A Practical Solution for the Classification in Interval-Valued Information Systems

AMAURY CABALLERO, KANG YEN, YECHANG FANG
Department of Electrical & Computer Engineering
Florida International University
10555 W. Flagler Street, Miami, Florida 33174
USA
caballer@fiu.edu, yenk@fiu.edu, yfang003@fiu.edu

Abstract.- Complex information systems have to manage enormous quantity of data, representing different measurements or attributes from various sources. Classification of this set of data is a difficult task that frequently requires special methods for being solved. The problem of classification in information systems has been studied by many authors, and different methods have been developed. The use of rough sets, fuzzy logic, neural networks, entropy, or the combination of these methods has been used widely. When information is diffuse and the number of obtained values for each attribute is large, so is the number of rules obtained for the classification. Even worst is hidden information in the data that makes the process complicated. Due to this fact, an interval of values is defined for each attribute, moving from the minimum to the maximum obtained values in the collected database. This is what is defined as interval-valued information systems. Differently from other works, the concept of information measure is used in this paper, together with a fuzzy logic discrimination tool. Using these concepts, initially an attribute reduction is obtained and then fuzzy logic is applied for discriminating among the possible solutions. This method is simpler than others, and provides accuracy not less than the usually employed methods.

Key Words.- Fuzzy Logic, Information Systems, Classification, Databases, Diffuse Information.

1 Introduction

When dealing with noisy signals or databases which involve variations of parameters A_k in the same object or class, it becomes mandatory to define intervals of variations for each parameter. As indicated by Y. Leung et al. [1], "This interval-valued attribute set, obtained by statistical methods may have non-empty intersection for distinct classes in the universe of discourse." For an attribute k, where l_i^k and u_i^k represent the minimum and maximum values for class i, and l_j^k and u_j^k are the corresponding values for class j, respectively. The region of coincidence or misclassification rate of the two attributes may vary from 0 when there is not coincidence at all, to 1 when the two attributes coincide completely.

In general, the probability that objects in class u_i be misclassified into class u_j according to attribute k has been modified from [1] and given by

$$\alpha_{ij}^k \equiv \alpha \quad if\, [l_i^k, u_i^k] \cap [l_j^k, u_j^k] = 0 \tag{1}$$

and

$$\alpha_{ij}^k = min\{(u_i^k - l_j^k, u_j^k - l_i^k)/(u_i^k - l_i^k)\},1\} \quad if\, [l_i^k, u_i^k] \cap [l_j^k, u_j^k] \neq 0 \tag{2}$$

where α is the permissible misclassification rate and α_{ij}^k is the probability that objects in class U_i are misclassified into class U_j for the attribute A_k. Note that in general $\alpha_{ij}^k \neq \alpha_{ji}^k$.

From the previous result, the maximum mutual classification error between classes u_i and u_j for attribute k can be defined as

$$\beta_{ij}^k = max\, \{\alpha_{ij}^k, \alpha_{ji}^k\} \tag{3}$$

where $\beta_{ij}^k = \beta_{ji}^k$.

It is logical to think that when these two classes coincide for some parameter k, the obtained information from this parameter for discriminating between classes i and j is 0, and this information increases as the coincidence diminishes. Using the definition of Shannon and Hartley, this information is expressed in terms of a logarithmic scale of based 10 by

$$I_{ij}^{k} = -(\log \beta_{ij}^{k}) \quad \text{[Hartley]} \qquad (4)$$

Similarly, the minimum information required for this classification between classes i and j for the attribute k is given by

$$I_{\alpha}^{k} = L = -\log \alpha \ \text{[Hartley]} \qquad (5)$$

where α is the permissible misclassification error between classes for any attribute. This value shall be defined from the beginning of describe a fuzzy set.

The major tasks of fuzzy-based patter classification are the extraction of knowledge from numerical data and construction of a rule base, which will permit the classification of new data members. One way of calculating the similarity is given below [2]:

Let $P^*(X)$ be a group of fuzzy sets with $A_i \neq 0$, and $A_i \neq X$. Defining two fuzzy sets from this family of sets, $A, B \in P^*(X)$, the expression

$$(A,B) = (A \bullet B) \wedge (\overline{A \oplus B}) \qquad (6)$$

describes the degree of similarity between these two sets A and B. When the approaching degree reaches the value of unity, these two sets will have higher degree of similarity and it is very difficult to differentiate one from another. On the other hand,

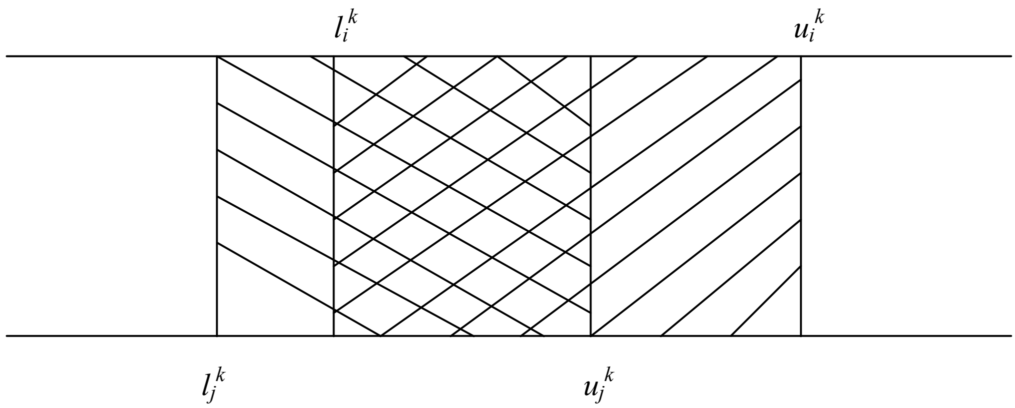

Figure1. Region of Coincidence of Classes i and j for the Attribute k

the classification, and should be larger than zero. If $I_{ij}^{k} \geq I_{\alpha}^{k}$, then these two classes can be differentiated by using the attribute k.

2 Fuzzy Sets for Pattern Recognition

Fuzzy logic, a multi-valued Boolean logic, helps describing concepts that are commonly encountered in the real world, by using linguistic variables. One of the basic concepts in fuzzy logic is the membership function. In general, any function $A: X \rightarrow [0, 1]$ can be used as a membership function to

if its value is closed to zero, it implies that the distinction between these two fuzzy sets has higher possibility.

For two Gaussian membership functions, with means a and b and standard deviations σ_a and σ_b, the previous expression becomes

$$(A,B) = \exp\left[\frac{-(a-b)^2}{(\sigma_a + \sigma_b)^2}\right] \wedge 1 \qquad (7)$$

In the case of multi-feature pattern recognition, many methods have been proposed by researchers. As indicated by Ross [3] the three popular approaches are:

(1) nearest neighbor classifier,
(2) nearest center classifier, and
(3) weighted approaching degree.

The first two methods are restricted to the recognition of crisp singleton data samples. The third method is the one used in this work.

Defining a new data sample with m fuzzy attributes, the approaching degree concept can be applied to compare the new data pattern

$$B = \{B_1, B_2, \ldots, B_m\}$$

with some known data pattern. Each of the known patterns A_i is characterized by the same m attributes and given by

$$A_i = \{A_{i1}, A_{i2}, \ldots, A_{im}\}$$

where $i = 1, 2, \ldots, k$ describes k-patterns. For each of the known k-patterns, the approaching degree expression is given by

$$(B, A_i) = \sum_{j=1}^{m} \omega_j (B_j, A_{ij}) \qquad (8)$$

where ω_j is a normalizing weighting factor, taken unitary in this work.

Then sample B is closest to sample A_j if

$$(B, A_j) = \max_{1 \le i \le k} \{(B, A_i)\} \qquad (9)$$

The collection of fuzzy sets

$$B = \{B_1, B_2, \ldots, B_m\}$$

can be reduced to a collection of crisp singletons, $B = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ where each sample ($\mathbf{x}_i$) is a vector of features,

$$\mathbf{x}_i = \{x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}\}$$

and the above mentioned weighted equation can be expressed as

$$\mu_{A_i}(\mathbf{x}) = \sum_{j=1}^{m} \omega_j \cdot \mu_{A_{ij}}(\mathbf{x}_j) \qquad (10)$$

In the maximum approaching degree, sample singleton $\mathbf{x}$, is closest to pattern A_j when

$$\mu_A(\mathbf{x}) = \max_{1 \le i \le k} \{\mu_{A_i}(\mathbf{x})\} \qquad (11)$$

3 Method for Classification of Diffuse Datasets

The proposed method, which differs from other methods mentioned above, uses the quantity of information obtained from the coincidence of two classes. From here, it is possible to find the minimal number of parameters that permit the classes definition. Using the minimum number of attributes and the defined rules, it is possible to classify any new obtained data in each case. This minimum number of attributes is defined as the reduct in rough sets theory [3, 4, 5].

Several authors solve the problem combining neural networks and fuzzy logic [6], or using the concept of information entropy and mutual information [7]. Other works present the use of rough sets and fuzzy logic together [8, 9,10] or simply fuzzy logic [11]. In the paper four databases are presented for testing the proposed method. The minimization of attribute and the rules extraction have been performed on each of them using the proposed algorithm.

3.1 Example with Interval-Valued Information System I

The method is introduced using the database from Table 1. In each cell has been written the minimum and maximum accepted values for each parameter and each class [12]. Here $U_1, \ldots, U_n$ are the different classes and $A_1, \ldots, A_m$ are the attributes.

Table 1. Interval-valued Information System

	A_1	A_2	A_3	A_4	A_5
U_1	4.1; 4.9	5.4; 6.8	2.2; 3.6	3.5; 4.8	4.8; 5.6
U_2	4.5; 5.2	4.9; 5.4	3.3; 4.8	2.8; 4.1	4.0; 4.7
U_3	2.1; 2.8	3.2; 4.1	2.1; 4.2	3.5; 5.6	3.3; 5.4
U_4	3.1; 5.4	4.1; 5.8	1.9; 2.8	2.1; 4.1	2.4; 4.0
U_5	1.5; 2.0	2.9; 3.8	2.6; 3.9	1.4; 2.8	3.1; 4.8

Step 1: Select the value for α and calculate the value of L from $L = -\log_{10}\alpha$.

In this example, the value of α has been selected as 0.1 arbitrarily. So the corresponding value of L is 1.

Step 2: Calculate all the values of I_{ij}^k and create Table 2.

(1) If $I_{ij}^k \geq L$, replace the calculated value by the value of L. Then for each attribute add all the values shown in the corresponding column. Its sum is recorded in the same column of the last row recorded as Σ in the table.

(2) The total number of rows is given from

$$C_2^n = (n)(n-1)/2,$$

where n is the number of classes.

Table 2. Classification Error between Classes U_i and U_j

I_{ij}^k	A_1	A_2	A_3	A_4	A_5	$\sum_{k=1}^{m} I_{ij}^k$
U_{12}	0.2	0.77	0.68	0.34	1	
U_{13}	1	1	0	0	0.12	
U_{14}	0	0.54	0.17	0.34	1	
U_{15}	1	0.08	0.08	1	1	
U_{23}	1	1	0.22	0.34	0.11	
U_{24}	0	0	1	0	1	
U_{25}	1	1	0.3	0	0	
U_{34}	1	1	0.17	0.52	0.36	
U_{35}	1	0.17	0	1	0.06	
U_{45}	1	1	0.66	0.19	0.25	
Σ	1.2	6.56	3.28	2.73	4.9	

Table 2 shows that every row contains at least one attribute with the value $L = 1$. This means that all the classes can be uniquely discriminated by using proper attributes.

Step 3: If $\sum_{k=1}^{m} I_{ij}^k < L$ for some calculated rows, mark those rows.

This means that classes U_i and U_j cannot be differentiated by the selected attributes. This situation does not exist in the current example.

Step 4: Check each row containing only one L value and select the attribute corresponding to this value. If the same attribute is repeated in more than one row, select it only once. Then mark all the rows containing the selected attribute. If there are other rows with the value "L" go to Step 5. If not, move to Step 6.

In Table 2 this occurs with attribute A_5 only. Rows (U_{12}), (U_{14}), (U_{15}) and (U_{24}) are marked.

Step 5: Among the non-selected attributes, select the attribute (column) with the value L shown in the largest number of rows. If this number is the same for more than one attribute, then select the attribute with the largest values in row Σ. Mark all the non-marked rows containing the selected attribute. Repeat Step 5, until no row containing the value L is analyzed, then go to the following step. If all the rows are marked, the selection process is completed and the useful attributes are those selected.

In this step, the first selected attribute is A_1. This is due to the fact that for A_1 the value of L is shown in rows corresponding to U_{13}, U_{15}, U_{23}, U_{25}, U_{34}, U_{35}, and U_{45}. The total number of times shown in the table is 7, which is larger than other non-selected attributes. The selection of A_1 and A_5 causes all rows in Table 2 marked. So the selection process is completed and each class can be differentiated from others.

In the above example, a complete differentiation can be carried out with attributes A_1 and A_5. The rules developed from the above results for differentiation among different U_i are:

*R*ule # 1: IF $A_1 \in [4.1; 4.9]$ and $A_5 \in [4.8; 5.6]$, THEN it is U_1.
*R*ule # 2: IF $A_1 \in [4.5; 5.2]$ and $A_5 \in [4.0; 4.7]$, THEN it is U_2.
*R*ule # 3: IF $A_1 \in [2.1; 2.8]$ and $A_5 \in [3.3; 5.4]$, THEN it is U_3.
*R*ule # 4: IF $A_1 \in [3.1; 5.4]$ and $A_5 \in [2.4; 4.0]$, THEN it is U_4.

*R*ule # 5: IF $A_1 \in [1.5; 2.0]$ and $A_5 \in [3.1; 4.8]$ THEN it is U_5.

In order to have a complete classification, it is necessary to formulate one rule for each class.

In the following, another database will be utilized to show that not all the classes can be completely differentiated from the given set of attributes.

3.2 Example 2 with Iris Database

The original database was presented by R. Fisher [12], and has been adopted in several papers [10, 13]. Table 3 shows the results after a proper elaboration of the original database.

Table 3. Iris Database and Attributes for Each Class

Attribute	**Setosa**			
	x_{av}	σ	Min	Max
SL	5.0	0.35	4.3	5.8
SW	3.42	0.38	2.3	4.4
PL	1.45	0.11	1.0	1.9
PW	0.24	0.11	0.1	0.6
Attribute	**Versicolor**			
	x_{av}	σ	Min	Max
SL	5.94	0.52	4.9	7.0
SW	2.77	0.31	2.0	3.4
PL	4.26	0.47	3.0	5.1
PW	1.33	0.20	1.0	1.8
Attribute	**Virginica**			
	x_{av}	σ	Min	Max
SL	6.59	0.64	4.9	7.9
SW	2.91	0.32	2.2	3.8
PL	5.55	0.55	4.5	6.9
PW	2.03	0.27	1.4	2.5

The three classes to be differentiated are U_1–Setosa, U_2–Versicolor and U_3–Virginica and the parameters taken into consideration are: A_1: SL-sepal length, A_2: SW-sepal width, A_3: PL-petal length, and A_4: PW-petal width. In this example the value of α has been selected as 0.2 and the corresponding value for L is 0.7 Hartley.

After completing Steps 1-5, rows U_{12} and U_{13} in Table 4 are marked and the selected attribute is A_3. This attribute permits to differentiate U_1 (Setosa) from the other two. From it, the following rule can be extracted:

*R*ule # 1: IF $A_3 \in [1; 1.9]$,
THEN it is U_1 – Setosa.

There are still non-marked rows, so let's move to the next step.

Table 4. Classification Error between Classes U_i and U_j, for Iris Database

I_{ij}^k	A_1	A_2	A_3	A_4	$\sum_{k=1}^{m} I_{ij}^k$
U_{12}	0.22	0.11	0.7	0.7	
U_{13}	0.22	0.03	0.7	0.7	
U_{23}	0.0	0.06	0.54	0.31	0.91
Σ	0.44	0.20	1.94	1.71	

Step 6: Select the value with maximum I_{ij}^k on the first non-marked row. If the corresponding attribute has not been selected, select it and look for the next maximum I_{ij}^k in the same row. Add them together and check whether the result satisfies $\Sigma I_{ij}^k \geq L$. If not, repeat the process with other attributes in the same row until the indicated condition is met. Mark this row. All the new selected attributes will be included in the final selection. Move to the next non-marked row until all the rows have been analyzed.

For this case, following Step 6, on the third row the attribute with maximum I_{ij}^k is A_3 ($0.54 < 0.7$). The next is A_4 ($0.31 < 0.7$), then we calculate $\Sigma I_{ij}^k = I_{23}^3 + I_{23}^4 = 0.85 > 0.7$. Since A_3 has already been selected, the only attribute selected in this Step is A_4. Different from the previous example, the discrimination between U_2 and U_3 cannot be uniquely defined due to the fact that each attribute does not provide enough information for the differentiation. Another rule for this case could be

*R*ule # 2: IF $A_3 \in [3.0, 6.]$ and $A_4 \in [1.0, 2.5]$, THEN it can be U_2–Versicolor or U_3–Virginica.

From the above rule the discrimination between Versicolor and Virginica is not possible at this point, so it is necessary to move to the next step.

From the above, rule the discrimination between Versicolor and Virginica is not possible at this point, so it will be necessary to move to the following step.

The same results can be obtained using rough sets [10]. For a given permissible misclassification rate $\alpha \in [0, 1]$ and an attribute subset $B \subseteq A$, a binary relation on U is defined by

$$R_B^{\alpha} = \{(u_i, u_j) \in U \times U \mid \beta_{ij}^k > \alpha, \forall a_k \in B\}$$

The errors that objects in class u_i being misclassified into class u_j in the system are defined as $\alpha_{ij} = min\{\alpha_{ij}^k : k \leq m\}$ and are given in Table 5.

Table 5. Misclassification Error of Object u_i into u_j

α_{ij}	U_1	U_2	U_3
U_1	1	0	0
U_2	0	1	0.25
U_3	0	0.29	1

The maximal mutual classification error between classes is defined by

$$\beta_{ij}^k = max\{\alpha_{ij}^k, \alpha_{ji}^k\}$$

In the present example is given by:

$\beta_{12}^1 = 0.6$ $\beta_{13}^1 = 0.6$ $\beta_{23}^1 = 1$
$\beta_{12}^2 = 0.78$ $\beta_{13}^2 = 0.94$ $\beta_{23}^2 = 0.86$
$\beta_{12}^3 = 0$ $\beta_{13}^3 = 0$ $\beta_{23}^3 = 0.29$
$\beta_{12}^4 = 0$ $\beta_{13}^4 = 0$ $\beta_{23}^4 = 0.5$

If $\alpha = 0.2$, the permissible misclassification rate for this example is shown in Table 6.

Table 6. Permissible Misclassification Rate between Classes u_i and u_j

β_{ij}	U_1	U_2	U_3
U_1	1	0	0
U_2	0	1	0.29
U_3	0	0.29	1

The matrix for the α-Tolerance relations can be fund as

$$R_A^{0.2} = \begin{bmatrix} 100 \\ 011 \\ 011 \end{bmatrix}$$

where all $\beta_{ij} > \alpha$ From this matrix, it is clear that object u_1-Setosa can be uniquely determined from the given attributes, but objects u_2-Versicolor and u_3-Virginica may not be separated. This situation can be expressed by

$$S_A^{0.2}(u_1) = \{u_1\}$$

and

$$S_A^{0.2}(u_2) = S_A^{0.2}(u_3) = \{u_2, u_3\}$$

where $S_A^{0.2}(u)$ denotes that these are the sets of objects which are possible indiscernible by A within u with the misclassification rate $\alpha = 0.2$. From the previous results, the 0.2-discernibility set is given 1n Table 7.

Table 7. Discernibility Set

	U_1	U_2	U_3
U_1			
U_2	a_3, a_4		
U_3	a_3, a_4		

The obtained functions are $f_1^{0.2} = a_3 \vee a_4$ and $f_2^{0.2} = a_3 \vee a_4$. Using rough sets, we have demonstrated that the important attributes for the classification are a_3: PL-petal length and a_4: PW-petal width.

From the above results, the following rules can be extracted:

$R(U_1)$: IF $a_3 \in [1, 1.9]$ or $a_4 \in [0.1, 0.6]$, THEN it is u_1 – Setosa.
$R(U_2)$: IF $a_3 \in [3.0, 5.1]$ or $a_4 \in [1.0, 1.8]$, THEN it can be u_2–Versicolor or u_3–Virginica.
$R(U_3)$: If $a_3 \in [4.5, 6.9]$ or $a_4 \in [1.4, 2.5]$, THEN it can be u_2 – Versicolor or u_3 – Virginica.

As can be noted, the results are equivalents using both methods.

Step 7: Apply Fuzzy Logic. Construct a membership function for each selected attribute for each class [14]. Evaluate the obtained attribute for each class using these

membership functions and applying equation (10).
For differentiating between versicolor and virginica, the table presented in [10] was used as the database. The first 40 rows in the original database are used as training instances and the last 10, the testing instances. Gaussian function has been used as the membership function of each attribute of the training instances, and the singleton value as the average value for each testing attribute. Table 8 presents the mean and standard deviation values for the training data, and Table 9, the mean value for the testing data

Table 8. Mean and Standard Deviation for 40 Training Data

Attributes	Versicolor		Virginica	
	x_{av}	σ	x_{av}	σ
SL	6.01	0.523	6.62	0.68
SW	2.78	0.33	2.96	0.34
PL	4.32	0.45	5.61	0.59
PW	1.35	0.21	1.99	0.27

Table 9. Mean Value for 10 Testing Data

Attributes	Versicolor	Virginica
	x_{av}	x_{av}
SL	5.64	6.45
SW	2.73	3.03
PL	4.03	5.33
PW	1.23	2.17

The approaching degree between the testing and training data is shown in Table 10. As can be seen from this table, for Versicolor testing data, the degree of similarity reaches the maximum of 0.9773 with Versicolor training instances, and for Virginica testing data the maximum of 0.9585 is obtained with Virginica testing instances. These results confirm that the method works properly.

Table 10. Approaching Degree between Testing and Training Data

Attributes	Testing Versicolor – Training Versicolor	Testing Versicolor – Training Virginica
SL	0.4614	0.1253
SW	0.9773	0.6328
PL	0.6601	0.0070
PW	0.7214	0.0040
	Testing Virginica – Training Virginica	**Testing Virginica – Training Versicolor**
SL	0.9394	0.4927
SW	0.9585	0.5633
PL	0.7983	0.0065
PW	0.6412	0

3.3 Example with Interval-Valued Information System II

The interval-valued information system presented by Y. Leung et al. is partially presented in Table 11, showing the first five classes only.

Table 11. Interval-Valued Information System II

	A_1	A_2	A_3	A_4	A_5
U_1	2.17; 2.86	2.45; 2.96	5.32; 7.23	3.21; 3.95	2.54; 3.12
U_2	3.37; 4.75	3.43; 4.85	7.24; 10.47	4.00; 5.77	3.24; 4.70
U_3	1.83; 2.70	1.78; 2.98	7.23; 10.27	2.96; 4.07	2.06; 2.79
U_4	1.35; 2.12	1.42; 2.09	2.59; 3.93	1.87; 2.62	1.67; 2.32
U_5	3.46; 5.35	3.37; 5.11	6.37; 10.28	3.76; 5.70	3.41; 5.28

As per Step 1, using $\alpha = 0.2$ ($I_\alpha = 0.7$), the results are shown in Table 12. In this table, from Step 2, it is obtained that classes U_2 and U_5 cannot be discriminated. Step 3 gives on the second row the attribute A_3, which is selected.

Table 12. Classification Error between Classes U_i and U_j

I_{ij}^k	A_1	A_2	A_3	A_4	A_5	$\sum_{k=1}^{m} I_{ij}^k$
U_{12}	0.7	0.7	0.7	0.7	0.7	
U_{13}	0.11	0	0.7	0	0.37	

U_{14}	0.7	0.7	0.7	0.7	0.7	

Applying Step 4, the first selected attribute is A_5. Rows U_{12}, U_{14}, U_{15}, U_{23}, U_{24}, U_{35}, and U_{45} result marked. The only non-marked row is U_{34}. The selected attributes are A_3 and A_5. The following rules could be proposed:

*R*ule # 1: IF $A_3 \in$ [5.32; 7.23] and $A_5 \in$ [2.54; 3.12], THEN it is U_1.
*R*ule # 1: IF $A_3 \in$ [5.32; 7.23] and $A_5 \in$ [2.54; 3.12], THEN it is U_1.
*R*ule # 2: IF $A_3 \in$ [7.23; 10.27] and $A_5 \in$ [2.06; 2.79], THEN it is U_3.
*R*ule # 3: IF $A_3 \in$ [2.59; 3.93] and $A_5 \in$

The above results show that it is not possible to discriminate between classes 2 and 5, because adding the information given by all the attributes, the obtained information is smaller than the necessary for the accepted mutual classification error.

3.4 Example with Abalone Database.

This database [15] has been created to predict the age of abalone from physical measurements. Here it is used only as an example, and the obtained conclusions do not necessarily reflect all the possibilities that could be obtained in the classification.

The authors randomly selected 150 entries from the database and divided them into three groups: Group I, from 6 to 10 years old, Group II, from 11 to 15 years old, and Group III, from16 to 20 years old. The attributes are as follows: *A*-Length, *B*-Diameter, *C*-Height, *D*-Whole Weight, *E*-Shucked Weight, *F*-Viscera Weight, and *G*-Shell Weight. he elaborated information is presented in Table 13.

Table 13. Attributes foe Each Class

Attribute	**Group I**			
	x_{av}	σ	Min	Max
A	0.46	0.09	0.27	0.59
B	0.37	0.06	0.2	0.47
C	0.12	0.02	0.07	0.18
D	0.54	0.24	0.12	1.06
E	0.22	0.37	0.05	0.49
F	0.12	0.06	0.03	0.27
G	0.17	0.07	0.07	0.34
Attribute	**Group II**			
	x_{av}	σ	Min	Max
A	0.57	0.06	0.38	0.72
B	0.46	0.05	0.31	1.06
C	0.16	0.03	0.10	0.24
D	1.04	0.37	0.28	2.55
E	0.41	0.15	0.11	1.07
F	0.23	0.08	0.07	0.54
G	0.32	0.12	0.10	0.76
Attribute	**Group III**			
	x_{av}	σ	Min	Max
A	0.62	0.05	0.54	0.74
B	0.50	0.04	0.42	0.58
C	0.19	0.02	0.16	0.24
D	1.43	0.41	0.94	2.50
E	0.53	0.25	0.35	0.93
F	0.29	0.08	0.17	0.49
G	0.47	0.15	0.27	0.78

Table 14 shows the maximum mutual classification error between classes in logarithmic form. It is clear that it is not possible to discriminate between groups II and III, using the given attributes ($\sum_{k=1}^{m} I_{ij}^{k} < I_{\alpha}$) and error of misclassification α = 0.2. It is possible to discriminate between groups I and III, using attributes *D* or *A*. Finally, there is a possibility of discriminating between Goups I and II, using fuzzy logic and practically all the attributes. But, in general as per the obtained results in the example, the conclusion could be that the selected attributes in this case are not useful for determining the abalone age, using intervals of 5 years. As per Table 14, the discrimination between all the selected classes can be done with an error of 0.9, which is in general not acceptable.

Table 14. Classification Error for Abalone Database

I_{ij}^{k}	U_{12}	U_{13}	U_{23}	Σ
A	0.18	0.7	0.05	0.23
B	0.23	0.51	0.0	0.74
C	0.14	0.6	0.0	0.74
D	0.08	0.7	0.0	0.08

E	0.07	0.06	0.0	0.13
F	0.07	0.07	0.0	0.14
G	0.05	0.58	0.04	0.68
$\sum_{k=1}^{m} I_{ij}^{k}$	0.82		0.09	

Conclusions

A method has been proposed for the attribute reduction and classification using the concept of information (or loss of information) together with fuzzy logic in interval-valued information systems. Several databases have been selected for introducing the method to discriminate between classes. The values of misclassification error selected by other researchers are also chosen in this study to representing in this paper to include the uncertainty in the differentiation between two classes. One of the advantages of the method is that from it we can easily determine how far the solution is for each established misclassification error, as well as finding the reduct in a simple way. Another advantage is that it is easily seen from the table whether it is possible to discriminate between any two classes or not.

Four examples have been solved by the proposed method. In the first one, it is necessary to use more than one attribute to differentiate between any two classes due to the fact that only one attribute is not enough to discriminate all classes, but the results are uniquely obtained applying steps 1 through 5.

The second example shows a situation, in where it is not possible uniquely to discriminate between those two classes. The is because in the third row of Table 4 we have(U_{23}), $I_{ij}^{k} < I_{\alpha}$, $\forall k$. And the application of steps 6 and 7 is mandatory.

In the third example, we show that it is not possible to discriminate between classes 2 and 5. This is due to by adding the information given by all the attributes, the obtained information is smaller than the necessary for the accepted mutual classification error.

The last database demonstrates that the discrimination between the second and third classes is not possible. However, Discrimination between classes U_1 and U_3 is possible by using attributes *D* or *A*. For the discrimination between classes I and II, it becomes necessary to include practically all the attributes and to use fuzzy logic. From this fact it is reasonably to conclude that it is not possible to discriminate the abalone age based on the given information in the database.

References

[1] Y. Leung, et Al., *A Rough Set Approach or the Discovery of Classification Rules in Interval-valued Information Systems*, Int'l J. of Approximate Reasoning, no. 47, 2007, pp. 233-246.

[2] T.J. Ross, *Fuzzy Logic with engineering Applications,* John Wiley and Sons, 2004

[3] Z. Pawlak, *Rough Set*, Int'l J. of Computer & Information Sciences. no. 11, 1982, pp.341-356.

[4] R. Jensen, *Combining Rough and Fuzzy Sets for Feature Selection*, PhD. Dissertation, U. of Edinburgh, 2005.

[5] Z. Pawlak, *Rough Sets*: *Theoretica Aspects of Reasoning About Data*, Dordrecht: Kluwer, 1991.

[6] E. Granger, et Al., *Classification of Incomplete Data* Using *the Fuzzy ARTMAP Neural Network*, Technical Report Boston U., Center for Adaptive Systems and Dept. of Adaptive Neural Systems, January 2000.

[7] T. Ying-Chieh, T. et Al., *Entropy-Based Fuzzy Rough Classification Approach for Extracting Classification Rules,* Expert Systems with Applications. no.31, 2006, pp.436-443.

[8] Martine De cock et al. *Fuzzy Rough Sets*: *The forgotten Step,* IEEE Transaction on Fuzzy Systems, Vol. 15, No 1, February 2007.

[9] Y. Y. Yao, *A Comparative Study of Fuzzy Sets and Rough Sets,* Information Sciences, Vol. 109, No. 1-4, pp.227-242, 1998.

[10] A. Caballero et al., *Classification with Diffuse or Incomplete Information,* WSEAS Transactions on systems and Control, Issue 6, Vol. 3, June 2008, pp.617-626.
[11] Caballero A., et Al. *Model for Clustering Objects under Specified Conditions,* WSEAS Transactions on Systems and Control, Issue 1, Vol. 2, January 2007. pp 86 – 91.
[12] Caballero A., K. Yen, Y. Fang, J. L. Abreu, *Method for Classification in Interval-Valued Information Systems. (Plenary Lecture).* Proceedings of the 12th WSEAS International Conference on Automatic control, Modeling and Simulation (ACMOS'10) May 29-31,2010, pp. 242-247, Catania, Italy
[13] R. Fisher, *The use of Multiple Measurements in Taxonomic Problem*, Annals of Eugenics Cambridge University Press, V. 7, pp.179-138, 1936.
[14] Shyi-Ming Chen and Yao-De Fang, *A New Approach for Handling the Iris Data Classification Problem,* International Journal of Applied Science and Engineering, 2005.3, 1:37-49.
[15] Kaleem F., et Al, *Intelligent Spatial Filters Based on Fuzzy Logic.* WSEAS Transactions on Circuits. Issue 1, Volume 2, January 2003. pp. 252- 258.
[16] Asuncion, A. & Newman, D.J. *UCI Machine Learning Repository* [http:// www.ics.uci.edu/~mlearn/MLRepository. Html]. Irvine CA: University of California, School of Information and Computer Science. (2007).