Asian Journal of Information Technology 5 (4) : 380-387, 2006

Supervised Machine Learning by Generation of Rules: Optimization of the Size of the Base of Rules of Training by the Method of Inclusion

[1,2]M. Kaddeche and [1]N. Doghmane
[1]Department of Electronic, Faculty of the engineer, Annaba University, Algeria
[2]Laboratory of informatic Research and Study, Reims University, French

Abstract: Within the framework of supervised induction, in this study, a method to minimize the number of rules for a training base is proposed based on optimisation criterion of the simultaneous functions. This method consists in determining the redundant terms (included) in order to lead to a non redundant base. The method is used in various fields such as the automatism to minimise the realisation costs of the simultaneous functions. By analogy, we try through this study, to apply this method to various basic examples in order to remove the redundant rules of a training base and to lead to a base made only prime rules. This method is based on the comparison of two of the same rules classifies. Each rule is made of two parts: (1) The membership area of attributes. (2) The degree of belief of these rules. For the first part, the inclusion notion is applied. However, for the second rule which is it is a number, the superiority or inferiority is used. We present here the experimental tests with the IRIS data base. The obtained principal results and their comparison with other method are given. These results are satisfactory and constitute an additional validation of our method.

Key words: Supervised induction, rules, training, simultaneous functions

INTRODUCTION

The inductive training process can be regarded as a search for plausible general descriptions which explain the data input and which are useful to predict new items.

In other words, the inductive inference tries to derive a complete and correct description of a phenomenon given to part of specific observations of this phenomenon.

In the case of the supervised training, or training starting from examples, we have a set of labelled data, or examples which were associated a class by a professor or an expert.

This set of examples constitutes the base of training.

The contribution of our work in the construction of these rules is in multi-attribute selection for the premises construction and also in attributes discretization.

The attributes selection is done by a research of the linear correlations between the components of the training vectors.

We were inspired at the same time by the work[10] on the supervised classification and of those[8] on the logic multi-valente, which enabled us to introduce a multi-value representation of knowledge.

The rules obtained are classification rules. Their conclusions are assumptions on the membership of a class. They are sullied with uncertainty which we represent by a degree of belief.

Several methods to optimize this base of training rules were developed, we can quote some:

- To increase the threshold in order to minimize the number of rules.
- To remove the rules with degree of confidence inferior to a certain value.
- Genetic algorithms.
- Etc.

It is in this precise context that our work is. We propose a method to minimize these rules. It is a question of decreasing the size of the rules bases by eliminating some of them.

In this article, we will start by presenting the method of generation of rules by detailing the construction of the premises. Then we will describe the detailed phase to minimize the number of rules. We present the principal experimental results obtained on the basis of data IRIS.

Their comparison with the initial data bases, given starting from two types of correlation (inter classes and will intra classes[3], of various cardinals of subdivision and various thresholds, will be also presented. Finally, we will conclude on the principal prospects for this study.

Corresponding Author: M. Kaddeche, Department of Electronic, Faculty of the engineer, Annaba University, Algeria

THE GENERATION OF RULES

We have a training base made of attribute-values examples already classified. That is to say y1..., yc classes defined by the expert and X1..., Xn attributes of the vectors representing the examples[2]. The rules generated starting from this training base is following form:

$$A_1 \text{ and } A_2 \text{ and... and } A_k \rightarrow y, \alpha$$

With

A_i a condition of the type: X j is in [a, b].
X_j the j th component of the vector representing an example.
[a, b] interval resulting from the discretization of the fields of variations of the attributes, here for Xj attribute.
y an assumption on the membership to a class.
α a degree of belief representing the uncertainty of the conclusion.

Regrouping of the attributes: The presented method can be seen within the framework of the polythetic numerical training[7]. The regrouping of the attributes in the premises of the rules corresponds to a multi-attribute selection. This selection is done by a research of the linear correlations between the components of the training vectors. The idea is to locate privileged correlations between the attributes of these vectors and to generate rules holding account of these correlations: the correlated attributes are gathered in the same premise. Such an approach makes it possible to take into account the possible predictive capacity of a conjunction of attributes taken simultaneously[2]. The first stage consists in calculating the matrix of linear correlations R between the components of the training set vectors. It thus acts of a symmetrical square matrix of dimension n*n (n is the total number of attributes). R is noted by $R = (r_{i,j})_{1 \le i \le n, 1 \le j \le n}$

$$R = \begin{bmatrix} 1 & r_{1,2} & \dots & r_{1,n} \\ r_{2,1} & 1 & \dots & \\ & \dots & & \\ r_{1,n} & \dots & & 1 \end{bmatrix}$$

The following stage is that of the thresholding of matrix R. We decide that two attributes Xi and Xj are correlated if the absolute value of $r_{i,j}$ is higher than a threshold θ that we fixed. The decision procedure is defined by:

$$\text{If} \quad |r_{i,j}| < \theta$$

Then Xi and Xj are not correlated and $r_{i,j}$ is affected `0`.
If not Xi and Xj are correlated and $r_{i,j}$ is affected `1`.

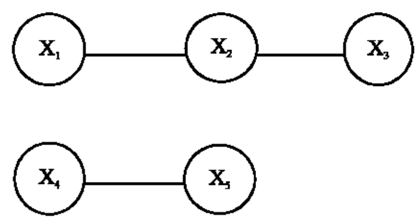
Thus we obtained, starting from the matrix R, a threshold matrix $R_{threshold} = (rs_{i,j})_{1 \le i \le n, 1 \le j \le n}$ composed of 0's and 1's. To determine the subsets of correlated attributes, this matrix can be seen like a graph:

The nodes of the graph are the attributes and two nodes are connected by an arc if and only if they are correlated. The maximum related components of the graph correspond then to the subsets of correlated attributes.

In order to illustrate this step, let us take a simple example. In the case of training examples are represented by 5 attributes {X1, X2, X3, X4, X5}, n = 5, the compitation of linear correlated matrix and its thresholding lead for example to the following binary matrix:

$$R_{seuil} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

The graph associated with this matrix is as follows:



The subsets of correlated attributes are then: {X1, X2, X3}, {X4, X5}. The attributes {X1, X2, X3} on the one hand and {X4, X5} on the other hand will be gathered in the same premises.

The method of search for correlations between attributes which we presented above is carried out on all the unit of training without distinction of class. In[4], we propose one second method, more original, which consists in seeking the linear correlations between the examples of the various classes: it is the search for correlations collate. To focus themselves on the examples belonging to the same class and to seek the correlations between these examples make it possible to characterize the class to which they belong, or at least to release some of its properties, which is interesting for the task of discrimination of the classes. Each class is thus characterized by its own matrix of linear correlations and its own subsets of correlated attributes[2].

Discretization of the attributes and construction of the premises: In this article, we limit ourselves to use a method of non supervised discretization: regular discretization.

We cuts out the field of variation of each attribute in a finished number M of subintervals with the same amplitude. For example, if the field of variation of the components is [0,255] and numbers it subdivisions M is 3, we obtain the following subintervals:

rg_0 = [0, 85], rg_1 = [85,170] and rg_2 = [170,255]

The premises of the rules are then built while considering for each subset of correlated attributes, a subinterval (rg_i) for each attribute and this with all the possible combinations. Thus, in our example, we generate the following premises:

(X1 in rg_a) and (X2 in rg_b) and (X3 in rg_c) is a premise with a, b, c in {0, 1, 2}.
(X4 in rg_d) and (X5 in rg_e) is a premise with d, e in {0, 1, 2}.

In fact, the premises parts of the rules constitute a partition of the space of the correlated attributes. Fig. 1 illustrates such a partition in the case of the subset of correlated attributes {X4, X5}. The regular partition is obtained with a size of M = 3 subdivision.

The completely ordered symbols rg_0, rg_1..., rg_(M-1) resulting from the discretization make it possible to establish a bond with the representation multi-valente used in logics of the same name[1,8]. These symbols represent a scale of degrees. According to the semantics of the attribute considered, they can be interpreted like satisfaction degrees of a measurable adjective and can be represented by scalar adverbs. The choice of the size of the subdivision M depends then on the significance of the attributes and the required precision.

For example the X4 attribute of Fig. 1 can be qualified as not very high if it is in rg_0, fairly high if it is in rg_1 and very high if it is in rg_2.

The introduction of such linguistic terms simplifies the rules and more precisely the premises, at least on the plan of their interpretability and their comprehension by the user. The discretization makes it possible to break up the dynamics of each attribute into several levels.

A level of attribute, such as for example a level of low, average or high gray, is more comprehensible than the data of a single precise numerical value, example gray = 57.

In addition, the discretization of the attributes leads to "non precise" premises: the rules thus built could be started with vague data.

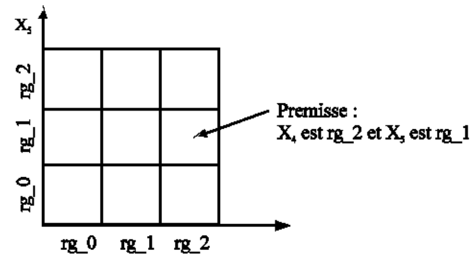


Fig. 1: Example of partition of the space of the correlated attributes

A new object to be classified which we do not know the exact value of the attributes but their orders of magnitude, i.e. their values in term of row, could be classified.

Conclusions of the rules: After the construction of the premises, it comes the generation phase for the rules conclusion part. The generated rules are rules of classification, i.e. they conclude with the membership or not of a class.

Each premise, built according to the method exposed above, leads to the generation of C rules, C being the total number of classes: for each premise, all the possible conclusions are generated.

The conclusions are an assumption on the membership of a class and are affected with uncertainty. The relevance of each generated rule is evaluated compared to the density of the examples belonging to the premise of the rule considered. It is characterized in the conclusion part of the rules by the evaluation of a degree of belief associated with each rule for each class.

This degree expresses the confidence which one can have in this conclusion when the premise is checked. In this article, we propose to represent the degree of belief of the rules by a traditional probability[9].

The degrees of belief are directly estimated on the training set. Such a degree is quantified by the conditional probability to obtain the conclusion when the premise is checked. This conditional probability is estimated on the unit of training according to frequency approach.

COST MINIMIZATION OF SIMULTANEOUS FUNCTIONS IN AUTOMATIC CONTROL

Two methods can be used to represent switching functions using logical gates or electric diagram.

- The first method consists in simplifying each function, then to separately carry out the simplified diagram for each function.

The major disadvantage of this method is that implementation using gates may lead to complex circuits and whose price of realization is very high.

- The second method which minimizes the price of realization is called: method of simultaneous function simplification and which we will illustrate through an example. Let us simplify the three following functions written in the first canonical form (sum of minterms):

  F1 (a, b, c, d) = R (6, 7, 8, 9, 13, 14, 15)
  F2 (a, b, c, d) = R (2, 3, 5, 6, 7, 10, 11, 14, 15)
  F3 (a, b, c, d) = R (2, 3, 5, 7, 8, 9, 10, 11, 13, 15)

It is a digital representation of the `' 1 " of the function (decimal digit 6 is written in binary as 0110 = $\bar{a}\,bc\,\bar{d}$ where a, b, c and d are binary variables).
a) Search for canonical forms:
A canonical couple is composed of two parts:

- The first part represents the combination of the input variables (a, b, c and d).
- The second is a combination of outputs, indicating the output functions to which these combinations of variables belong.

Thus, the research for canonical forms consists in determining the first complete set of basis for the functions: F1, F2, F3, F1*F2, F1*F3, F2*F3 and F1*F2*F3. By using the Karnaugh map, one can obtain the following results:

$$F1 = b\,c + a\,b\,d + a\,\bar{c}\,d + a\bar{b}\bar{c}$$
$$F2 = c + \bar{a}b\,d$$
$$F3 = c\,d + \bar{b}\,c + b\,d + a\,d + a\,\bar{b}$$
$$F1*F2 = b\,c$$
$$F1*F3 = b\,c\,d + a\,b\,d + a\,\bar{c}\,d + a\bar{b}\bar{c}$$
$$F2*F3 = c\,d + \bar{b}\,c + \bar{a}\,b\,d$$
$$F1*F2*F3 = b\,c\,d$$

We will have the following canonical couples composed of two parts (binary combinaison and function of membership):

$b\,c\,d$ F1F2F3	$a\,b\,d$ F1F3	$\bar{b}\,c$ F3	$\bar{a}\,b\,d$ F2
$c\,d$ F2F3	$a\,\bar{c}\,d$ F1F3	$b\,d$ F3	$b\,c$ F1
$\bar{b}\,c$ F2F3	$a\,\bar{b}\,\bar{c}$ F1F3	$a\,d$ F3	$a\,b\,d$ F1
$\bar{a}\,b\,d$ F2F3	$b\,c$ F1F2	$a\,\bar{b}$ F3	$a\,\bar{c}\,d$ F1
$b\,c\,d$ F1F3	$c\,d$ F3	c F2	$a\,\bar{b}\,\bar{c}$ F1

b) Search for first canonical couples:
A couple (f F) is made up of a part f which is a combination of binary variables and a second part F, the function of membership of this combination.
It is said that the couple (f F) is a canonical couple first:
If there is not any couple (f ' F') such as:

$$f' = f \text{ or } f' \subset f$$
$$\text{and } F' \subset F$$

These conditions enable us to eliminate all the non primary canonical couples (redundant couples) and to keep only the primary ones which are involved in the realization of these simultaneous functions.

Example: is not a primary canonical couple, it is a redundant couple and it should be rejected because there is a couple among the canonical couples such as:

abd = abd (f = f') and F1 ⊂ F1F3.

Among the twenty canonical couples cited previously, there remain only twelve primary canonical couples which intervene in the function of minimal realization and which are:

$b\,c\,d$ F1F2F3	$a\,b\,d$ F1F3	$b\,d$ F3
$c\,d$ F2F3	$a\,\bar{c}\,d$ F1F3	$a\,d$ F3
$\bar{b}\,c$ F2F3	$a\,\bar{b}\,\bar{c}$ F1F3	$a\,\bar{b}$ F3
$\bar{a}\,b\,d$ F2F3	$b\,c$ F1F2	c F2

The eight other couples are rejected because they are not essentially primary.
In conclusion, the function of realization of these simultaneous functions is reduced and the cost of realization is minimized by using this method.

SEARCH METHOD FOR THE PRIMARY RULES OF A TRAINING BASE

This method uses the rules coming from the union of the results obtained from attributes correlation (mixed correlation[5], inter and intra classes).
Examples with three classes C1, C2 and C3:

Group 1: If X1 is in rg-0 then C1 with 0.65
If X3 is in rg-2 then C3 with 0.72
……………………………
……………………………

Groupe2: If X1 is in rg-1 and if X2 is in rg-2 then C2 with 0.84.
If X2 is in rg-2 and if X4 is in rg-0 then C4 with 0.73.
………………………………………………….
…………………………………………………

Groupe3: If X1 is in rg-0 and if X2 is in rg-1 and if X4 is in rg-1 then C1 with 0.62.
If X2 is in rg-2 and if X3 is in rg-1 and if X4 is in rg-0 then C2 with 0.87.

One notes by the functions:
F: Expressions in the form (If X1 is in rg-0 and if X2 is in rg-2).
Ci: class i.
α: The degree of confidence.

It is said that the rule:
F' Ci α ' is not prime (redundant)
If and only if $\exists$ F Cj α such as:
{ F' $\subseteq$ F and i = j and α ' $\leq$ α }.
(The concept of inclusion is used in the sense of space inclusion)

Example:
1st Rule: If X1 is in rg-0 and X2 is in rg-1 then C1 with a degree of confidence $\alpha = 0.68$.
2nd Rule: If X1 is in rg-0 then C1 with a degree of confidence $\alpha = 0.70$.

It is concluded that the 1st Rule is redundant and it is removed from the training rule base.

This study of comparison of rules is done in the following way:

Rules relative to an attribute are compared to rules with 2 attributes, then 3 attributes, then 4 attributes, etc.

- And also of all the inter classes attributes correlations (C1 U C2, C1 U C3, C2 U C3 and C1 U C2 U C3).

One classifies according to a column all the rules in the following way:

- Groupe1: rules related to only 1attribute.
- Group 2: rules related to 2 attributes.
- Group 3: rules related to 3 attributes.
- Etc.

has as many Groups as attributes.

- Then, rules relative to two attributes are compared to rules with three attributes, four attributes, etc.
- Then, rules relative to three attributes are compared to rules with four attributes, etc.
- Etc.

At the end of the comparison, the rules included in other rules are known as redundant (rejected rules) and the new optimal base (prime base) is built included from the remaining rules known as prime rules.

SYSTEM OF INFERENCE

The inference engine receives as input the base of rules, as well as a vector describing the object to be classified. The inference engine associates a class this vector.

More precisely, the inference engine manages the uncertainty of the rules and makes it possible to obtain a degree of confidence associated with each class for the new object.

The triggered rules are those whose premises are matched exactly by the new vector to classify. The engine must manage the uncertainty of the rules and ensure its taking into account of inference dynamics.

If the degrees of belief are probabilities, the reasoning being used is of Bayesian type. It is of type MYCIN using the theory of the confirmation if the degrees of belief are coefficients of certainty.

Once the rules are triggered, the problem is initially to define how the degrees of various rules concluding with the same assumption, i.e. with the same class, are combined to obtain a final degree of confidence in this class.

If the uncertainty of the rules is represented by a probability, we propose to use, for the treatment of uncertainty a triangular co-norm. The final confidence coefficient associated with each class yi is the result of the calculation of this co-norm on the probabilities associated with the fixed rules and conclusive with yi.

The triangular standards and co-norms are operators used within the framework of the treatment of uncertain knowledge and decision-making [YAG 85]. Their origin is found in the study of probabilistic metric spaces.

They were then introduced into the theory of the fuzzy subsets. We wish to combine degrees of belief coming from rules concluding all to the same class. We are interested in the co-norms, which constitute an example of function of aggregation [DUB 85]. For further details, the interested reader will be able to refer to article [GUP 91].

Certain T-norms or T-co-norms can prove more effective than others, in particular in the processes of decision-making. We chose to test two different co-norms:

- The co-norm of Zadeh, S (p, q) = max (p, Q).
- The probabilistic co-norm whose principle is to reinforce the degree of belief, S (p, q) = p + q-p*q.

It would be interesting to test other co-norms with the different characteristics like the co-norm from Lukasiewicz (S(p, q) = min (p + q, 1)) or the co-norm of Weber (S(p, q) = p if q = 0, S(p, q) = q if p = 0, S(p, q) = 1 if not).

In a more general way, one could study and take into account other functions of aggregation. This point constitutes a perspective for our future work.

IMPLEMENTATION AND RESULTS

The tests: We carried out simulations, using MATLAB tool, on a real dataset. The base used comes from UCI server of the University of Irvine (California, the USA). We considered the Iris data base.

For the tests, we fixed in an empirical way the values of the parameters. The objective is to find the optimal rules base.

The various values of the parameters tested are:

- Threshold of correlation (θ): 0.5, 0.6, 0.7, 0.8, 0.9
- cardinal of the subdivision (M): 3, 5, 7
- Cross-country race validation of order 10.

We chose to test the probabilistic co-norm whose principle is to reinforce the degree of confidence, S (p, Q) = p + q-p*q.

Experimental results: The results obtained, using this method of optimization which is based on the concept of inclusion, are represented below by the various tables. For the IRIS base, we compare our results with the results found by other authors[3] and that using a chart.

It will be noted however that the rules, of which the degree of belief is lower than a threshold of confidence than one the fixed at 0.43, are eliminated Table 1, 2 and 3).

According to the results given by tables 1, 2 and 3, we will note that:

The method of mixed correlation leads for an average size of subdivision (M = 5), independently of the value of the threshold, to the size of the most optimal rule base.

Table 1: A number of rules optimized for a cardinal of M = 3 subdivision (bases data IRIS)

Threshold of correlation (θ)	0.5	0.8	0.9	0.95
Number of rules in initial rule base	492	240	96	96
Redundant rules	251	125	52	52
Rules whose rate is lower than the threshold of confidence	177	88	36	36
A number of rules of the optimal base	64	27	8	8
Rate of classification	96.78	97.52	96.72	96.72

Table 2: A number of rules optimized for a cardinal of M = 5 subdivision (bases data IRIS)

Threshold of correlation (θ)	0.5	0.8	0.9	0.95
Number of rules in initial rule base	2385	712	213	213
Redundant rules	1334	425	127	127
Rules whose rate is lower than the threshold of confidence	725	149	67	67
A number of rules of the optimal base	326	138	19	19
Rate of classification	93.12	93.68	94.26	94.26

Table 3: A number of rules optimized for a cardinal of M = 7 subdivision (bases data IRIS)

Threshold of correlation (θ)	0.5	0.8	0.9	0.95
Number of rules in initial rule base	8483	1478	284	284
Redundant rules	5219	863	187	187
Rules whose rate is lower than the threshold of confidence	2759	391	66	66
A number of rules of the optimal base	906	224	31	31
Rate of classification	92.85	91.11	90.16	90.15

Table 4: The rate of classification (%) according to the cardinal of subdivision (the highest rate for values = 0.5 to 0.95)

	Rate of classification (%)		
Cardinal of subdivision	inter classes correlation	intra classes correlation	Mixed correlation
M = 3	98.00	97.33	97.52
M = 5	93.33	94.00	94.26
M = 7	92.67	94.67	92.85

Table 5: The rate of classification (%) according to the threshold of correlation (the highest rate for values of the cardinal of subdivision M = 0.3, 0.5 and 0.7)

	Rate of classification (%)		
Threshold of correlation (σ)	Correlation inter classes	Correlation will intra classes	Mixed correlation
0.5	96.67	94.67	96.78
0.8	96.67	97.33	97.52
0.90	98.00	97.33	96.72
0.95	98.00	97.33	96.72

For a subdivision of a reduced size, one will note that for high thresholds (σ = 0.9 to 0.95),

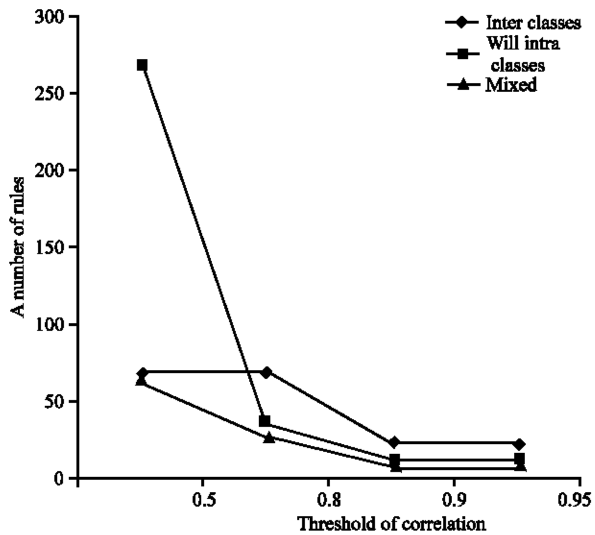


Fig. 2: A number of rules according to the threshold of correlation (M = 3)

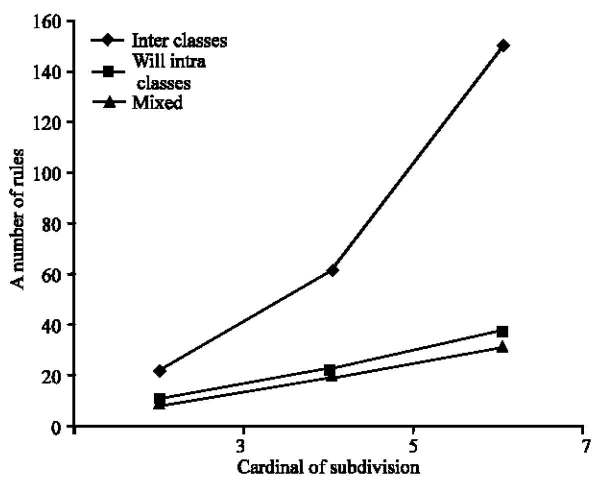


Fig. 3: A number of rules according to the cardinal of subdivision (threshold = 0.95)

The optimal rule base will remain unchanged as found by the method of mixed correlation; but the optimal rate of classification is however similar to that given by the inter classes method (all confused classes).

Lastly, for important sizes of subdivision (M = 7), the size of the optimal rule base is all the time that similar to that given by the mixed correlation; but the optimal rate of classification is obtained by the method of intra classes correlations for weak thresholds (σ = 0.5 to 0.8).

By observing the graphs of Fig. 2 and 3 and in comparison with the results presented in the thesis under the reference[3] (results of the correlations in intra classes and inter classes), the method of mixed correlation

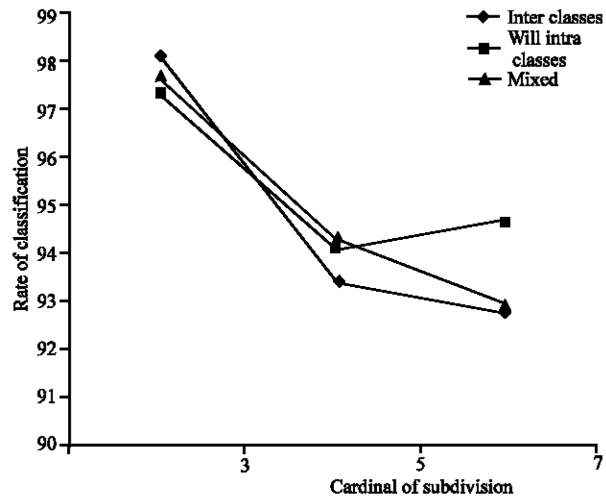


Fig. 4: Rate of classification according to the cardinal of subdivision

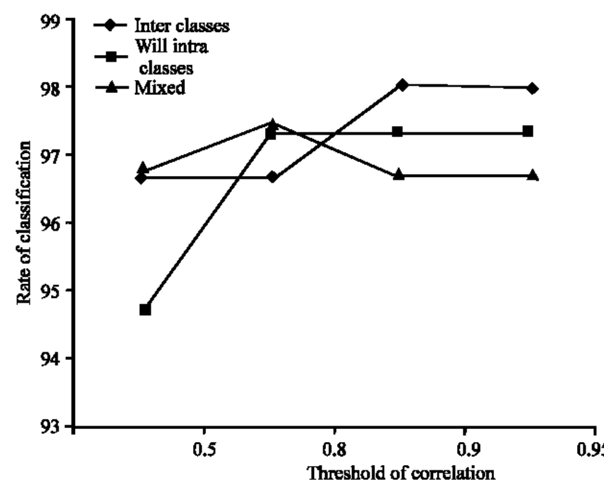


Fig. 5: Rate of classification according to the threshold of correlation

introduced in this study gives the smallest size for the training rule base independently of both the cardinal of subdivision (M) and the threshold of correlation (σ).

One obtains a very high rate of classification for the three methods of correlation for a weak cardinal of subdivision (M = 3) (Table 4 and Fig. 4).

However, this optimal rate is obtained with lower threshold as regard to the mixed and intra classes correlations and with higher threshold for the inter class correlation.

For an average cardinal of subdivision (M = 5), the optimal rate of classification for the three methods of correlation is obtained for higher thresholds. The suggested mixed correlation method gives the best result.

It is noticed that for regular discretization by great cardinal of subdivision, the method of intra class correlation is that which gives best classification.

One notices that our method is less powerful for higher thresholds of correlations, that is explained by the simplification of many rules which are significant (rejection of the rules whose rate of classification is lower than the fixed threshold of confidence) and a part is necessary for a good classification.

CONCLUSION

We presented a polythetic method of supervised training. The inductive training consists in automatically generating rules of classification to multi-valent premise. These rules are an explicit form of representation of knowledge. The rules thus built are exploited by a system of inference.

This system manages the uncertainty of the rules and is able to take into account inferential dynamics.

An interesting point of our approach is the use of T-co-norms for the aggregation of the degrees of belief. The results obtained were compared with different other approaches tested on IRIS data base.

The results obtained are very satisfactory, they constitute an additional validation of our approach and open the way to new experiments.

One of the advantages of our approach resides indeed in the use of a very small number of rules, which makes their interpretation very easy.

One could also supplement our work for example by eliminating certain rules, those which do not intervene or are not very influential at the time of the phase of recognition. For this purpose, one could have recourse to a method of rules selection by introducing a forgetting term or using genetic algorithms.

Another technique that may constitute a perspective for our method of mixed correlation is possible, for higher threshold values of correlation and which consists in testing thresholds of confidence lower than the threshold used in our simulations.

Admittedly, with the suppression of certain rules, one could expect a reduction in the performance, a compromise must then be found between effectiveness, in term of good classification and legibility of the results.

REFERENCES

1. Akdag, H., M. De Glas and D. Pacholczyk, 1992. A Qualitative Theory of Uncertainty. Fundamenta Informaticae, 17: 333-362.
2. Borgi, A., J.M. Bazin and H. Akdag, 1997. Classification supervisée d'images par génération automatique de règles. Société Francophone de Classification, Lyon.
3. Borgi, A., 1999. Apprentissage supervisé par génération de règles : le système SUCRAGE. Thèse de doctorat en informatique de l'Université Paris 6.
4. Borgi, A., J.M. Bazin and H. Akdag, 1998. Two methods of linear correlation search for a knowledge based supervised classification. 11th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems '98, IEA/AIE-98.
5. Kaddeche, N., 2005. Supervised machine learning by generation of rules: Optimization of the rate of classification by the search for mixed correlation. Asian J. Inform. Tech., 4: 152-161.
6. Borgi, A. and H. Akdag, 2001. Apprentissage supervisé et raisonnement approximatif, l'hypothèse des imperfections'. Revue d'Intelligence Artificielle, Editions Hermès. 15: 55-85.
7. Di Palma, S., J. Da Rugna and D.A. Zighed, 1997. Apprentissage supervisé polythétique: Une évaluation. Cinquièmes Rencontres de la Société Francophone de Classification, SFC'97, Lyon.
8. Ginsberg, M.L., 1988. Multivalued logics: A uniform approach to reasoning in artificial intelligence. Computational Intelligence 4: 265-316.
9. Pearl, J., 1988. Probabilistic reasoning in intelligent systems: Networks of Plausible Inference. Morgan Kaufman Publishers, Inc. Computer Science, Artificial Intelligence.
10. Vernazza. G., 1993. Image Classification By Extended Certainty Factors. Pattern Recognition, Pergamon Press Ltd, 26: 1683-1694.