Method for Optimizing the Number and Precision of Interval-Valued Parameters in a Multi-Object System

AMAURY CABALLERO*, KANG YEN*, *, JOSE L. ABREU**, ALDO PARDO***
*Department of Electrical & Computer Engineering
modeling, simulation, rapid prototyping, model in the loop, software in the loop, hardware in the loop, controller, hydro applicationFlorida International University
10555 W. Flagler Street, Miami, Florida 33174
**CJ Tech, Miami, Florida, 33145
USA
*** University of Pamplona, Colombia
caballer@fiu.edu, yenk@fiu.edu, cjtechno@yahoo.com, apardo13@unipamplona.edu.co

Abstract : In any decision-making process, it is necessary to evaluate the merit of different parameters and disclose the effect they have on the solution in order to optimize it. If the criteria are mathematically quantifiable, a mathematical model may be created for the evaluation process. The application of rough sets, neural networks, fuzzy logic or information theory are widely used mathematical tools for the solution of this type of problem. In this work two tasks are approached: (1) To minimize the number of required parameters(attributes) by discriminating different objects (classes), for the case when there is an overlap in the collected information from parameters (interval-valued information); (2) To calculate the minimum accuracy necessary in the selected attributes to discriminate all of the objects. This approach is very useful in both reducing the cost of the communication channels and in eliminating unnecessary stored information.

Key Words: Information Systems, Classification, Databases, Diffuse Information.

1 Introduction

Extraction and discovery of automated information in a database has been growing due to the development of factors as the introduction in the factory of wire or wireless communication channels, PLCs, and accurate and inexpensive sensors.

Frequently, in the differentiation of different objects or conditions, it is necessary to use more than one attribute. As the number of objects is increased and their properties are more similar, the number of differentiating parameters is also increased.

Due to disturbances, the collected data can have variations into some interval.

Several works have been devoted to the solution of this kind of problem [1], [2], [3] [4]. Among the different solutions to this problem, one method employing information theory. can provide very simple results.

Let us have the following definitions:

A_k ------ attribute # k

U_i , U_j ---- object (class) i or j respectively

l_i^k, l_j^k ---- minimum values for classes i and j respectively

u_i^k, u_j^k ---- maximum values for classes i and j respectively.

The region of coincidence or misclassification rate of the two attributes may vary from 0 when there is no coincidence at all, to 1 when the two attributes coincide completely.

The probability that objects in class U_i are misclassified into class U_j according to attribute k [5] has been modified and given below

$$\alpha_{ij}^k \equiv \alpha \quad if\,[l_i^k, u_i^k] \cap [l_j^k, u_j^k] = \phi; \qquad (1)$$

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

Where α_{ij}^k is the probability that objects in class U_i are misclassified into class U_j as per the attribute A_k,.

Note that in general

$\alpha_{ij}^k \neq \alpha_{ji}^k$.

From the previous result, it was defined the maximum mutual classification error between classes U_i and U_j for attribute k as

$$\beta_{ij}^k = max\{ \alpha_{ij}^k, \alpha_{ji}^k\} \quad (3)$$

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

It is logical to think that when the two classes coincide for some parameter k, the obtained information from this parameter for discriminating between classes i and j is 0, and that value increases as the coincidence diminishes. This leads to the representation of this information, from Shannon and Hartley definition, using a logarithmic scale. Here the logarithm is used to provide the additivity characteristic for independent uncertainty.
For expressing it with logarithms base 10, it is given as

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

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

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

where α is the permissible misclassification error between classes for any attribute. This value shall be defined from the beginning of the classification, and should be larger than zero.
If $I_{ij}^k \geq I_\alpha^k$, the two classes can be separated using the attribute k with the classification error given by α.
If $(l_i - l_j)$ and $(u_j - u_i) \leq 0$, for any i, j combination and for any k, make $I_{ij}^k = 0$

2 Proposed Method for the Parameter's Minimization

The method proposed was developed by the same authors in [6].
The following elements will be defined for the computer program:

L ---- Localization for [$\log_{10} \alpha$]
T1 ---- Table for α_{ij}^k
T2 ---- Table for calculated I_{ij}^k
SA ---- Location in T2, showing the addition of all the I_{ij}^k for each row
SO ---- Location in T2, showing the addition of all the I_{ij}^k for each column
Vector ATU ---- Vector showing the attributes that discriminate selected objects without any uncertainty for the accepted error α
Vector ATD ---- Vector showing the attributes that discriminate selected objects with uncertainty for the accepted error α
Vector UD ---- Vector showing the discriminated objects without any uncertainty for the accepted error α
Vector NUD ---- Vector showing the discriminated objects with some uncertainty for the accepted error α
Vector ND ---- Vector showing objects that cannot be discriminated for the accepted error α

The generalized block diagram is presented in Figure 1. With the application of this method a minimum of attributes is selected for obtaining the parameters classification under the permitted misclassification error α.

3 Application of the Method: Iris Classification

This example is proposed because of its simplicity
The original database presented by R. Fisher [7] has been tested by different methods [8,9].
In this example, the value of α has been selected as 0.1 arbitrarily. So the corresponding value of L is $-\log_{10}(0.1) = 1$ Hartley.
From [7] the attributes for each class (object) have been calculated, and are presented in Table 1.

Table 1. Attributes for Each Class

Attribute	Setosa			
	x_{av}	σ	Min	Max
SL	5.0	0.35	4.3	5.7
SW	3.42	0.38	2.66	4.18
PL	1.45	0.11	1.23	1.67
PW	0.24	0.11	0.1	0.6
Attribute	Versicolor			
	x_{av}	σ	Min	Max
SL	5.94	0.52	4.9	6.98
SW	2.77	0.31	2.15	3.39
PL	4.26	0.47	3.32	5.2
PW	1.33	0.20	0.93	1.73
Attribute	Virginica			
	x_{av}	σ	Min	Max
SL	6.59	0.64	5.31	7.87
SW	2.91	0.32	2.27	3.55
PL	5.55	0.55	4.45	6.65
PW	2.03	0.27	1.49	2.57

The three classes to be differentiated are:
U_1 -- Setosa; U_2 -- Versicolor; U_3 -- Virginica.
The attributes taken into consideration are:
A_1: Sepal Length (SL); A_2: Sepal Width (SW);
A_3: Petal Length (PL); A_4: Petal Width (PW).
Table 2 shows the classification errors between classes, expressed in Hartley.

Table 2. Classification Error between Classes

I_{ij}^k	A_1	A_2	A_3	A_4	∑
U_{12}	0.22	0.23	1	1	
U_{13}	0.22	0.16	1	1	
U_{23}	0	0.05	0.4	0.52	0.97
∑	0.44	0.43	2.11	2.52	

In this table the following conditions are observed:

1) If $I_{ij}^k \geq L$, replace the calculated value by the value of L.
2) If I_{ij}^k, expressed in (4), is less than L locate the calculated value in the corresponding cell of Table 2.

Then for each attribute we add all the values shown in the corresponding column and row. Its sum is recorded in the same column or row of the last row or column named Σ in the table.
The total number of rows is given from

$$C_2^n = (n)(n\text{-}1)/2,$$

where n is the number of classes.
Attributes A_3 and A_4 permit to differentiate U_1 from the other two. Select A_4, because it offers higher information in total. From here, the first rule is obtained

Rule # 1: If A_4 is in the interval [0.1 : 0.6] THEN it is U_1 (Setosa)

As can be seen from Table 2, the differentiation between classes 2 and 3 is not possible for α =0.1, because even in the case of including the three attributes, the obtained information is less than the needed one. In this case, there are two possible solutions:

1. To select a bigger α-value, which will affect the precision of the whole solution.
2. To find out for the selected attributes the minimum necessary precision.

The second approach has been selected because in this case, the precision of the whole model is not affected, and it is not necessary to take action on all the parameters. The new calculation will affect only the selected attributes.
Under certain conditions, it is possible to reduce the standard deviation in the values obtained for all the different attributes. Doing this, the objects discrimination can be improved, but this increases the cost of the whole system. If the action of minimizing the standard deviation is presented only in the parameters that permit to discriminate all the objects, the cost is reduced to a minimum.
In the present case, the standard deviation minimization only should be done on parameter 4.
In general it is possible to find the standard deviation on selected parameters that will permit to discriminate all the objects.
In Figure 2, it is shown the region of coincidence of objects (classes) i and j for the attribute k. Taking the maximum and minimum values for any attribute as $x^k + 2\sigma$ for the maximum and $x^k - 2\sigma$ for the minimum, where x represents the average value for class k . The following assumptions are accepted:

(1) The standard deviation (σ) is the same for the attributes i and j,
(2) The average value for both attributes under analysis does not change.
(3) The variations comply with the normal distribution.

From equation (3), the value β_{ij}^k is calculated as

$$\beta_{ij}^k = (u_i^k - l_j^k) / (u_i^k - l_i^k),$$
$$= [(x_i^k + 2\sigma) - (x_j^k - 2\sigma)/ [(x_i^k + 2\sigma) - (x_i^k - 2\sigma)]$$
$$= [(x_i^k - x_j^k) + 4\sigma]/ 4\sigma$$

But $(x_j^k - x_i^k) = K$ (constant)

$$\beta_{ij}^k = (4\sigma - K)/ 4\sigma \quad (6)$$

It is necessary to reduce the standard deviation to a new value σ_F to comply with the condition given by

$$\alpha = (4\sigma_F - K)/ 4\sigma_F \quad (7)$$

Comparing Equations (6) and (7):

$$\sigma_F = \sigma[(1 - \beta_{ij}^k)/(1 - \alpha)$$

If it is necessary to have the maximum and minimum new values, it can be obtained from

$$u_i^k = x_i^k + 2\sigma_F$$
$$l_j^k = x_j^k - 2\sigma_F$$

Solving for the presented example, the new calculated maximum standard deviation σ_F should be:

$\sigma_F = \sigma[(1 - \beta_{23}^4) / (1 - \alpha)]$
$\sigma_F = 0.27\ [(1- 0.3)/(1 - 0.1)] = 0.21$

4 Conclusions

An algorithm has been presented for attribute reduction in a classification process, using the concept of information (or loss of information) in diffuse databases. In this paper a logarithmic form for expressing the uncertainty in the differentiation between two classes has been used. One of the advantages of the method is that it can be easy to determine how far the solution is from 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 2 whether it is possible to discriminate between two classes or not.

In such cases, where not all the objects (classes) can be discriminated, it is possible to calculate the minimum necessary standard deviation for certain selected attributes under the established conditions of acceptable error, which will permit to discriminate all the objects. This is possible only when there are technical reasons creating the overlap, for example, noise in the communication channel or lack of precision of the equipment. When the overlap is due to other factors, as in the case of surveys or other social situations, this approach is not useful. Applying this procedure after the number of attributes has been minimized results in savings of resources.

The obtained rules serve as a model, and the database can be used for training purposes to discriminate further received information from objects in the same process, under the same conditions.

References

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

[2] Battiti R. Using mutual information for Selecting Features in Supervised Neural Net Learning, *IEEE Transactions on Neural Networks* No. 5,1991, pp. 537-550

[3] 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

[4] Hu ming-li and Lin-li. Research on the Interval-Valued Multiple attribute Decision Model Based on the α-β Dominance Relation. *National Natural Scienc foundation,2013 IEEE.*

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

[6] Caballero A.et Al. A practical Solution for the Classification in Interval-Valued Information Systems, *WSEAS Transactions on Systems and Contro*l, Issue 9, Vol. 5, September 2010, pp.735-744.

[7]] R. Fisher, "The use of Multiple Measurements in Taxonomic Problem," *Annals of Eugenics Cambridge University Press, V. 7, pp.179-138, 1936.*

[8] 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.*

[9] Caballero et al., "Classification with Diffuse or Incomplete Information", *WSEAS Transactions on systems and Control, Issue 6, Vol. 3, June 2008, pp.617-626.*

Objects can't be discriminated

- Find out values in table T2, which are smaller than "*L*" (column SA)
- Locate the objects(U_{ij}) corresponding to the rows previously identified in vector ND and write "0" in all these rows
- If all the rows are "0", the discrimination is finished. If not, move to the next step

There are rows that posses only one value "*L*"

- Select the attributes (A_k) corresponding to these cells and locate it in vector ATU. Locate the objects U_{ij} complying with this condition in vector UD and write "0" in all these rows

There are rows with more than one value "*L*" in some cells

- For each row, select the attribute A_k containing the value "*L*" that presents the largest value in cell SO, and locate it in vector ATU. If there exist this value in more than one cell SO, then select the first encountered attribute in the row.
- Locate the objects U_{ij} that possess the value "*L*" for this attribute in vector UD, and write "0" in all these rows If all the rows are "0", the discrimination is finished. If not, move to the next step

It is necessary more than one attribute for objects discrimination

- For each row, select the cell with maximum I_{ij}^{k}, add this value to the one with the nearest value, and check whether the resultant value is ≥ "*L*". Repeat the process until this condition is met. Locate the corresponding parameters in vector ATD, and the objects U_{ij} complying with this condition in vector NUD. For discriminating among the objects in this condition it becomes necessary to apply some other method.

Figure 1. Generalized Block Diagram

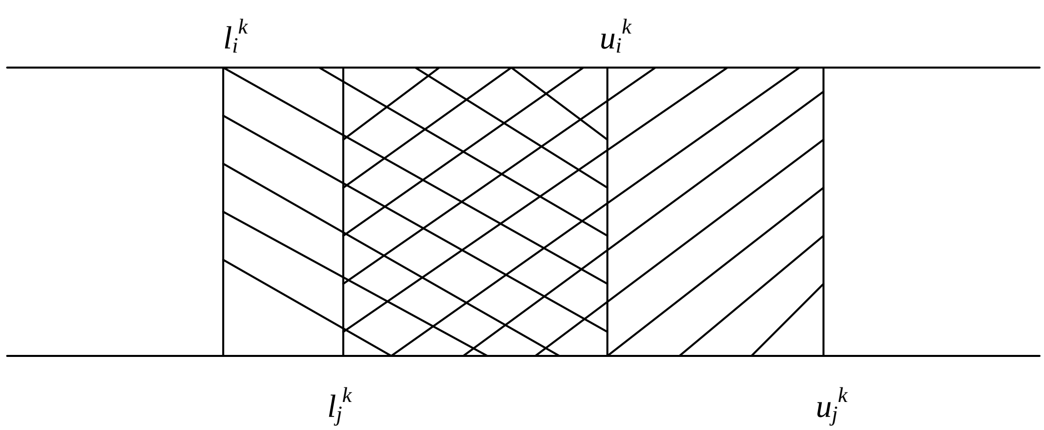

Fig.2. Region of Coincidence of Classes i and j for the Attribute k