



HAL
open science

A feature selection method based on Choquet Integral and Typicality Analysis

Cyril Mazaud, Jan Rendek, Vincent Bombardier, Laurent Wendling

► **To cite this version:**

Cyril Mazaud, Jan Rendek, Vincent Bombardier, Laurent Wendling. A feature selection method based on Choquet Integral and Typicality Analysis. 16th International Conference on Fuzzy Systems - FUZZIEEE'07, Jul 2007, Londres, United Kingdom. pp.1703-1708, 10.1109/FUZZY.2007.4295623 . hal-00137983

HAL Id: hal-00137983

<https://hal.science/hal-00137983>

Submitted on 7 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A feature selection method based on Choquet Integral and Typicality Analysis

Cyril Mazaud, Jan Rendek, Vincent Bombardier and Laurent Wendling

Abstract— An iterative feature selection method based on feature typicality and interactivity analysis is presented in this paper. The aim is to enhance model interpretability by selecting the best significant features among a list extracted from images. The inference mechanism uses a fuzzy linguistic rule-based system. This method is applied here to a wood defect classification problem. Nowadays, feature selection is expertise-driven and most of the time, expert uses features by habits which not always represent the best ones to use. The proposed approach aims to replace expert selection by automatically choosing a suitable set of features to the recognition problem.

I. INTRODUCTION

In many pattern recognition applications, a feature selection scheme is fundamental to focus on most significant data while decreasing the dimensionality of the problem under consideration. The information to be extract from the images is not always trivial, and to ensure that the maximum amount of information is obtained, the number of extracted features can strongly increase. The feature selection area of interest consists in reducing the problem dimension. It can be described as an optimization problem where a feature subset is searched in order to maximize the classification performance of the recognition system. Among the several approaches used to perform such optimization, genetic algorithms are widely used [1-5]. Feature selection can thus be seen from two points of view: selecting the most representative features thanks to data mining techniques or by using expert knowledge about the concerned field of study.

A. Current feature selection

Our pattern recognition method relies on a fuzzy linguistic rule-based system. Its complexity depends on the number of features used to generate it. The more are features used as input to the system, the more are rules generated during the learning stage. More details about the recognition method are given in section III. Until now, feature selection is done by expert knowledge of the studied field which

concerns wood defect classification. Two kinds of complementary knowledge is used to choose the most appropriated features: knowledge about the defect definition (wood expert knowledge) allowing to recognize the defects in naked eyes; and knowledge about the industrial vision system allowing to define what kind of features could be used to recognize the wished defects.

B. Problematic

The problem of such selection relies on the validation of this decision. It requires to be sure that all the chosen features are the best ones and to be sure that they carry all the useful information to make the right recognition. Moreover, this validation could decrease the model size (reducing the number of input features) allowing to keep the recognition mechanism interpretability given by the use of fuzzy linguistic rules [6].

C. Proposed feature selection method

Many classifier combination systems have been proposed and compared in the literature [7-10]. The method we propose consists in analyzing the learning database used to generate the rule-based system. We extract efficient subsets of features using the typicality analysis and the Choquet integral. Typicality analysis [11] is based on the resemblance and the dissimilarity of one sample compared to the defect classes. The Choquet integral is part of the aggregation techniques based on fuzzy integrals. Fuzzy integrals, and the Choquet integral in particular, have been successfully used as fusion operators in various applications [12, 13] including Content-Based Image Retrieval [14] and speech recognition [15].

First an initialisation is done (step 0). Then an iterative global feature selection process is performed and can be roughly split into two steps (1 and 2). Figure 1 illustrates the feature selection process proposed here.

- (Step 0) The application of typicality analysis allows to propose a primary set of features, and also to validate this choice by training and testing the recognition model with them. A global recognition rate is obtained and assumed to be a reference set.
- (Step 1) From this first set of features, a feature interactivity process is applied to determine the less representative ones.
- (Step 2) Generate the recognition model without the first less representative features and test it. The reached recognition rate is stored. The process is repeated using the k next less representative features.

Cyril Mazaud is with the Research Centre for Automatic Control, CRAN, CNRS, UMR 7039, Université Henri Poincaré, Campus scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France (email: cyril.mazaud@cran.uhp-nancy.fr).

Jan Rendek is with the Lorraine Laboratory of IT Research and Its Application, LORIA, UMR 7053, Université Henri Poincaré, Campus scientifique, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France (email: jan.rendek@loria.fr).

Vincent Bombardier is with the Research Centre for Automatic Control, CRAN, France (phone: +33-3-83-68-44-52; fax: +33-3-83-68-44-59; e-mail: vincent.bombardier@cran.uhp-nancy.fr).

Laurent Wendling is with the Lorraine Laboratory of IT Research and Its Application, LORIA, France (email: Laurent.wendling@loria.fr)

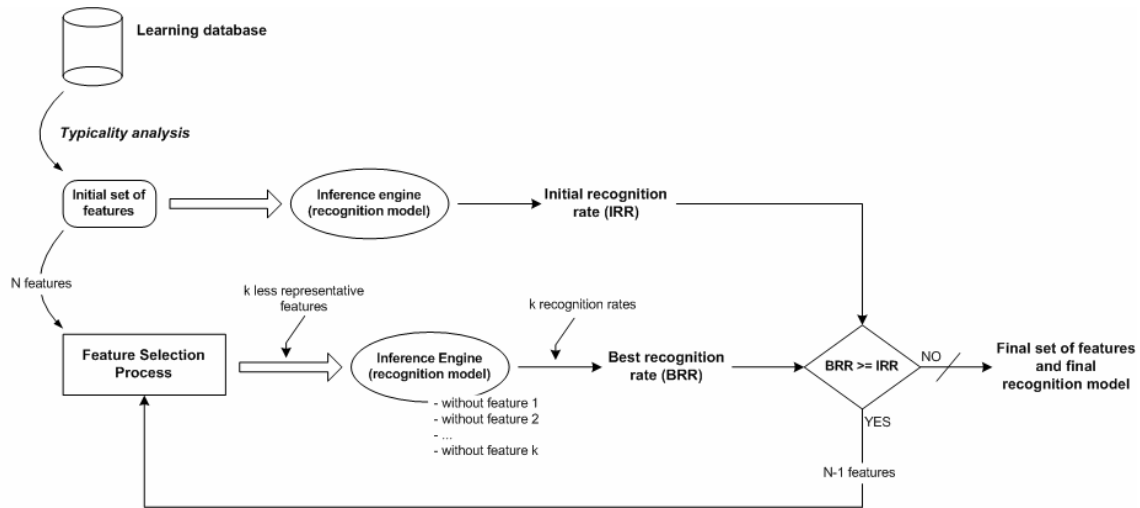


Fig. 1. Proposed feature selection process

If one of the recognition rates is better or equal to the initial one (reference rate obtained in step 0), the weaker associated feature is removed from the list and go to Step 1 (with $n-1$ features) and so on. Else, the first set of features given by the step 0 is kept and the associated recognition model is used for the final pattern recognition application.

II. FEATURE SELECTION METHODS

In this section, we introduce the two feature selection methods used for our proposed feature selection process. The first one is the typicality analysis which consists in extracting the initial set of features from the learning database. The second one is the feature interactivity analysis. It consists in analyzing the initial list of features to extract the three less representative ones. The result of this analysis (as shown in section IV) is a matrix which coefficients indicate how significant the features are.

A. Typicality analysis

Typicality gives information about the representativeness of a sample for a feature. In our case, as each sample of the learning database is labelled, the typicality notion returns information about the representativeness of a defect class (set of the samples composing the class) for a feature. It is evaluated with the calculation of the *typicality coefficient* for the different features. The typicality calculation (3) is based on the intern dissimilarity (1) and extern likeness (2) evaluations for all samples.

$$R(V_a^u) = \frac{\sum_{i=1}^n \frac{1}{d(V_a^u, V_a^{f_i})}}{n} \quad (1)$$

$$D(V_a^u) = \frac{\sum_{i=1}^m \frac{1}{1 - d(V_a^u, V_a^{e_i})}}{m} \quad (2)$$

$$T(V_a^u) = \frac{R \cdot D}{R \cdot D + (1 - R) \cdot (1 - D)} \quad (3)$$

where V_a^u is the value of the feature a for the point u ;

$V_a^{f_i}$ is the value of the feature a for the “friend” point

f of the class i ;

$V_a^{e_i}$ is the value of the feature a for the “enemy” point e of the class i ;

$d(x, y)$ is the Euclidian distance between x and y ;

n is the number of “friend” points;

m is the number of “enemy” points;

R and D correspond to $R(V_a^u)$ and $D(V_a^u)$.

The figures 2 and 3 show how a feature can be considered as significant. The decision is made thanks to a visual interpretation of the obtained graphs which represent the classes’ distribution (i.e. the sample set distribution).

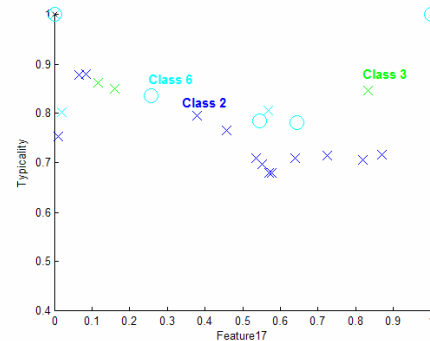


Fig. 2. Example of an insignificant feature according to typicality analysis

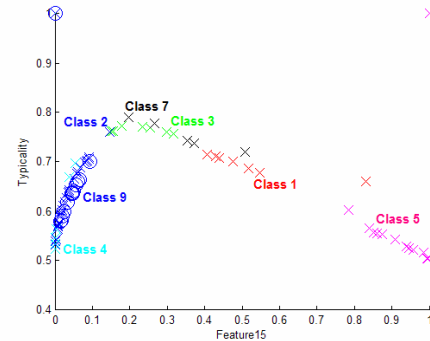


Fig. 3. Example of a significant feature according to typicality analysis

B. Determination of suitable features using Choquet Integral

1) *Choquet integral*: The Choquet integral was first introduced in capacity theory [16]. For the sake of clarity, we will only present here the definitions useful to explain our work. In particular, we consider discrete spaces only. For a thorough description of the Choquet integral, the reader may refer to [17-19]. Let us consider m classes, C_1, \dots, C_m , and n decision rules (DRs) D_1, \dots, D_n . When a new pattern x_θ is observed, we wish to find the class it most likely belongs to. Labelling this unknown pattern is a three-steps process. Firstly, for each decision rule j we need to compute φ_j the degree of confidence in the statement “According to D_j , x_θ belongs to the class C ”. Secondly, we combine all these partial confidence degrees into a global confidence degree by choosing a suitable aggregation operator.

Let us denote $X = \{D_1, \dots, D_n\}$ the set of decision rules, and P the power set of X , i.e. the set of all subsets of X .

Definition 1 A fuzzy measure or capacity, μ , defined on X is a set function μ :

$P(X) \rightarrow [0, 1]$, verifying the following axioms:

1. $\mu(\emptyset) = 0, \mu(X) = 1$
2. $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

Fuzzy measures generalize additive measures, by replacing the additivity axiom ($\mu(A \cup B) = \mu(A) + \mu(B)$, $A \cap B = \emptyset$) by a weaker one. Fuzzy measures embed particular cases including probability measure, possibility and necessity measures, or belief and plausibility functions. In our context of decision rules fusion, $\mu(A)$ represents the weight of importance, or the degree of trust in the decision provided by the subset A of DRs. The next step in building a final decision is to combine the partial confidence degree according to each DR into a global confidence degree.

Definition 2 Let μ be a fuzzy measure on X . The discrete Choquet integral of $\varphi = [\varphi_1, \dots, \varphi_n]_t$ with respect to μ , noted $C_\mu(x)$, is defined by:

$$C_\mu(\varphi) = \sum_{j=1, n} \varphi(j) \left[\mu(A_{(j)}) - \mu(A_{(j+1)}) \right] \quad (4)$$

where $(.)$ is a permutation, such as $\varphi_{(1)} \leq \dots \leq \varphi_{(n)}$. Also $A_{(j)} = \{(j), \dots, (n)\}$ represents the $[j..n]$ associated criteria in increasing order and $A_{(n+1)} = \emptyset$.

2) *Fuzzy measure*: From the above definitions, we can see that the behavior of the Choquet integral as an aggregation operator entirely relies on the fuzzy measure used. There are several methods to determine the most adequate fuzzy measure to be used for a given application (see [17]). To cope with initialization problems (ill-conditioned matrices, convergence, processing time...) “heuristic” algorithms have been developed. Their goal is to find an approximation of the fuzzy measure that minimizes an error criterion. To our

knowledge, the algorithm providing the best approximation is the one proposed by Grabisch in [20]. It assumes that in the absence of any information, the most reasonable way of aggregation is the arithmetic mean, thus a Choquet integral with respect to an additive equidistributed fuzzy measure. Any input of information tends to move away the fuzzy measure from this equilibrium point. This means that, in case of few data, coefficients of the fuzzy measure which are not concerned with the data are kept as near as possible to the equilibrium point. Thus, this algorithm is still efficient when training data are limited. It also has a low computing time and a low memory cost.

3) *Indices*: Once the fuzzy measure is learned, it is possible to interpret the contribution of each decision rule in the final decision. Several indexes can be extracted from the fuzzy measure, helping to analyze its behavior.

Importance index. The importance index, also called the *Shapley index*, is based on the definition proposed by Shapley in game theory [21] and introduced in a fuzzy measure context by Murofushi and Soneda [22]. It is defined for a fuzzy measure μ :

$$\sigma(\mu, i) = \frac{1}{n} \sum_{t=0}^{n-1} \frac{1}{\binom{n-1}{t}} \sum_{\substack{T \subseteq X \setminus i \\ |T|=t}} [\mu(T \cup i) - \mu(T)] \quad (5)$$

The Shapley value can be interpreted as a weighted average value of the marginal contribution $\mu(T \cup i) - \mu(T)$ of the decision rule i alone in all combinations. A property worthy to be noted is that the sum of the indexes of all DRs equals to one. In other words $\sum_{i=1, n} \sigma(\mu, i) = 1$. Hence, a DR with an importance index value lesser than $1/n$ can be interpreted as an importance below in the final decision.

Interaction index. The interaction index, also called the *Murofushi and Soneda* index represents the degree of interaction between two decision rules. If the fuzzy measure is non-additive then some sources interact. The *marginal interaction* between i and j , conditioned to the presence of elements of the combination $T \subseteq X \setminus ij$ is:

$$(\Delta_{ij}\mu)(T) = \mu(T \cup ij) + \mu(T) - \mu(T \cup i) - \mu(T \cup j) \quad (6)$$

Averaging this criterion over all the subsets of $T \subseteq X \setminus ij$ gives the interaction index of sources i and j , as defined by Murofushi and Soneda [22].

$$I(\mu, ij) = \sum_{T \subseteq X \setminus ij} \frac{(n-t-2)!t!}{(n-1)!} (\Delta_{ij}\mu)(T) \quad (7)$$

A positive interaction index for two DRs i and j means that the importance of one DR is reinforced by the second. In other words, both DRs are complementary and their combined use better the final decision. The magnitude of this complementarity is given by the value of the index. A negative interaction index indicates that the sources are

antagonist, and their combined use impairs the final decision. Given a training set composed of a set of input vectors and the expected values of their Choquet integral with an ideal measure. Grabisch's algorithm [20], used here, learns the fuzzy measure using a gradient descent algorithm. This algorithm tries to minimize the mean square error between the values of the Choquet integral with respect to the fuzzy measure being learned, and the expected values. For a training sample, the parameter vector is the current values of the fuzzy measure along the path dictated by the ordering of the training vector coordinates. This parameter vector is translated along the direction of the gradient, with a magnitude proportional to the error, thus updating the values along the path.

4) *Improving the final decision using confusion matrices:* While retaining Grabisch's algorithm, we modified the way the labeled training patterns are processed to compose a training set. A training pattern yields to m training samples Φ_1, \dots, Φ_m , with $\Phi_i = (\varphi_{i1}, \dots, \varphi_{im})$ where φ_{ij} represents the confidence in the fact that the sample belongs to class i , according to DR j . For each of these samples, a target value must be assigned. For techniques that use a different fuzzy measure per class, the optimal target value minimizing the quadratic error is known [20]. For techniques that use a single fuzzy measure no such formula exists and most often the following one is used: $C\mu(\Phi_i) = 1$, if sample belongs to class i , and 0 otherwise. This is where our approach differs. The bigger the error between the real value of the Choquet integral is the more the fuzzy measure coefficient will be modified. The Choquet integral being an averaging operator output values of 0 or 1 will never be reached on real data. Hence, each learning sample will move the fuzzy measure away from the weighted arithmetic mean, whether it is already correctly recognized or not.

Our idea is to link the target value assigned to a sample with an estimation of the confusion between classes, that is the probability for each pair of classes (C_i, C_j) that an element of C_i is recognized as belonging to C_j . We want the learning algorithm to put the emphasis on distinguishing classes that are confused. If little confusion exists between two classes, we assume that the Choquet integral based on the current fuzzy measure leads to acceptable results, since its initialization emulates a weighted arithmetic mean. This translates into assigning specific target values to the outcome of the Choquet integral according to the class a training sample belongs to, and the class we try to match it against. The confusion between classes is estimated by first building the confusion matrix [23] for each DR. Then an average confusion matrix is built by averaging these matrices. Confusion matrices are used as a visualization tool to see if a recognition system is confusing two classes, i.e. routinely mislabeling one as an other. Following the global confusion between classes, a decreasing function is defined to take into account it. Thus, the target value for the sample associated with the class having the least confusion is the outcome of the Choquet integral. With such a target value, this samples leaves the fuzzy measure unchanged, when processed by the learning algorithm. On the contrary, the

target for the sample associated with the class having the most confusion is set to zero, implying the biggest modification possible on the fuzzy measure.

Automatic extraction of subsets of DRs. Once the lattice is learn, we analyze the individual performance of each DR in the produced fuzzy measure. This analysis is performed using the importance and interaction indexes, namely the Shapley and Murofushi-Soneda indexes. We wish to track the DRs that are the least important in the final decision, and that positively interact the least with the other rules. We assume that such DR blurs the final decision. We implemented a two step selection scheme to discard such DR. First, the Shapley value is scaled by the number of DRs, n . A DR with a scaled importance index greater than 1 describes a DR more important than the average. We select the set of low significant rules S_L having an importance index lower than 1:

$$S_L = \{k/n \cdot \sigma(\mu, k) < 1\} \quad (8)$$

Then, we extract from S_L the subset of decision rules having the least positive synergy with the others. For each rule S_{iL} , the values of its interaction with others are averaged to estimate its global interaction. Finally the subset of rules to be removed MS_L is composed of the rules from S_L that have an interaction index lower than the mean of the interaction indexes of all rules of S_L :

$$MS_L = \left\{ k / \sum_{j=1, n} I(\mu, k_j) < m \right\}, k \in S_L \quad (9)$$

with the global mean interaction index:

$$m = 1 / |S_L| \sum_{k \in S_L} \sum_{j=1, n} I(\mu, k_j) \quad (10)$$

Putting it all together. The application of our learning algorithm followed by the extraction of the most relevant DRs forms a training epoch. After each epoch, the recognition results from the fusion of the DRs using a Choquet integral based the fuzzy measure trained is evaluated. This evaluation can either be done on some labeled samples that were not part of the training set, or on the full training set itself if it is too small to be split. If the epoch has improved the recognition, another epoch is performed. Otherwise, the fuzzy measure computed before the current epoch is kept, and the algorithm stops. Selected features are then provided to the next step.

III. FUZZY INFERENCE ENGINE

The choice of a fuzzy logic-based method for our application in the wood defect detection field could be justified by three main reasons. Firstly, the defects to be recognized are intrinsically fuzzy (gradual transition between clear wood and defects). The features extracted from the images are thus uncertain (but precisely calculated) and the use of fuzzy logic allows to take it into account. Secondly, the customer expresses his needs under a nominal

form; the output classes are thus subjective and often not separated (non strict boundary between the class representing a small knot and the class representing a large knot, for example). Finally, the customer needs and the human operator experience are subjective and mainly expressed in natural language.

The implemented inference mechanism uses a fuzzy linguistic rules-based system where each rule is automatically generated from the learning database according to the iterative form of the Ishibuchi-Nozaki-Tanaka's algorithm [24]. The obtained rules, under a matrix form, are used to classify the different non-labeled samples. The figure 4 represents the rough principle of the Fuzzy Reasoning Classifier (FRC) developed for our application, underlying the training and the generalization steps.

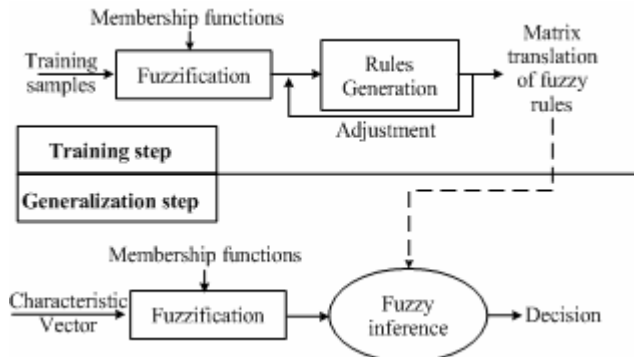


Fig. 4. Inference mechanism principle

For more details about the inference mechanism implemented, the reader can refer to [25].

IV. EXPERIMENTAL RESULTS

The results presented in this section are based on a real set of samples collected from a wood industrial case. The database is composed of 877 samples divided in nine classes. The learning database, where the typicality and interactivity analysis are made, is composed of 250 samples. The database used to test and validate the feature selection is composed of 627 samples. The fuzzy inference engine consists in a single inference where all the features are in input of the model and all the classes to recognize in output.

Initially, the feature selection was done using expertise (i.e. using expert knowledge). In the present case, the expert has chosen six features (among nineteen extracted from the images) classically used to recognize the defects present in the database. The obtained recognition rate in generalization reached 74.2% and 242 rules were generated.

In the proposed feature selection method, the first step of the method is the use of the typicality analysis to choose the first set of features. The method gives nine features which seem significant to recognize all the classes. Among these features, four of them represents a notion of shape (features n°2, n°3, n°8, n°9), one of them a measure of size (feature n°6) and four of them a measure of color (features n°10, n°12, n°13, n°15). The result of such an analysis is twice: the recognition rate increases to 77.8%, but 4 641 rules were generated, implying real difficulties to interpret the model.

The interest of the iterative part of our method (interactivity analysis) is to decrease the number of features to use (i.e. decrease the model complexity) while preserving the high recognition rate achieved with the typicality analysis. At each step of the iterative part, the method provides a complete analysis, under a matrix form, of the feature interactivity and the feature significance. The table 1 illustrates one of the matrices generated during our tests where values in grey mean potential insignificant features (M1 corresponding to a binary output and M2 integrating the confusion between sources).

TABLE I
MATRIX OF A FEATURE INTERACTIVITY ANALYSIS

Feature	N°9	N°8	N°13	N°12	N°3	N°6	N°10
Shapley M1	1.04	0.8	0.73	0.84	0.54	1.68	1.4
Muro M1		2		3	1		
Shapley M2	1.09	1.06	0.89	1.18	0.46	1.25	1.05
Muro M2			2		1		

In the present case, the iterative part allows to remove three features from the initial list given by the typicality analysis. These features concern one feature of color and two features of shape. A final recognition rate of 78.6% is reached. We can denote that it is better than the one obtained by expertise or by the typicality analysis. Moreover the number of generated rules decreases to 468 (divided by 10 compared to the typicality analysis, but increased by 2 compared to the expertise feature choice).

The figure 5 and 6 summarizes respectively the different recognition rates and the number of rules obtained with the different feature selection methods.

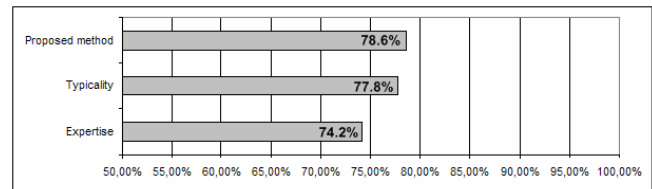


Fig. 5. Recognition rates obtained with three methods

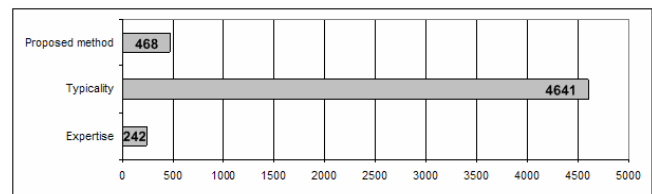


Fig. 6. Number of generated rules with the three methods

V. CONCLUSIONS

In this article, we have presented an iterative feature selection method based on a learning database analysis. The proposed method is composed of three steps: first, the initialization part, which consists in choosing a first subset of features thanks to a feature typicality analysis. The recognition rate obtained with this subset (composed of N features) is then taken as a reference for the rest of the method. Second and third steps are the iterative part of the

method which principle is to analyze the feature interactivity among the initial subset and determine the k less representative features. The combination of k features ($N - k^{\text{th}}$ features) is tested using the fuzzy inference engine and the best recognition rate is retained. If this recognition rate is better or equal than those obtained in the initial step, the k^{th} associated feature is removed from the subset and the feature selection method is applied again until a final subset is found.

The results obtained in the experimental section shows that our method can choose an optimal subset of features significantly increasing the recognition rate compared to the expert feature choice while keeping a certain interpretability of the model. However, even if the recognition rate is better, the interpretability of the model is not easy with 468 rules to interpret. Further investigations will be performed to reduce the number of generated rules. For that, we will apply the proposed method to analyze the most significant features for each class and not for the whole classes as done for the moment. This improvement is the basis of a feature selection adapted to the use of a higher level decision making process which gives better results in term of recognition rate [25], and in term of interpretability. This model is based on a tree structure where the configuration is led by expertise and expert knowledge integration.

ACKNOWLEDGMENT

The authors would like to thank the LuxScan Technologies company for providing the industrial database which allowed this study on real samples.

REFERENCES

- [1] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann and L. Magdalena, "Ten years if genetic fuzzy systems: current framework and new trends", *Fuzzy Sets and Systems*, vol. 141, pp. 5-31, 2004.
- [2] H. Handels, Th. Roß, J. Kreuzsch, H.H. Wolff and S.J. Pöppel, "Feature selection for optimized skin tumor recognition using genetic algorithms", *Artificial Intelligence in Medicine*, vol. 16, pp.283-297, 1999.
- [3] N.K. Jain, V.K. Jain and K. Deb, "Optimization of process parameters of mechanical type advanced machining processes using genetic algorithms", *International Journal of Machine Tools & Manufacture*, vol. 47, pp. 900-919, 2007.
- [4] N.R. Siriwardene and B.J.C. Perera, "Selection of genetic algorithm operators for urban drainage model parameter optimization", *Mathematical and Compute Modelling*, vol. 44, pp. 415-429, 2006.
- [5] C.H. Huang and C-J. Wang, "A GA-based feature selection and parameters optimization for support vector machines", *Expert Systems with Applications*, vol. 31, pp. 231-240, 2006.
- [6] D. Dubois and H. Prade, "What are Fuzzy rules and how to use them?", *Fuzzy Sets and Systems*, vol. 84, pp. 169-185, 1996.
- [7] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers", *IEEE Transactions on PAMI*, vol. 20(3), pp. 226-239, 1998.
- [8] L.I. Kuncheva and C.J. Whitaker, "Measures of diversity in classifier ensembles", *Machine Learning*, vol. 51, pp. 181-207, 2003.
- [9] D. Ruta and B. Gabrys, "An overview of classifier fusion methods", *Computing and Information System*, vol. 7, pp. 1-10, 2000.
- [10] Z. Stejic, Y. Takama, and K. Hirota, "Mathematical aggregation operators in image retrieval: effect on retrieval performance and role in relevance feedback", *SignalProcessing*, vol. 85(2), pp. 297-324, 2005.
- [11] J. Forest, M. Rifqi and B. Bouchon-Meunier, "Class Segmentation to Improve Fuzzy Prototype Construction: Visualization and Characterization of Non Homogeneous Classes (in French)," *Rencontres Francophones sur la Logique Floue et ses Applications LFA*, Toulouse, France, pp. 29-36, 2006.
- [12] S. B. Cho and J. H. Kim, "Combining multiple neural networks by fuzzy integral for robust classification", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 25(2), pp. 380-384, February 1995.
- [13] H. Tahani and J.M. Keller, "Information fusion in computer vision using the fuzzy integral", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 20(3), May/June 1990.
- [14] Y. S. Choi and D. Kim, "Relevance feedback for content-based image retrieval using the choquet integral", *IEEE trans. on Knowledge and Data Engineering*, vol. 16(10), pp. 1185-1199, 2004.
- [15] S.Chang and S.Greenberg, "Application of fuzzy-integration-based multiple information aggregation in automatic speech recognition", In *In Proceedings of the IEEE Conference on Fuzzy Integration Processing, Beijing*, 2003.
- [16] G. Choquet, "Theory of capacities", *Annales de l'Institut Fourier*, vol. 5, pp. 131-295, 1953.
- [17] M. Grabisch and J.M. Nicolas, "Classification by fuzzy integral - performance and tests", *Fuzzy Sets and Systems, Special Issue on Pattern Recognition*, vol. 65, pp. 255-271, 1994.
- [18] J.-L. Marichal, "Aggregation of interacting criteria by means of the discrete choquet integral", In *Aggregation operators: new trends and applications*, pages 224-244. Physica-Verlag GmbH, Heidelberg, Germany, 2002.
- [19] T. Murofushi and M. Sugeno, "A theory of fuzzy measures: representations, the Choquet integral, and null sets", *Journal of Mathematical Analysis and Applications*, vol. 159, pp. 532-549, 1991.
- [20] M. Grabisch, "A new algorithm for identifying fuzzy measures and its application to pattern recognition", In *International Joint Conference of the 4th IEEE International Conference on Fuzzy Systems and the 2nd International Fuzzy Engineering Symposium*, pp. 145-150, Yokohama, Japan, March 1995.
- [21] L. Shapley, "A value for n-person games", In H. Khun and A. Tucker, editors, *Contributions to the Theory of Games, Annals of Mathematics Studies*, pp. 307-317. Princeton University Press, 1953.
- [22] T.Murofushi and S. Soneda, "Techniques for reading fuzzy measures(iii): interaction index.", In *Proceedings of the 9th Fuzzy System Symposium, Sapporo, Japan*, pp. 693-696, May 1993. In Japanese.
- [23] R. Kohavi and F. Provost, "Glossary of terms", *Machine Learning*, vol. 30, pp. 271-274, 1998.
- [24] H. Ishibuchi, K. Nozaki, and H. Tanaka, "A Simple but powerful heuristic method for generating fuzzy rules from numerical data", *Fuzzy sets and systems*, vol. 86, 1997, pp. 251-270.
- [25] E. Schmitt, C. Mazaud, V. Bombardier and P. Lhoste, "A Fuzzy Reasoning Classification Method for Pattern Recognition," *Fuzzy IEEE*, Vancouver, Canada, 2006.