}(\hat{\alpha}) = \hat{\alpha},\ \forall \alpha \in [0,1]$

Proof:

i. $\mu_{\underline{\delta}(A)}(x_i) = inf_{x_j\in U}\ \mathcal{J}[1-\delta_{ij},\mu_A(x_j)]$
$\le \mathcal{J}[1-\delta_{ii},\mu_A(x_i)]$
$= \mathcal{J}[1,\mu_A(x_i)],\ since\ \delta_{ii} = 0$
$= \mu_A(x_i), \forall x_i \in U$, as $\mathcal{J}$ is a border implicator

Also, $\mu_{\overline{\delta}(A)}(x_i) = sup_{x_j\in U}\ \mathcal{T}[1-\delta_{ij},\mu_A(x_j)]$
$\ge \mathcal{T}[1-\delta_{ii},\mu_A(x_i)]$
$= \mathcal{T}[1,\mu_A(x_i)]$
$= \mu_A(x_i), \forall\ x_i \in U$, since $\mathcal{T}(1,x) = x$

Thus, $\underline{\delta}(A) \subseteq A \subseteq \overline{\delta}(A), \forall A \in \mathcal{F}(U)$.

ii. By property (1), $\underline{\delta}(\phi) \subseteq \phi$. Also, $\phi \subseteq \underline{\delta}(\phi)$. Hence, $\underline{\delta}(\phi) = \phi$. Again, $\mu_\phi(x_j) = 0, \forall x_j \in U$.

So, $\mathcal{T}[1-\delta_{ij},\mu_{\phi}(x_j)] = \mathcal{T}[1-\delta_{ij},0]$.
Now, as $\mathcal{T}$ is an increasing function, the value of $\mathcal{T}[1-\delta_{ij},0]$ will be the supremum, when $(1-\delta_{ij})$ will be the maximum. The maximum value of $(1-\delta_{ij})$ is 1. So, $\mu_{\overline{\delta}(\phi)}(x_i) = \mathcal{T}[1,0] = 0, \forall x_i \in U$.
Thus, $\underline{\delta}(\phi) = \phi = \overline{\delta}(\phi)$.

iii.

a. We have, $\mu_U(x_j) = 1,\ \forall x_j \in U$.
So, $\mathcal{I}[1-\delta_{ij},\mu_U(x_j)] = \mathcal{I}[1-\delta_{ij},1], \forall\, x_j \in U$.
If $\mathcal{I}$ is left monotonic, then $\mathcal{I}[1-\delta_{ij},1]$ will be minimum when $(1-\delta_{ij})$ is maximum. Hence,
$\mu_{\underline{\delta}(U)}(x_i) = inf_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},1]$
$= \mathcal{I}[1,1] = 1, \forall\, x_i \in U$.
Thus, $\underline{\delta}(U) = U$.

b. $\mu_{\overline{\delta}(U)}(x_i) = sup_{x_j\in U}\,\mathcal{T}[1-\delta_{ij},\mu_U(x_j)]$
$= sup_{x_j\in U}\,\mathcal{T}[1-\delta_{ij},1]$
$= sup_{x_j\in U}\,[1-\delta_{ij}] = 1$.
Thus, $\overline{\delta}(U) = U$.

iv.

a. If $A \subseteq B$, then $\mu_A(x_j) \le \mu_B(x_j), \forall x_j \in U$.
So, if $\mathcal{I}$ is right monotonic, then $\forall x_j \in U$,
$\mathcal{I}[1-\delta_{ij},\mu_A(x_j)] \le \mathcal{I}[1-\delta_{ij},\mu_B(x_j)]$.
Therefore, $inf_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mu_A(x_j)]$
$\le inf_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mu_B(x_j)]$.
Thus, $\underline{\delta}(A) \subseteq \underline{\delta}(B)$.

b. By definition, $\mathcal{T}$ is an increasing function.
Also, $\mu_A(x_j) \le \mu_B(x_j), \forall x_j \in U$. So,
$\mathcal{T}[1-\delta_{ij},\mu_A(x_j)] \le \mathcal{T}[1-\delta_{ij},\mu_B(x_j)],\ \forall x_j \in U$.
Therefore, $sup_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mu_A(x_j)]$
$\le sup_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mu_B(x_j)]$.
Thus, $\underline{\delta}(A) \subseteq \overline{\delta}(B)$.

v.

a. For all fuzzy constants $\hat{\alpha}$, $\mu_{\hat{\alpha}}(x_j) = \alpha, \forall x_j \in U$. If $\mathcal{I}$ is left monotonic, $\mathcal{I}[1-\delta_{ij},\mu_{\hat{\alpha}}(x_j)]$ decreases as $(1-\delta_{ij})$ increases. Therefore, the infimum corresponds to the maximum value of $(1-\delta_{ij})$, which is 1. Thus, $\mu_{\underline{\delta}(\hat{\alpha})}(x_i) = \mathcal{I}[1,\alpha] = \alpha$.
Thus $\underline{\delta}(\hat{\alpha}) = \hat{\alpha}$, $\forall \alpha \in [0,1]$.

b. Since, $\mu_{\hat{\alpha}}(x_j) = \alpha, \forall x_j \in U$ and $\mathcal{T}$ is an increasing function, the supremum of $\mathcal{T}[1-\delta_{ij},\mu_{\hat{\alpha}}(x_j)]$ corresponds to the maximum value of $(1-\delta_{ij})$, which is 1. Hence, $sup_{x_j\in U}\,\mathcal{T}[1-\delta_{ij},\alpha] = \alpha$.
Therefore, $\overline{\delta}(\hat{\alpha}) = \hat{\alpha},\ \forall \alpha \in [0,1]$.

Corollary 3.6: All S-implicators and R-implicators satisfy properties (i) to (v) and all QL implicators satisfy properties (i), (ii), (iiib), (iv) and (vb) of theorem 1.

Proof:

The result follows directly from the fact that all S-implicators and R-implicators are hybrid monotonic and all QL-implicators are right monotonic.

Lemma 3.7: If $\mathcal{N}$ is an involutive fuzzy complement, then $\mathcal{N}[inf_{i\in J}(a_i)] = sup_{i\in J}[\mathcal{N}(a_i)]$, where J is an index set and $a_i \in [0,1], \forall\, i \in J$.

Proof: $\mathcal{N}$ is a decreasing function on [0,1]. Therefore, we have, $inf_{i\in J}(a_i) \le a_i \Rightarrow \mathcal{N}(inf_{i\in J}(a_i)) \ge \mathcal{N}(a_i), \forall i \in J$.

This means that $\mathcal{N}(inf_{i\in J}(a_i))$ is an upper bound for the set $\{\mathcal{N}(a_i) : i \in J\}$. Let k be any upper bound for this set. Then $k \ge \mathcal{N}(a_i), \forall i \in J$. Since $\mathcal{N}$ is decreasing and involutive, $\mathcal{N}(k) \le a_i, \forall i \in J$. Hence, $\mathcal{N}(k) \le inf_{i\in J}(a_i)$. Therefore, $k \ge \mathcal{N}(inf_{i\in J}(a_i))$. Thus, $\mathcal{N}(inf_{i\in J}(a_i))$ is the least upper bound. That is, $\mathcal{N}(inf_{i\in J}(a_i)) = sup_{i\in J}[\mathcal{N}(a_i)]$.

Theorem 3.8: Consider a fuzzy t-norm $\mathcal{T}$, a fuzzy negator $\mathcal{N}$ and a fuzzy implicator $\mathcal{I}$. If $\mathcal{N}(\mathcal{I}(a,b)) \ge \mathcal{T}(a,\mathcal{N}(y))$, then $\forall A \in \mathcal{F}(C)$,

i. $(\underline{\delta}(A^C))^C \subseteq \overline{\delta}(A), \forall A \in \mathcal{F}(C)$
ii. $(\overline{\delta}(A^C))^C \subseteq \underline{\delta}(A), \forall A \in \mathcal{F}(C)$

Proof: $\mu_{(\underline{\delta}(A^C))^C}(x_i) = \mathcal{N}\{\, inf_{x_j\in U}\,\mathcal{I}\,[1-\delta_{ij},\mu_{A^C}(x_j)]$
$= \mathcal{N}\{inf_{x_j\in U}\mathcal{I}[1-\delta_{ij},\mathcal{N}(\mu_A(x_j))]\}$
$= sup_{x_j\in U}\mathcal{N}\{\mathcal{I}\,[1-\delta_{ij},\mathcal{N}(\mu_A(x_j))]\}$
$\ge sup_{x_j\in U}\,\mathcal{T}\{1-\delta_{ij},\mathcal{N}[\mathcal{N}(\mu_A(x_j))]\}$, by assumption
$= sup_{x_j\in U}\,\mathcal{T}\{1-\delta_{ij},\mu_A(x_j)\}$
$= \mu_{\overline{\delta}(A)}(x_i)$
Therefore, $(\underline{\delta}(A^C))^C \subseteq \overline{\delta}(A), \forall A \in \mathcal{F}(U)$
Similarly, $(\overline{\delta}(A^C))^C \subseteq \underline{\delta}(A), \forall A \in \mathcal{F}(U)$.

Theorem3.9: Let $(\mathcal{T},\mathcal{S},\mathcal{N})$ be a dual triple, where $\mathcal{T}$ is a fuzzy t-norm, $\mathcal{S}$ is a fuzzy t-conorm and $\mathcal{N}$ is an involutive fuzzy complement such that $\mathcal{T}$ and $\mathcal{S}$ are dual with respect to $\mathcal{N}$. Then the divergence based fuzzy rough lower and upper approximations are dual to each other if the impicator is the S-implicator determined by the fuzzy t-conorm $\mathcal{S}$.

Proof: $\mu_{(\underline{\delta}(A^C))^C}(x_i) = \mathcal{N}\{inf_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mu_{A^C}(x_j)]$
$= \mathcal{N}\{inf_{x_j\in U}\,\mathcal{I}[1-\delta_{ij},\mathcal{N}(\mu_A(x_j))]\}$
$= sup_{x_j\in U}\mathcal{N}\{\mathcal{I}[1-\delta_{ij},\mathcal{N}(\mu_A(x_j))]\}$
$= sup_{x_j\in U}\mathcal{T}\{1-\delta_{ij},\mathcal{N}[\mathcal{N}(\mu_A(x_j))]\}$,
since $\mathcal{T}$ and $\mathcal{S}$ are dual w.r.t $\mathcal{N}$
$= sup_{x_j\in U}\,\mathcal{T}\{1-\delta_{ij},\mu_A(x_j)\}$
$= \mu_{\overline{\delta}(A)}(x_i)$

Corollary3.10: If $\mathcal{T}$, $\mathcal{S}$ and $\mathcal{N}$ represent the standard fuzzy intersection, union and complement respectively and $\mathcal{I}$ is the

S-implicator based on $\mathcal{T}$. Then, the divergence based fuzzy rough lower and upper approximations are dual to each other.

Theorem3.11: Let $(\mathcal{T}, \mathcal{S}, \mathcal{N})$ be a dual triple, where $\mathcal{T}$ is a fuzzy t-norm, $\mathcal{S}$ is a fuzzy t-conorm and $\mathcal{N}$ is an involutive fuzzy complement such that $\mathcal{T}$ and $\mathcal{S}$ are dual with respect to $\mathcal{N}$. Then the algebraic properties of the fuzzy rough lower and upper approximations with respect to δ are given below:

i. If $\mathcal{I}[a, \mathcal{T}(b,c)] \geq \mathcal{T}[\mathcal{I}(a,b), \mathcal{I}(a,c)]$, then $\underline{\delta}(A \cap B) \supseteq \underline{\delta}(A) \cap \underline{\delta}(B)$
ii. If $\mathcal{T}[a, \mathcal{T}(b,c)] \geq \mathcal{T}[\mathcal{T}(a,b), \mathcal{T}(a,c)]$, then $\overline{\delta}(A \cap B) \supseteq \overline{\delta}(A) \cap \overline{\delta}(B)$
iii. $\underline{\delta}(A \cup B) \supseteq \underline{\delta}(A) \cup \underline{\delta}(B)$, if $\mathcal{I}$ and $\mathcal{S}$ satisfy $\mathcal{I}(a, \mathcal{S}(b,c) \geq \mathcal{S}(\mathcal{I}(a,b), \mathcal{I}(a,c)$
iv. $\overline{\delta}(A \cup B) \subseteq \overline{\delta}(A) \cup \overline{\delta}(B)$, if $\mathcal{T}$ and $\mathcal{S}$ satisfy distributive laws.

Proof

i. $\mu_{\underline{\delta}(A\cap B)}(x_i) = inf_{x_j \in U}\, \mathcal{I}[1-\delta_{ij}, \mu_{A\cap B}(x_j)]$
$= inf_{x_j \in U}\, \mathcal{I}[1-\delta_{ij}, \mathcal{T}(\mu_A(x_j), \mu_B(x_j)]$
$\geq inf_{x_j \in U} \mathcal{T}[\mathcal{I}(1-\delta_{ij}, \mu_A(x_j), \mathcal{I}(1-\delta_{ij}, \mu_B(x_j)]$
$\geq \mathcal{T}[inf_{x_j \in U} \mathcal{I}(1-\delta_{ij}, \mu_A(x_j),$
$inf_{x_j \in U} \mathcal{I}(1-\delta_{ij}, \mu_B(x_j)]$
$= \mathcal{T}(\mu_{\underline{\delta}(A)}(x_i), \mu_{\underline{\delta}(B)}(x_i))$
$= \mu_{\underline{\delta}(A)\cap\underline{\delta}(B)}(x_i))$
Therefore, $\underline{\delta}(A \cap B) \supseteq \underline{\delta}(A) \cap \underline{\delta}(B)$

ii. $\mu_{\overline{\delta}((A\cap B))}(x_i) = sup_{x_j \in U}\, \mathcal{T}[1-\delta_{ij}, \mu_{(A\cap B)}(x_j)]$
$= sup_{x_j \in U}\, \mathcal{T}[1-\delta_{ij}, \mathcal{T}(\mu_A(x_j), \mu_B(x_j)]$
$\leq sup_{x_j \in U}\, \mathcal{T}[\mathcal{T}(1-\delta_{ij}, \mu_A(x_j), \mathcal{T}(1-\delta_{ij}, \mu_B(x_j)]$
$\leq \mathcal{T}[sup_{x_j \in U}\, \mathcal{T}(1-\delta_{ij}, \mu_A(x_j),$
$sup_{x_j \in U}\, \mathcal{T}(1-\delta_{ij}, \mu_B(x_j)]$
$= \mathcal{T}\left(\mu_{\overline{\delta}(A)}(x_i), \mu_{\overline{\delta}(B)}(x_i)\right)$
$= \mu_{\overline{\delta}(A)\cap\overline{\delta}(B)}(x_i))$
Thus, $\overline{\delta}(A \cap B) \supseteq \overline{\delta}(A) \cap \overline{\delta}(B)$

iii. $\mu_{\underline{\delta}(A\cup B)}(x_i) = inf_{x_j \in U}\, \mathcal{I}[1-\delta_{ij}, \mu_{A\cup B}(x_j)]$
$= inf_{x_j \in U}\, \mathcal{I}[1-\delta_{ij}, \mathcal{S}(\mu_A(x_j), \mu_B(x_j)]$
$\geq inf_{x_j \in U} \mathcal{S}[\mathcal{I}(1-\delta_{ij}, \mu_A(x_j), \mathcal{I}(1-\delta_{ij}, \mu_B(x_j)]$
$\geq \mathcal{S}\,[inf_{x_j \in U} \mathcal{I}(1-\delta_{ij}, \mu_A(x_j),$
$inf_{x_j \in U} \mathcal{I}(1-\delta_{ij}, \mu_B(x_j)]$
$= \mathcal{S}(\mu_{\underline{\delta}(A)}(x_i), \mu_{\underline{\delta}(B)}(x_i))$
$= \mu_{\underline{\delta}(A)\cup\underline{\delta}(B)}(x_i))$.
Thus, $\underline{\delta}(A \cup B) \supseteq \underline{\delta}(A) \cup \underline{\delta}(B)$

iv. $\mu_{\overline{\delta}((A\cup B))}(x_i) = sup_{x_j \in U}\, \mathcal{T}[1\, \delta_{ij}, \mu_{(A\cup B)}(x_j)]$
$= sup_{x_j \in U}\, \mathcal{T}[1-\delta_{ij}, \mathcal{S}(\mu_A(x_j), \mu_B(x_j)]$
$\leq sup_{x_j \in U}\, \mathcal{S}[\mathcal{T}(1-\delta_{ij}, \mu_A(x_j), \mathcal{T}(1-\delta_{ij}, \mu_B(x_j)]$
$\leq \mathcal{S}\,[sup_{x_j \in U}\, \mathcal{T}(1-\delta_{ij}, \mu_A(x_j),$
$sup_{x_j \in U}\, \mathcal{T}(1-\delta_{ij}, \mu_B(x_j)]$
$= \mathcal{S}(\mu_{\overline{\delta}(A)}(x_i), \mu_{\overline{\delta}(B)}(x_i))$
$= \mu_{\overline{\delta}(A)\cup\overline{\delta}(B)}(x_i))$

Corollary3.12: If $\mathcal{T}$, $\mathcal{S}$ and $\mathcal{N}$ are the standard fuzzy intersection, union and complement respectively and $\mathcal{I}$ is the S-implicator based on $\mathcal{T}$. Then,

i. $\underline{\delta}(A \cap B) = \underline{\delta}(A) \cap \underline{\delta}(B)$
ii. $\overline{\delta}(A \cup B) = \overline{\delta}(A) \cup \overline{\delta}(B)$

Theorem3.13: If δ^1 and δ^2 are two divergence measures on $\mathcal{F}(P)$ with $\delta^1(A,B) \leq \delta^2(A,B)$, $\forall A,B \in \mathcal{F}(P)$, then $\underline{\delta^1}(A) \geq \underline{\delta^2}(A)$

Proof: $\delta^1(A,B) \leq \delta^2(A,B), \forall A,B \in \mathcal{F}(P) \Rightarrow \delta^1{}_{ij} \leq \delta^2{}_{ij}$
$\Rightarrow 1-\delta^1{}_{ij} \geq 1-\delta^2{}_{ij}$
$\Rightarrow inf_{x_j \in U} \mathcal{I}\left(1-\delta^1{}_{ij}, \mu_A(x_j)\right)$
$\geq inf_{x_j \in U} \mathcal{I}(1-\delta^2{}_{ij}, \mu_A(x_j))$
$\Rightarrow \mu_{\underline{\delta^1}(A)}(x_i) \geq \mu_{\underline{\delta^2}(A)}(x_i)$
Thus, $\underline{\delta^1}(A) \geq \underline{\delta^2}(A)$. Similarly, $\overline{\delta^1}(A) \geq \overline{\delta^2}(A)$

IV. FEATURE SELECTION USING FUZZY BOUNDARY REGION

This section describes an application of the divergence based fuzzy rough approximations to feature selection. A feature selection algorithm using fuzzy boundary regions of the decision classes is presented. Both the lower and upper approximations are taken into consideration.

Consider an information system having $U = \{x_1, x_2, \ldots, x_n\}$, $C = \{a_1, a_2, \ldots, a_m\}$ and $D = \{d_1, d_2, \ldots, d_k\}$. Assume that all the conditional attributes are fuzzy. The real valued conditional attributes can be converted into fuzzy sets by the transformation $a^*(x) = \frac{x-p}{q-p}$, where $p = min_{x\in\, U} a_i(x)$ and $a = max_{x\in\, U} a_i(x)$.

For crisp decision attributes, the characteristic functions of the equivalence classes serve as the membership functions of the decision classes. In [12], a feature selection method using the divergence based fuzzy positive region is presented. The fuzzy positive region may be considered as an expression of the certainty of the membership of an object to a given class. Meanwhile, the boundary region gives information regarding the uncertainty of the membership of an object to a concept. This information is also used to select relevant features in an information system.

Definition4.1: The *divergence based fuzzy boundary region* of a fuzzy set X on U with respect to the attribute subset P of C is defined as

$$\mu_{BND(X)}(x_i) = \mu_{\overline{\delta}(A)}(x_i) - \mu_{\underline{\delta}(A)}(x_i). \quad (9)$$

Definition4.2: The *uncertainty degree* of a fuzzy set X on U with respect to P is given by

$$\eta_P(X) = \frac{\sum_{i=1}^{n} \mu_{BND(X)}(x_i)}{n}. \quad (10)$$

Definition4.3: The *total uncertainty degree* of the decision classes in a decision system with respect to P is given by $\rho_P(D) = \sum_{X \in U/D} \eta_P(X)$ (11)

Algorithm4.4: The algorithm for finding the total uncertainty degree of D with respect to $P \subseteq C$,

1. Input the decision table and $P \subseteq C$
2. Input $\mathcal{J}$ and $\mathcal{T}$.
3. Find the divergence matrix δ_P
4. Find the decision classes $U/D = \{X_1, X_2, \dots, X_r\}$
5. For $l = 1,2,..,r,\ i = 1,2,\dots,n$, find $\mu_{BND(X_l)}(x_i)$
6. Compute $\rho_P(D)$

Algorithm4.4: The following is the algorithm to find the set of features to be selected for the decision table reduction.

1. Input the fuzzy decision system
2. Initialise $C \leftarrow \{C_1, C_2, \dots, C_m\},\ R = \phi$
3. For each $a_i \in C - R$, generate the divergence matrix with respect to $R \cup \{a_i\}$.
4. Calculate the the total uncertainty degree of D, $\rho_{R\cup\{a_i\}}(D)$ for each each $a_i \in C - R$.
5. Find the attribute a_i that makes $\rho_{R\cup\{a_i\}}(D)$ the minimum.
6. When $\rho_{R\cup\{a_i\}}(D) < \rho_R(D)$, assign $C \leftarrow C - R\ and\ R \leftarrow R \cup \{a_i\}$.

In the first stage, the uncertainty value is computed n times, where n represents the number of conditional features in the data set. The feature with lowest uncertainty value is selected and the process is repeated for pairs of the selected feature with the remaining (n-1) features. In the worst case, this process is terminated when the whole feature set has been exhausted. Hence, the maximum possible number of computation of the total uncertainty value is $n + (n-1) + (n-2) + \dots + 1 = (n^2 + n)/2$. Thus, the maximum time complexity of the proposed algorithm is $o(n^2)$. Also, at the initial stage, n divergence matrices are computed and stored corresponding to each individual features. The space for all the subsequent matrices and local variables can be reused. So, the space complexity of the proposed algorithm is $o(n)$.

V. EXPERIMENTAL RESULTS

The results from the experimental study of the proposed algorithm using the divergence based fuzzy boundary region is presented in this section. Eleven data sets from the UCI Machine Learning repository[26] and one from the website of Milano Chemometrics and QSAR Research Group have been used for the experimentation. The data sets consist of objects ranging from 120 to 4898 and decision classes ranging from 2 to 34 and real valued features ranging from 5 to 166 in number. The description of the data sets is given in table 1.

Table 1. Data Set Description

Dataset	Objects	Features	Decision classes	Description
Olitos[27]	120	26	4	Chemical analysis
Sonar-mines/ rocks	208	61	2	Mine/rock recognition
Glass	214	10	7	Glass identification
Knowledge[28]	258	6	4	Knowledge level classification
Ionosphere	351	35	2	Structure analysis
Musk	476	166	2	Musk/non-musk classification
Energy efficiency[29]	768	10	2	Energy analysis
Plant leaves[30]	1600	65	34	Plant leaves identification
Steel plate faults	1941	28	7	Steel plates fault diagnosis
Segment	2310	20	7	Image segmentation
Statlog	4435	37	7	Land sat satellite data
Wine quality-white[31]	4898	12	7	Wine quality analysis

The uncertainty degree corresponding to each single attribute sets are computed first and the attribute with minimum uncertainty degree is selected. Then, pairs of the selected feature with the remaining features are considered and the pair having the minimum value of uncertainty degree is selected. This process is repeated unless there is no further increase in the dependency value. In the worst case, the process is terminated when the whole feature set has been exhausted. The feature selection process is implemented using a program in OCTAVE and the results are presented in table 2.

Table 2: Reduct size

Dataset	Objects	Features	Reduct size	Uncertainty degree
Olitos	120	26	17	0.716
Sonar-mines/rocks	208	61	31	0.659
Glass	214	10	7	0.671
Knowledge	258	6	5	0.518
Ionosphere	351	35	29	0.586
Musk	476	166	39	0.164
Energy efficiency	769	10	6	0.319
Plant leaves	1600	65	34	0.930
Steel plates faults	1941	28	15	0.549
Segment	2310	20	11	0.419
Statlog	4435	37	21	0.72
Wine quality-white	4898	12	9	0.912

The data presented in table 2 shows that the size of the selected feature set (reduct size) is significantly less than the original number of attributes in almost all the cases. The algorithm converges even for data sets consisting of around 5000 objects.

VI. CONCLUSION

Divergence measures are fuzzy measures that express the extent of dissimilarity between fuzzy sets. In this paper, the generalized fuzzy rough lower and upper approximations of a fuzzy set in a fuzzy information system based on divergence measure have been defined and their properties were investigated. Also, an algorithm for feature selection using the fuzzy boundary region has been presented. The proposed feature selection method was implemented with twelve real world data sets using an OCTAVE program. The data sets consisted of objects ranging from 120 to 4898 and decision classes ranging from 2 to 34 and real valued features ranging from 5 to 166 in number. The experimental results showed that the number of features in almost all the data sets considered was considerably reduced by applying the proposed algorithm. The algorithm converged even for data sets containing approximately 5000 objects. Future work includes a comparison of the methods using different divergence measures and different fuzzy logical operators.

REFERENCES

[1] S. Sumathi and S.N. Sivanandam, "Introduction to Data Mining and its Applications", Springer, Berlin, 2006.

[2] Sonali Suskar, S. D. Babar , "*Survey on Feature Selection for Text Categorization*", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, Vol.3, Issue.4, pp.261-266, 2018.

[3] P. Rutravigneshwaran , "*A Study of Intrusion Detection System using Efficient Data Mining Techniques*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.6, pp.5-8, 2017.

[4] L. A. Zadeh, "*Fuzzy Sets*", Information and Control, Vol.8, Issue.3, pp.338-353, 1965.

[5] Z. Pawlak, "*Rough Sets*", International Journal of Computer and Information Sciences, Vol.11, Issue.5, pp.341 - 356, 1982.

[6] S. Vluymans, L. D'eer, Y. Saeys, C. Cornelis, "*Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey*", Fundamenta Informaticae, pp.1-34, 2015.

[7] D. Dubois, H. Prade, "*Rough Fuzzy Sets and Fuzzy Rough Sets"*, International Journal of General Systems, Vol.17, pp.191-209, 1990.

[8] J. S. Mi, Y. Leung and T. Feng, "*Generalized Fuzzy Rough Sets Determined by a Triangular Norm*", Information Sciences, Vol.178, pp.3203-3213, 2008.

[9] N. M. Parthalain, R. Jensen and Q. Shen, "Fuzzy entropy assisted fuzzy-rough feature selection", in Proceedings of 2006 IEEE International Conference on Fuzzy Systems, pp.1499-1506, 2006.

[10] A. Radzikowska and E. E. Kerre, "*A comparative Study of Fuzzy Rough Sets*", Fuzzy Sets and Systems, Vol.126, pp.137-155, 2002.

[11] W. Z. Wu, J. S. Mi and W. X. Zhang, "*Generalized Fuzzy Rough Sets"*, Information Sciences, Vol.151, pp.263-282, 2003.

[12] Sheeja T. K. and A. Sunny Kuriakose, "*A novel feature selection method using fuzzy rough sets*", Computers in Industry, Vol.97, pp.111 - 116, 2018.

[13] J. Klir and B. Yuan, "Fuzzy Sets and Fuzzy Logic", Prentice Hall, New Jersey,1995.

[14] Z. Pawlak, "Rough Sets - Theoretical Aspect of Reasoning About Data*"*, Kluwer Academic Publishers, The Netherlands, 1991.

[15] S. Montes, I. Couso, P. Gil and C. Bertoluzza, "*Divergence Measure Between Fuzzy Sets*", International Journal of Approximate Reasoning, Vol.30, pp.91-105, 2002.

[16] S. An, Q. Hu and D. Yu, "*A Robust Rough Set Model Based on Minimum Enclosing Ball", In* Rough Sets and Knowledge Discovery, LNAI 6401 eds J. Yu et al., pp.102-109, Springer - Verlag, Berlin 2010.

[17] C. Chornelis, N. Verdiest and R. Jensen, "*Ordered Weighted Average Based Fuzzy Rough Sets*", In Rough Sets and Knowledge Discovery, LNAI 6401, eds J. Yu et al., pp.78-85, (Springer - Verlag, Berlin 2010.

[18] Q. Hu, D. Chen, D. Yu and W. Pedrycz, "*Kernelized Fuzzy Rough Sets*", In Rough Sets and Knowledge Discovery, LNCS 5589, eds P. Wen et al. pp.304-311, 2009 (Springer-Verlag, Berlin).

[19] L. Deer, N. Verbiest, C. Cornelis and L. Godo, "*A comprehensive study of implicator-conjunctor-based and noise-tolerant fuzzy rough sets: Definitions, properties and robustness analysis"*, Fuzzy Sets and Systems, Vol.275, pp.1-38, 2015.

[20] R. Jensen and Q. Shen, "*Fuzzy-rough attribute reduction with application to web categorization*", Fuzzy Sets and Systems, Vol.141, issue.3, pp.469-485, 2004.

[21] R. Jensen and Q. Shen, "*Fuzzy-rough data reduction with ant colony optimization*", Fuzzy Sets and Systems, Vol.149, Issue.1, pp.5- 20, 2005.

[22] R. Jensen and Q. Shen, "*New approaches to fuzzy-rough feature selection*", IEEE Transactions on Fuzzy Systems, Vol.17, Issue.4, pp.824-838, 2009.

[23] R. Jensen and Q. Shen, "*Fuzzy-rough sets assisted feature selection*", IEEE Transactions on Fuzzy Systems, Vol.15, Issue.1, pp.73 - 89, 2007.

[24] C. C. Eric, D. Chen, D.S.Yeung, X. Z. Wang and W.T. John, "*Attributes reduction using fuzzy rough sets*", IEEE Transactions on Fuzzy Systems, Vol.16, Issue.5, pp.1130-1141, 2008.

[25] N. M.Parthalain, Q. Shen and R. Jensen, "*A distance measure approach to exploring the rough set boundary region for attribute reduction*", IEEE Transactions on Knowledge and Data Engineering, Vol.22, Issue.3, pp.305-317, 2010.

[26] M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science 2013.

[27] C. Armanino, R. Leardi, S. Lanteri and G. Modi, "Chemometrics and Intelligent Laboratory Systems" Vol.5, pp.343-354, 1989.

[28] H. T. Kahraman, S. Sagiroglu and I.Colak, "*Developing intuitive knowledge classifier and modeling of users' domain dependent data in web*", Knowledge Based Systems, Vol.37, pp.283-295, 2013.

[29] A. Tsanas and A. Xifara, "*Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools*", Energy and Buildings, Vol.49, pp.560-567, 2012.

[30] C. Mallah, J. Cope and J. Orwell, "*Plant Leaf Classification Using Probabilistic Integration of Shape, Texture and Margin Features*", Signal Processing, Pattern Recognition and Applications, 2013.

[31] P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis, "*Modeling wine preferences by data mining from physicochemical properties*", Decision Support Systems, Elsevier, Vol.47, Issue.4, pp.547-553, 2009.

Authors Profile

Mrs. Sheeja T. K. obtained her MSc degree in Mathematics from Mahatma Gandhi University, Kottayam in 2000. She has fifteen years of teaching experience. She is presently working as an Assistant professor under the Collegiate Education Department of Govt. of Kerala and pursuing Ph.D. under the supervision and guidance of Dr. Sunny Kuriakose A.

Dr. Sunny Kuriakose A. was awrded PhD degree by Cochin University of Science and Technology, Kerala, India in 1995. He has more than three decades of teaching experience. He is currently serving as the Dean and a Professor at Federal Institute of Science and technology, Angamaly, Kerala, India. Fourteen scholars have been awarded Ph.D degree under his supervision. He has published more than sixty research papers in various national and international journals. He authored two books and edited a number of volumes. His research interest includes Fuzzy Logic, Graph Theory, Decision Theory etc. He served the Kerala Mathematical Association as its General Secretary for eight years. Presently, he is the Academic Secretary of the Association.

