KNOWLEDGE BASED SUPERVISED CLASSIFICATION : AN APPLICATION TO IMAGE PROCESSING

A. Borgi (*, **)

H. Akdag (*, **)

(*) LIP6, Université P. & M. Curie
4, Place Jussieu, Boîte 169,
75252 Paris Cedex 05, FRANCE
Amel.Borgi@lip6.fr

(**) LERI, Université de Reims
Rue des Crayères, B.P. 1035
51687 Reims Cedex 2, FRANCE
Herman.Akdag@lip6.fr

Summary

We present a knowledge based supervised classification method. Our modelisation is based on automatic generation of classification rules. The classification function is directly given in the form of production rules base.
The proposed learning method is multi-features, it allows to take into account the possible predictive power of a simultaneously considered features conjunction. On the other hand, the feature space partition allows a multi-valued representation of the features and data imprecision integration. The rules conclusions are accompanied by belief degrees. This uncertainty is managed in the learning phase as well as in the recognition one. To introduce more flexibility and overcome the boundary problem due to the discretisation, we propose to use approximate reasoning. We introduce, in this purpose, an adequate distance to compare neighboring facts. This distance, measuring imprecision, combined with uncertainty of classification decisions represented by belief degrees, drives the approximate inference.
The proposed method was implemented in a tool called SUCRAGE and confronted with a real application in the field of image processing. The obtained results are very satisfactory. They validate our approach and allow us to consider other application fields.

Keywords : supervised classification, approximate reasoning, imprecision and uncertainty treatment, image processing.

1. INTRODUCTION

Facing the increase of data amount recorded daily, the detection of both structures and specific links between them, the organisation and the search of exploitable knowledge in this information become a strategic stake for decision holding and prediction task.
This complex problem, also known as « Data Mining » has multiple aspects. We focus on one of them : supervised learning. We propose a learning method from examples situated at the junction of statistical methods and those based on Artificial Intelligence techniques. Our modelisation is based on automatic generation of classification rules. The classification function is directly given in the form of production rules base. This ensures the transparency and easy interpretation of the classifier.

The construction of production rules *IF [premise] THEN [conclusion]* using the knowledge and the know-how of an expert is a very difficult task. The complexity and cost of such a knowledge acquisition have led to an important development of learning methods used for an automatic knowledge extraction [9] [7].
In the pattern recognition domain, expert's rules allow to determine the belonging of a pixel to a class. For instance, in a human thigh cryosection image a pixel will be classified as bone, muscle, ... In the medical domain, it is practically impossible to obtain from an image classification given by a medical specialist a coherent and complete set of rules. The only trusted informations he can give us, are relations between pixels and classes such as $P_i \in C_j$, where P_i is a pixel and C_j is a class defined by the domain specialist. A learning method based only on such information allows to limit mistakes during knowledge acquisition. Therefore, starting from the expertise given by a medical expert such as $P_i \in C_j$, we propose to build classification rules automatically. So, we present an image classification system based on a supervised learning method. Our package is composed by two sub-systems. The first one, the generator, ensures the learning phase by automatic rules construction : « knowledge acquisition » from training pixels is automatic. The proposed learning method is multi-features, it allows to take into account the possible predictive power of a simultaneously considered features conjunction. The feature space partition allows a multi-valued representation of the

features and data imprecision integration. The second sub-system, the inference engine, uses these rules to classify new pixels. To introduce more flexibility and overcome the boundary problem due to the discretisation, we propose to use an approximate reasoning. The proposed approximate reasoning, used as an inference mode, allows to manage imprecise knowledge as well as rules uncertainty.

The proposed method was implemented in a tool called SUCRAGE and confronted with a real application in the field of image processing (multi-components image segmentation). The obtained results are very satisfactory. They validate our approach and allow us to consider other application fields.

2. BUILDING THE RULES

One of the particular interest of our work consists in this base of rules automatically constructed. To achieve this goal, the picture processing expert represents a pixel by a vector : each pixel feature corresponds to a vector component. We have then a set of classified vectors, the training set, that we use to build the base of rules.

Let C_1, ..., C_c be the possible classes defined by the medical expert. Let X_1, ..., X_n be the components of the vector representing the pixel features. The generated rules are of the type :

$$\mathbf{A_1 \text{ and } A_2 \text{ and ... and } A_k \longrightarrow C, \alpha}$$

where A_i is a condition of the form : X_j is in [a,b], X_j is the j^{th} vector component, the interval [a,b] is issued from the discretization technique. Following Vernazza [15] and Ginsberg [6], we introduce, by this discretization, a multi-valued representation of knowledge. C is a hypothesis about membership in a class with a belief degree α.

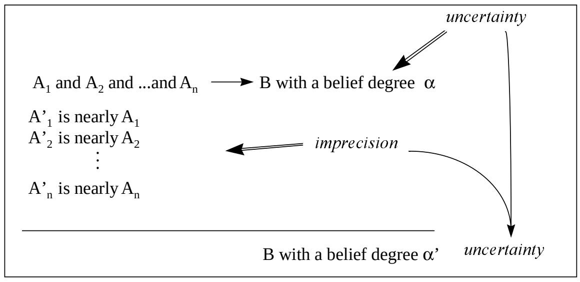
The building of the premises of the production rules is realized by linear correlation search among the training set elements [5]. To determine the grouping of components of different rules, we use the correlation matrix $M=(r_{i,j})_{n*n}$, where r_{ij} is the coefficient of linear correlation between X_i and X_j vector components. To decide which components are correlated, we define a threshold θ, and we consider that (X_i) and (X_j) are independent if $|r_{ij}|<\theta$. So, high correlated components are grouped in the same premises.

Then, to determine the interval [a,b], we decompose the global range of a component into a finite number M of sub-ranges of equal width. Premises of the rules are obtained by considering for each correlated components subset, an interval for each component in all possible combinations [3].

As in most applications, picture recognition is concerned with uncertainty [8] [10]. In fact, the conclusion parts of the rules are hypotheses about membership of a given class. Each hypothesis is accompanied by a belief degree α that can be computed by different methods (probability [11], certainty factors [13] or possibility degree [16]).

3. THE INFERENCE SYSTEM

The rules were built for the purpose of a further classification use. In fact, an inference engine will exploit the base of rules to classify new pixels. The main difference between our system and more classical ones is in the treatment of uncertainty. Moreover, the originality of our model is not only in the management of uncertainty but especially in the taking account of imprecision. In fact, we propose two reasoning models. The first one, called *exact reasoning* allows the inference engine to fire only the rules with which the new pixel components match exactly. The second, called *approximate reasoning* consists in firing also some neighborhood rules, and consequently treats imprecision [4]. This kind of approximate reasoning manipulates concepts developed in works related to uncertainty management techniques such us multi-valued logic [1], fuzzy logic [16],... and more particularly uses the neighborhood degree and distance between observations and premises. Our modelisation corresponds to the Generalized Modus Ponens schema. In our particular case, it is expressed by the following diagram :



In order to be productive, we need to formalize the neighboring notion, which allows to take observation imprecision into account. The conclusion degree α' of above schema, must be a function of both observation imprecision and rule uncertainty.

The approximate inference is only possible when a meta-knowledge exists in the system and allows it to run. In our case the meta-knowledge gives the possibility to bind imprecision (observation and premise of rule) to uncertainty (conclusion degree).

This meta knowledge, fundamental in approximate reasoning, has two important aspects :

1. A weak difference between observation and premise induces that the conclusion part is not

significantly modified. For every rule, a stability area exists around the premise of the rule. This is a quality that we have to find in any approximate reasoning.

2. If the distance between observation and premise increases, then uncertainty of the conclusion increases too. A maximal distance must give a complete uncertainty.

3.1 DISTANCES BETWEEN NEIGHBORING FACTS

To assess the proximity between an observation and a premise, we introduce a local distance between an element of observation (A'_i) and an element of premise (A_i).

In our application, rule premise (A_i) associates discrete values to pixel components. But observations (A'_i) are pixels with numerical values. Two kinds of solutions exist for computing distance between A_i and A'_i. The first consists in extracting expertise in order to obtain an estimation of symbolic distance. Such an expertise extraction seems very hard to us. The second kind of solution, that we use, consists in a numerical-symbolic interface [4].

These local distances are then aggregated into a global distance that reflect the closeness between premise and observation.

In order to take into account the values dispersion, we don't use tools like min-max functions. We are interested only by the measure of neighboring facts, that is facts which are not too distant. Thus, it is necessary to use an aggregated distance that is very sensitive to little variations of neighboring facts. We propose for that a global distance that measures proximity between approximately equal vectors [4]. This distance is insensitive when facts are very far and is a representation of neighboring-dispersion values. We note that the approximate inference can only be realized when the distance between rule and observation is acceptable.

3.2 THE APPROXIMATE INFERENCE

The global distance concept allows to model the representation of the neighboring facts. But how can we compute the effect of this distance on the conclusion of the rule? As we said above, the approximate inference is only possible when a meta-knowledge exists in the system and allows it to run. In our case, this point consists in linking imprecision (observation and premise of rule) with uncertainty (conclusion degree).

The conclusion degree is weakened in accordance with the global distance. In our model, belief degrees (α) associated to rules are numerical, so it is hoped, for the whole coherence, to conserve a numerical final degree (α'). In our work, we compute the final belief degree using the following method : Let R be a rule of the form « If A_1 and A_2 and ... A_n then (B, α) » and A an observed fact, d is the global distance between A and R (d is a symbolic distance). In order to compute the final degree α' of B, associated to the observation A, we proceed in two steps.

Step 1 :

We numerize the symbolic distance d using a function δ :

$$\delta : \begin{array}{l} D \longrightarrow [0,1] \subset IR \\ d \longrightarrow d/(p-1) \end{array}$$

For p possible values of d in D=[0 .. p-1], we will have p equidistant values between 0 and 1.

Step 2 :

We compute α' using the following formula :

$$\alpha' = \alpha * (1-\delta(d))$$

This formula includes the two aspects of the meta-knowledge hypothesis mentioned above. It's easy to observe that little imprecisions (in cases where d=0) don't modify uncertainty. On the other hand, a maximal distance (d = p-1) induces a complete uncertainty (α'=0).

The proposed function also verify the properties of a Modus Ponens generating function [14].

3.3 THE CLASSIFIER DYNAMIC

For a new pixel to classify, the system begins by computing its discretized components. This is done according to the discretization technique used in the building of the rules premises. Then the inference engine regroups the fired rules according to the class of their conclusion parts. In other words, for every class C_i we obtain a set of rules, denoted Rules(C_i), containing the fired rules that conclude to the class C_i.

We then have to compute a final belief degree associated to each class, for this we use a co-norm S such as :

$$S(p,q) = \max (p,q)$$

or

$$S(p,q) = p+q-p*q$$

We remind that a co-norm is a function used in knowledge uncertainty treatment. It constitutes an example of aggregation function [1] [16]. A co-norm is a real function

$$S : [0,1]x[0,1] \rightarrow [0,1]$$

having the following properties :

c1 : $S(1,1)=1$
c2 : $S(p,0)=S(0,p)=p$
c3 : $S(p,q)=S(q,p)$
c4 : $S(p,q)\leq S(r,c)$ if $p\leq r$ and $q\leq c$
c5 : $S(p,S(q,r))=S(S(p,q),r)$

Finally, the winner class associated to the new pixel is the class where the final belief degree is maximum.

4. RESULTS AND CONCLUSION

To validate our system, we test different kind of data [2]. The experimental results, number of rules and good classification rates, were obtained using the cross-reference validation method.

We first test our system on the well-known Iris data of Fisher (the number of classes is three and observations are represented by four components vectors). The experimental results are satisfactory and reach 97.33% of good classification. A comparative study with other learning methods, such as decision trees [12] and induction graphs [17], shows that our approach generally allows an improvement of the classification rates.

We have also confronted our approach to image data and tested our system on a cryo-section of human thigh image. In this case, the number of classes is four (bone, muscle, fat and marrow) and pixels are represented by 5 attributes. The obtained classification rates are very satisfying and reach 99% of good recognition. An other kind of picture, a colored butterfly, was tested too and leads to 85% of good classification. This rate is also satisfying according to the kind of the considered image where classes boundaries are not well defined. These satisfying results are the consequences of the adequate representation of the expert's knowledge by the automatic generated rules. That means that our building system is successful in the translation of the elementary medical expert knowledge into production rules. Moreover, the inference system exploits nearly-perfectly these rules, as the low error rate proves. These encouraging results lead us to test our modelisation in other fields of expertise where objects to classify can be represented by vectors.

References

[1] H. Akdag, M. De Glas and D. Pacholczyk. *A Qualitative Theory of Uncertainty*. Fundamenta Informaticae 17, pp 333-362, 1992.

[2] A. Borgi. *Apprentissage supervisé par génération de règles : le système SUCRAGE*. Phd Thesis, Université Pierre et Marie Curie, Paris 6, France.

[3] A. Borgi, J-M. Bazin, H. Akdag. *Supervised Classification by Automatic Rules Generation*. The Fourth World Congress On Expert Systems, Application of Advanced Information Technologies, Mexico 1998.

[4] A. Borgi, J-M. Bazin, H. Akdag. *A numerical approach to approximate reasoning via a symbolic interface. Application to image classification*. Fuzzy-Neuro Systems '98, Computational Intelligence. Munich, 1998.

[5] A. Borgi, H. Akdag. *Induction supervisée de règles : le système SUCRAGE*. Conférence d'Apprentissage, CAP'99. Plate-forme AFIA. Paris, Juin 1999.

[6] M. L. Ginsberg. *Multivalued logics : a uniform approach to reasoning in artificial intelligence*. Computational Intelligence Vol. 4, pp 265-316, 1988.

[7] Y. Kodratoff and E. Diday, editors. *Induction Symbolique et Numérique à Partir de Données*, volume 1. Cépaduès-Editions, 1991.

[8] T. Law, H. Itoh, H. Seki. *Image Filtering, Edge Detection, and Edge Tracing Using Fuzzy Reasoning*. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.18, N.5,1996.

[9] R. S. Michalski, R.L. Chilausky. *Knowledge acquisition by encoding expert rules versus computer induction from examples : a case study involving soybean pathology*. Fuzzy Reasoning and its Applications, edited by E. H. Mamdani and B.R. Gaines, pp 247-271, 1981.

[10] A. Mogre, R. McLaren, J. Keller and R. Krishnapuram. *Uncertainty Management for Rule-Based Systems with Applications to Image Analysis*. IEEE Transactions on Systems, Man, and Cybernetics, VOL. 24, No. 3, March 1994.

[11] J. Pearl. *Numerical Uncertainty In Expert Systems*. Readings in Uncertain Reasoning. Edited by G. Shafer and J. Pearl. Morgan Kaufman publishers. Inc. California 1990.

[12] J. R. Quinlan. *Induction of Decision Trees*. Machine Learning, Vol. 1, p 82-106, Kluwer Academic Publishers, Boston, 1986.

[13] E. H. Shortliff and B.G. Buchanan. *A Model of Inexact Reasoning in Medicine*. Readings in Uncertain Reasoning. Morgan Kaufman publishers. Inc. California 1990.

[14] E. Trillas and L. Valverde. *On Mode and Implication in Approximate Reasoning*. Approximate Reasoning in Expert Systems, pp 157-166. M. M. Gupta, A. Kandel, W. Bandler, J. B. Kiszka editors, Elsevier Science Publishers, 1985.

[15] G. Vernazza. *Image Classification By Extended Certainty Factors*. Pattern Recognition, Vol. 26, No. 11, pp. 1683-1694. Pergamon Press Ltd, 1993.

[16] L.A. Zadeh. *Fuzzy Sets as a Basis for a Theory of Possibility*. Fuzzy Sets and Systems 1, pp. 3-28. North-holland Publishing company, 1978.

[17] D. A. Zighed, J. P. Auray, G. Duru. *SIPINA : Méthode et logiciel*. Editions Alexandre Lacassagne, Mathématiques appliquées n°2. 1992.