Probabilistic Rough indices in Information Systems under Intuitionistic Fuzziness

D.Rekha(1) , K.Thangadurai(2)
(1) Department of Computer Applications, St.Peter's University, Chennai, India
(2)Department of Computer Science, Government College, Karur, Tamilnadu, India

Abstract - The concept of classifying the records of the information system has been due to Two Way Approach [ie, lower and upper approximations] of Pawlak's rough sets model. But the approximation does to take into consideration the degree of contribution of the basic categories. This deficiency was eliminated in early nineties by Ziarko who has proposed VPRS model and later on various efforts were made in defining a new Probabilistic Rough Set Model. In 2004, G.Ganesan et.al., have introduced the concept of classifying the records of the information system with fuzzy decision attributes using a threshold. Later, G. Ganesan extended this algorithm for any information system with intuitionistic fuzzy decision attributes. In this paper, we extended the work of G.Ganesan et.al., for the Probabilistic Rough Set Model to improve the efficiency of rough indices in the information system with intuitionistic fuzzy decision attributes.

Keywords: *information system, rough set, probabilistic rough set, rough index.*

*****

1. INTRODUCTION

In 1982, Z. Pawlak defined a mathematical model namely Rough Sets [6,7] which has applications in several areas, including the fields of knowledge acquisition and discovery, information retrieval etc. According to this theory, using either of the two ways of performing union of Basic categories, a given input or concept can be approximated. This limitation has been eliminated by Ziarko in 1993. Later this model has further been extended by Bing Zhou, YY Yao, Slezak and others.

Later, in 2005, G. Ganesan et, al., discussed the importance of defining the threshold [3] in computing the rough fuzziness. In 2008, G.Ganesan et.al., introduced the concept of rough classification [4] in information systems using these threshold in any information systems with fuzzy decision attributes. The initial approach of Naïve Bayesian Rough Set Model [10] was discussed in [9] by Slezak. Recently, Yiyu Yao and Bing Zhou discussed the same approach. In this paper, we extended the work of G.Ganesan et. al., on rough indexing to the information systems with intuitionistic fuzzy [1] decision attributes and functions with Probabilistic Naïve Bayesian Rough Set Model.

2. DECISION THEORETIC AND PROBABILISTIC ROUGH SETS

Pawlak's rough sets theory defines two way approximations [6,7] namely lower and upper approximations for a given input. For a given finite universe of discourse U and an equivalence relation E, we define the equivalence class of any $x \in U$ to
is a partition of the universe U.

For a given concept C, Pawlak defined the lower approximation as $\underline{apr}_E(C)=\{x \in U / [x]_E \subseteq C\}$ and upper approximation as $\overline{apr}_E(C)=\{x \in U / [x]_E \cap C \neq \Phi\}$.

Some of the researchers quote this model using Three Way Approach namely positive, negative and boundary regions which are defined as follows:

Positive Region: $POS_E(C)=\{x \in U / [x]_E \subseteq C\}$

Boundary: $BND_E(C)=\{x \in U / [x]_E \cap C \neq \Phi \wedge [x]_E \not\subseteq C\}$

Negative region: $NEG_E(C)=\{x \in U / [x]_E \cap C = \Phi\}$

Since Pawlak's model is restrictive, several researchers focused on generalizing the approach towards parameterized rough set model, probabilistic rough set model, Variable Precision rough set model and generalized rough set model.

In 1994, Pawlak and Skowron [8] defined rough membership function by considering degrees of overlap between equivalence classes and a concept C to be approximated and is viewed as the conditional probability of an object belongs to C given that the object is in [x] (for simplicity, we denote $[x]_E$ with [x]) which is given as $Pr\left(C/[x]\right)=\frac{|C \cap [x]|}{|[x]|}$

Using the definition quoted above, in [10], the positive, boundary and negative regions are defined as follows:

$$POS(C)=\left\{x \in U/ Pr\left(C/[x]\right)=1\right\}$$

$$BND(C)=\left\{x \in U/ 0 < Pr\left(C/[x]\right)<1\right\}$$

$$NEG(C)=\left\{x \in U/ Pr\left(C/[x]\right)=0\right\}$$

In 2009, Greco et.al [5] discussed the parameterized rough set model by generalizing the above said definitions. In this model, two thresholds namely α and β are used to define the probabilistic regions and the positive, boundary and negative regions are modified as follows:

$$POS(\alpha,\beta)\,(C) = \left\{ x \in {}^{U}\!\Big/ Pr\left({}^{C}\!/_{[x]} \right) \geq \alpha \right\}$$

$$BND(\alpha,\beta)(C) = \left\{ x\epsilon\, {}^{U}\!\Big/ \beta < Pr\left({}^{C}\!/_{[x]} \right) < \alpha \right\}$$

$$NEG(\alpha,\beta)(C) = \left\{ x\epsilon\, {}^{U}\!\Big/ Pr\left({}^{C}\!/_{[x]} \right) \leq \beta \right\}$$

These Probabilistic regions will lead three way decisions namely acceptance, deferment and rejection respectively for any object x in U. But, however, in several cases, it is easy to compute the probability of the existence of a category [x] for a given concept C using $Pr\left({}^{[x]}\!/_{C} \right) = \frac{|[x] \cap C|}{|C|}$

Hence, by Baye's Theorem, the Positive, Boundary and Negative Regions are given by

$$POS^{B}_{(\alpha',\beta')}(C) = \left\{ x\epsilon\, {}^{U}\!\Bigg/ \log \frac{Pr\left({}^{[x]}\!/_{C} \right)}{Pr\left({}^{[x]}\!/_{C^c} \right)} \geq \alpha' \right\}$$

$$BND^{B}_{(\alpha',\beta')}(C) = \left\{ x\epsilon\, {}^{U}\!\Bigg/ \beta' < \log \frac{Pr\left({}^{[x]}\!/_{C} \right)}{Pr\left(\left({}^{[x]}\!/_{C^c} \right) \right)} < \alpha' \right\}$$

$$NEG^{B}_{(\alpha',\beta')}(C) = \left\{ x\epsilon\, {}^{U}\!\Bigg/ \log \frac{Pr\left({}^{[x]}\!/_{C} \right)}{Pr\left({}^{[x]}\!/_{C^c} \right)} \leq \beta' \right\}$$

$$where \quad \alpha' = \log \frac{\Pr(C^C)}{\Pr(C)} + \log \frac{\alpha}{1-\alpha}$$

$$and \quad \beta' = \log \frac{\Pr(C^c)}{\Pr(C)} + \log \frac{\beta}{1-\beta}$$

Now, Now, we shall discuss the conventional approach on dealing the intuitionistic fuzzy sets to approximate under rough computing, which was discussed in [2]

3. ANALYSIS OF INTUITIONISTIC FUZZY SET USING A THRESHOLD

Consider a set D, called **IFR-domain** [2], satisfying the following properties:

a) $D \subset (0,1)$

b) If a intuitionistic fuzzy concept C is under computation, eliminate the values $\mu_C{}^c(x)$ and $\eta_C{}^c(x)$ $\forall\, x \in U$ from the domain D, if they exist, where μ_C and η_C represent the corresponding membership and non-membership grades respectively in the set C with the property that $0 \leq \mu_C + \eta_C \leq 1$.

c) After the computation using C, the values removed in (b) may be included in D provided C must not involve in further computation

Consider the universe of discourse $U=\{x_1,x_2,\ldots,x_n\}$. Let $\alpha,\alpha_1,\alpha_2,\beta,\beta_1$ and β_2 be the thresholds assume one of the values from the domain D, where D is constructed using the intuitionistic fuzzy concepts A and B.

For a given thresholds and with + <1 and a intuitionistic fuzzy set A, the Strong (α,β)-Cut is given by $A[\alpha, \beta] = \{x \in U \,/\, \mu_A(x) > \alpha$ *and* $\eta_A(x) < \beta\}$

Further, in this paper, we use the concept of hedges [11] which were introduced by Zadeh [12] with a modification for including non membership grade.

Hedges are commonly used in fuzzy logic to emphasize the grade of membership of any argument. Here, we use the hedge to both membership and non membership values.

For example, for the linguistic variable 'low' with the membership function α and non membership β, the hedges 'very' and 'very very' emphasis the efficiency of the variable with the corresponding membership values α^2 and α^4 respectively and the non membership values $\beta^{1/2}$ and $\beta^{1/4}$ respectively. They are called **concentration**, whereas the hedges 'slightly' and 'more slightly' dilutes the efficiency of the linguistic variables with the corresponding membership values $\alpha^{1/2}$ and $\alpha^{1/4}$ and the non membership values β^2 and β^4. They are called **dilation**.

3.1. Rough Approximations on intuitionistic fuzzy sets using α, β

Let Ψ be any partition of U, say $\{B_1,B_2,\ldots,B_t\}$. For the given intuitionistic fuzzy concept, the lower and upper approximations with respect to α can be defined as

$_{(\alpha,\beta)}C = (C[\alpha, \beta])$ and $^{(\alpha,\beta)}C = (C[\alpha, \beta])$ respectively.

4. NAÏVE BAYESIAN PROBABILISTIC ROUGH SETS MODEL FOR AN INUITIONISTIC FUZZY CONCEPT

Since, in the above both sections, the same thresholds α and β are used, hereafter, we use (δ,γ) cuts on intuitionistic fuzzy sets.

Hence, for a given intuitionistic fuzzy concept F with the thresholds δ and γ, the probabilistic positive, boundary and negative regions are respectively defined on the approximation space U/E as

$$POS_{(\delta,\gamma)}(F) = \left\{ x \in U \middle/ Pr\left(F[\delta,\gamma] / [x] \right) = 1 \right\}$$

$$BND_{(\delta,\gamma)}(F) = \left\{ x \in U \middle/ 0 < Pr\left(F[\delta,\gamma] / [x] \right) < 1 \right\}$$

$$NEG_{(\delta,\gamma)}(F) = \left\{ x \epsilon U \middle/ Pr\left(F[\delta,\gamma] / [x] \right) = 0 \right\}$$

For given parameters α and β, the regions of the parameterized rough sets model are given by

$$POS_{(\alpha,\beta,\delta,\gamma)}(F) = \left\{ x \epsilon U \middle/ Pr\left(F[\delta,\gamma] / [x] \right) \geq \alpha \right\}$$

$$BND_{(\alpha,\beta,\delta,\gamma)}(F) = \left\{ x \epsilon U \middle/ \beta < Pr\left(F[\delta,\gamma] / [x] \right) < \alpha \right\}$$

$$NEG_{(\alpha,\beta,\delta,\gamma)}(F) = \left\{ x \epsilon U \middle/ Pr\left(F[\delta,\gamma] / [x] \right) \leq \beta \right\}$$

and the regions of Naïve Bayesian Rough Sets Model are given by

$$POS^{B}_{(\alpha',\beta',\delta,\gamma)}(F) = \left\{ x \in U \middle/ log \frac{Pr\left([x] / F[\delta,\gamma] \right)}{Pr\left([x] / (F[\delta,\gamma])^{C} \right)} \geq \alpha' \right\}$$

$$BND^{B}_{(\alpha',\beta',\delta,\gamma)}(F) = \left\{ x \in U \middle/ \beta' < log \frac{Pr\left([x] / F[\delta,\gamma] \right)}{Pr\left([x] / (F[\delta,\gamma])^{C} \right)} < \alpha' \right\}$$

$$NEG^{B}_{(\alpha',\beta',\delta,\gamma)}(F) = \left\{ x \in U \middle/ log \frac{Pr\left([x] / F[\delta,\gamma] \right)}{Pr\left([x] / (F[\delta,\gamma])^{C} \right)} \leq \beta' \right\}$$

where $\alpha' = \log \frac{\Pr(C^{c})}{\Pr(C)} + \log \frac{\alpha}{1-\alpha}$

and $\beta' = \log \frac{\Pr(C^{c})}{\Pr(C)} + \log \frac{\beta}{1-\beta}$

5. ROUGH INDICES

Algorithm Naïve Bayesian_ rough index (x,A,α,β,δ,γ)
//Algorithm returns Naïve Bayesian_rough index of x

1. Let x_index be an integer initialized to 0
2. Pick the equivalence class K containing x.
3. If $\mu_A(y)=0$ and $\gamma_A(y)=1$ for all $y \in K$
   x_index=-M goto 7
4. If $\mu_A(x)=1$ and $\gamma_A(y)=0$ for all $y \in K$
   x_index=M goto 7
5. compute $POS^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$, and $NEG^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$
6. If $x \in POS^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$

Begin
 x_index=M
while ($x \in POS^{B}_{(\alpha',\beta',\delta,\gamma)}$
 (A)) begin
 δ=dil(δ) //dilation of δ
 γ=dil(γ) //dilation of γ
 x_index=x_index+1 compute $POS^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$
end
end
end
 else if $x \in NEG^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$ begin
 x_index=-M
while ($x \in NEG^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$) begin
 δ=con(δ) //concentration of δ
 γ=con(γ) //Concentration of γ
 x_index=x_index-1

compute $NEG^{B}_{(\alpha',\beta',\delta,\gamma)}(A)$
end
end
end
else

$a^2 + b^2 = c^2$; x1Let $\pi_1=\delta$; $\pi_2=\delta$; $\theta_1=\gamma$; $\theta_2=\gamma$
Let $\pi_1=\delta$; $\pi_2=\delta$; $\theta_1=\gamma$; $\theta_2=\gamma$
while $(x \notin (POS^{B}_{(\alpha',\beta',\pi 1,\theta 1)}(A) \cup NEG^{B}_{(\alpha',\beta',\pi 2,\theta 2)}(A)))$
begin
$\pi_1=Con(\pi_1)$; $\theta_1=Con(\theta_1)$;
$\pi_2=dil(\pi_2)$; $\theta_2=dil(\theta_2)$;
x_index=x_index+1 end
if $x \in POS^{B}_{(\alpha',\beta',\pi 1,\theta 1)}(A)$ x_index= -x_index
end
7. return x_index

6. NAÏVE BAYESIAN INDEXING IN INFORMATION SYSTEM WITH INTUITIONISTIC FUZZY DECISION ATTRIBUTES

Consider an information system given by T=(U, A, C, D), where U is the universe of discourse, A is a set of primitive attributes, C and D are the subsets of A called condition and decision features respectively [C and D may not exist in a few of the information systems].
Let $C=\{a_1,a_2,\dots,a_n\}$ and $D=\{d_1,d_2,\dots,d_s\}$ with the records $U=\{x_1,x_2,\dots,x_m\}$. For any index key 'a' in C, the indiscernibility relation is given by $x_i \approx_{ak} x_j$ (read as x_i is related to x_j with respect to a_k) if and only if $a_k(x_i)=a_k(x_j)$. Clearly, this indiscernibility relation partitions the universe of discourse U. However, the procedure of selecting the appropriate minimal attributes [reducts] for effectiveness is not discussed in this paper.
For example, consider the decision table with C={a,b,c,d} and D={E}.

	a	b	c	d	E
x_1	1	0	2	1	1
x_2	1	0	2	0	1
x_3	1	2	0	0	2
x_4	1	2	2	1	0
x_5	2	1	0	0	2
x_6	2	1	1	0	2
x_7	2	1	2	1	1

Let us consider the index key as 'c'. As x_1,x_2,x_4,x_7 have the values 2; x_3,x_5 have the values 0 and x_6 has the value 1. Hence, the partition on U with respect to c can be defined as $\{\{x_1,x_2,x_4,x_7\},\{x_3,x_5\},\{x_6\}\}$. However, in real time systems we can find several information systems with fuzzy or intuitionistic fuzzy decision attributes. The Naïve Bayesian rough indices algorithm discussed earlier can be applied for any information system with intuitionistic fuzzy decision attributes.

7. EXPERIMENTAL ANALYSIS

Consider data table with C={a,b,c,d} and D={E} where E is of intuitionistic fuzzy natured.

	a	b	c	d	$\mu_E(x_i)$	$\eta_E(x_i)$
x_1	1	0	2	1	0.45	0.54
x_2	1	0	2	0	0.7	0.2
x_3	1	2	0	0	0.65	0.3
x_4	1	2	2	1	0.1	0.6
x_5	2	1	0	0	0.91	0.03
x_6	2	1	1	0	0.6	0.31
x_7	2	1	2	1	0.35	0.6

On considering 'c' as the index key, the partition obtained is $\{\{x_1,x_2,x_4,x_7\},\{x_3,x_5\},\{x_6\}\}$. Let $\delta=0.5$ and $\gamma=0.39$. Here, $E[\delta,\gamma]=\{x_2,x_3,x_5,x_6\}$.
Let $\alpha'=0.999$ and $\beta'=0$. Choose the element x_1.
Here, $x_1 \in POS(E)$ where POS represents the Positive region as quoted in the algorithm and hence, initially x_index will be assigned M. On applying the dilation on δ and γ, we obtain $\delta=0.7071$ and $\gamma=0.1541$, and hence $E[\delta,\gamma]=\{x_5\}$. As $x_1 \notin POS(E)$, the algorithm returns the index of x_1 to be M+1

8. CONCLUSION

In this paper, by using the concept of Naïve Bayesian rough sets the approach of indexing the records of the information system is dealt. These rough indices are useful to analyze and index a database when the intuitionistic fuzzy information about the entire key values is obtained.

REFERENCES

[2] Atanassov KT, Intuitionistic Fuzzy Sets, Fuzzy Sets and Systems 20, pp:87-96, 1986
[3] G.Ganesan et.al., Analysis of Intuitionistic Fuzzy Sets using Thresholds-An Approach of Rough Sets, International Journal of Computing and Mathematical Applications, Vol: 2, pp: 263-270, 2008
[4] G.Ganesan, C.Raghavendra Rao, Rough Set: Analysis of Fuzzy Sets using thresholds, UGC Sponsored National Conference on Recent Tends in Computational Mathematics, March 2004, Gandhigram Rural Institute, Tamilnadu, by Narosa Publishers, pp:81-87, 2005
[5] G.Ganesan, D.Latha, C.Raghavendra Rao, Rough Index in Information System with Fuzziness in Decision Attributes, International journal of Fuzzy Mathematics, Vol: 16, No: 4, 2008
[6] Greco, S., Matarazzo, B. and Slowinski, R. Parameterized rough set model using rough membership and Bayesian con_rmation measures, International Journal of Approximate Reasoning, 49, 285-300, 2009

[7] Pawlak, Z. Rough sets, International Journal of Computer and Information Sciences, 11, 341-356, 1982.
[8] Pawlak, Z. Rough Sets, Theoretical Aspects of Reasoning about Data, Dordrecht: Kluwer Academic Publishers, 1991.
[9] Pawlak, Z. and Skowron, A. Rough membership functions, in: Yager, R.R., Fedrizzi, M. and Kacprzyk, J., Eds., Advances in the Dempster-Shafer Theory of Evidence,John Wiley and Sons, New York, 251-271, 1994.
[10] Slezak, D. and Ziarko, W. The investigation of the Bayesian rough set model, International Journal of Approximate Reasoning, 40, 81-91, 2005.
[11] Yiyu Yao and Bing Zhou, Naive Bayesian Rough Sets, Proceedings of RSKT 2010, LNAI 6401, pp. 719-726, 2010.
[12] Kohavi R, Useful feature subsets and Rough set reducts, Proceedings, Third International Workshop on Rough Set and Soft Computing, pp:310-317, 1994
[13] Zadeh L.A. Fuzzy Sets, Journal of Information and Control, 8, pp. 338-353, 1965