



**HAL**  
open science

# Fuzzy Rule Iterative Feature Selection (FRIFS) with Respect to the Choquet Integral Apply to Fabric Defect Recognition

Emmanuel Schmitt, Vincent Bombardier, Laurent Wendling

► **To cite this version:**

Emmanuel Schmitt, Vincent Bombardier, Laurent Wendling. Fuzzy Rule Iterative Feature Selection (FRIFS) with Respect to the Choquet Integral Apply to Fabric Defect Recognition. 10th International Conference on Advanced Concepts for Intelligent Vision Systems, ACIVS 2008, Oct 2008, Juan les Pins, France. pp.708-718. hal-00341647

**HAL Id: hal-00341647**

**<https://hal.science/hal-00341647>**

Submitted on 26 Nov 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.

# Fuzzy Rule Iterative Feature Selection (FRIFS) with respect to the Choquet Integral apply to fabric defect recognition

Emmanuel Schmitt<sup>1</sup>, Vincent. Bombardier<sup>1</sup>, Laurent Wendling<sup>2</sup>

<sup>1</sup>CRAN/<sup>2</sup>LORIA, Université Henri Poincaré  
54500 Vandœuvre-lès-Nancy, France

{schmitt,bombardier}@cran.uhp-nancy.fr  
wendling@loria.fr

**Abstract.** An iterative method to select suitable features in an industrial fabric defect recognition context is proposed in this paper. It combines a global feature selection method based on the Choquet integral and a fuzzy linguistic rule classifier. The experimental study shows the wanted behaviour of this approach: the feature number decreases whereas the recognition rate increases. Thus, the number of generated fuzzy rules is reduced.

## 1 Introduction

The feature selection method proposed in this paper takes place in a problematic of complexity reduction. The application domain relates to quality control in a technical textile industry (carbon fibers, Kevlar...). The aim of the global vision system is to identify fabric defects in a continuous mode during the production. Fabric defects recognition involves two steps, the defect detection and the defect identification. The first part does not concern this work because this confidential step is provided by the industrialist. So this paper focuses on the second part.

Because of this specific industrial context, there are many constraints. One constraint is the necessity of working with very small training data sets (sometimes, there is only one or two samples for a defect class because of its rareness). Another difficulty is to respect the real time constraint in the industrial production system [27], [29]. So, low complexity must be kept for the recognition model. Such a classification problem has been relatively poorly investigated in the early years [28], [30], [31].

Thus, this work takes place on a “small scale” domain according to [1], [2] definition because of the weak number of used features. Moreover, defects are intrinsically fuzzy. For example, there is not always a strict boundary between “sound” fabric and a “defective” region; this transition is more or less gradual. The segmentation step provides an accurate “defective” region, i.e. the calculated characteristics are accurate but uncertain (the same defect could be processed twice without obtaining the same characteristic vector). Thus, the recognition method must take these specificities into

account. Using fuzzy logic minimizes this effect to obtain a measure less sensible to these uncertainties. So a classifier based on Fuzzy Linguistic Rules has been chosen.

The second part of this paper focuses on the selection of suitable parameters, an associated weak classifier, in order to decrease the number of rules. Handling with several classifiers allows for integrating their discriminatory aspect to improve the recognition step [3]. Despite pattern recognition methods are generally independently built. Their combination may lead to positive correlations because both aim at achieve the same goal and both are based on the same learning data [4][5]. Nonetheless even if approaches like Adaboost, arcing [6] and boosting [7]... try to limit this dependence by reinforcing the diversity it is difficult to measure it in order to efficiently incorporate it in the classification process [3][8].

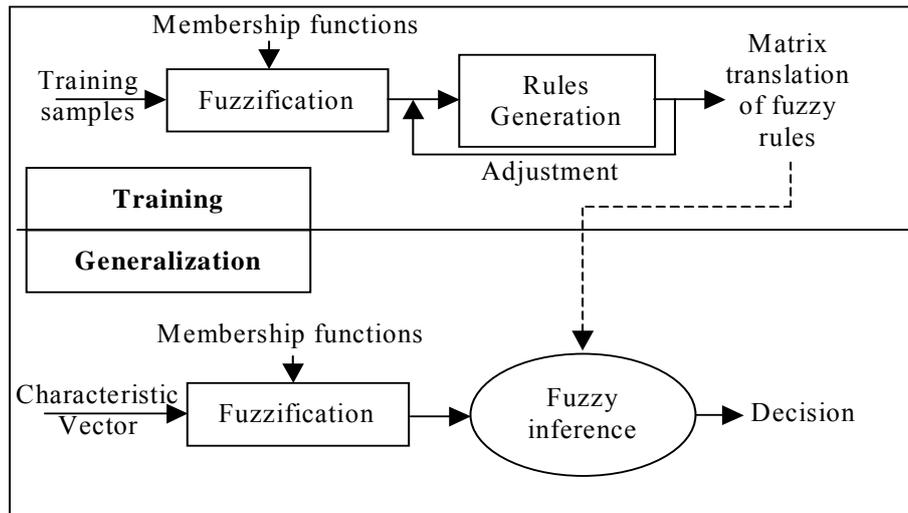
In this context, we handle with both few sets of learning data and simple parameters considering processing time constraint. Furthermore data may be inconsistent due to the fast acquisition step. Ways to aggregate such parameters have been studied in this context.

A fuzzy measure learning scheme is used with respect to the Choquet integral as it allows consistent learning even if only a few samples per class are processed. Such an algorithm is suitable for this specific application where few learning data are provided by the industrialist and also to handle ambiguous features. As numerous rules are provided by the system it is neither easy to determine which features are not important nor which features are redundant.

Some backgrounds on the Fuzzy Linguistic Rule Classifier are introduced. Then, a scheme to discard weaker parameters using the Choquet Integral is given and finally the global method and its application in an industrial context are presented.

## **2 Fuzzy Linguistic Rule Classifier**

The used classifier (F.R.C.: Fuzzy Reasoning Classifier) [12] is based on fuzzy linguistic rule mechanism; which is well adapted to our industrial application. Indeed, it presents a very good and efficient generalisation from a few sample set and is able to provide gradual membership for output classes [10]. Its satisfactory behaviour has been shown in [10] by several comparisons with other classifiers such as k Nearest Neighbor (k-NN), Neural Networks (NN) or Support Vector machine (SVM). This implemented algorithm for the fuzzy recognition method is a supervised learning mechanism which can be decomposed into three parts: Input fuzzification (features of the characteristic vector), Fuzzy rule generation and Rule adjustment (Fig. 1).



**Fig. 1:** Overall description of the fuzzy recognition system.

## 2.1 Input Fuzzification step

The fuzzification step aims to translate variables into linguistic variables [13]. This fuzzification step defines the decomposition number of the considered variable to provide the fuzzy rule premises. The different terms are chosen in relation to the expert vocabulary. The number of terms used to qualify a linguistic variable is generally empirically defined. But, the industrial user, who is not an expert in pattern recognition, often chooses a regular distribution of the terms, generally having more terms than are needed. However, whenever the number of terms increases, so does the number of rules and thus the overall complexity of the entire system. An automatic fuzzification method can also be used. Classical automatic methods are based on Genetic Algorithm [14] or Clustering [15]. But, this kind of methods needs large number of training samples to succeed. Moreover, if the partition of the input variable space is not fit with the real data, the terms and the number of terms will be inappropriate.

The chosen fuzzification method is based on the study of the output class typicality. The typicality measure  $T(V)$  is computed from extern dissimilarity and intern likeness according to the output classes [16]. From the Typicality measure  $T(V)$ , the correlation (Corr) and the cross-correlation (Xcorr) coefficients are computed for each output classes. Then, from the ratio  $\text{Corr}/\text{Xcorr}$ , which characterizes the inter-classes similarity, the number of terms is determined. Their positions are obtained by calculating the mean value of the samples belonging to the considered output classes [10]. The main interest takes place in the automatic adaptation of the fuzzification step which makes the tuning of the system easier.

## 2.2 Fuzzy Rule Generation

This second step allows to define the “If... Then...” fuzzy rules. Each rule describes the perceived defect related to the system. Such rules can be classified into two categories. The conjunctive rules regroup the possibility rules and the anti-gradual rules. The implicative rules regroup the certitude rules and the gradual rules. The conjunctive rules are derived from the data analysis field where reasoning mechanisms are led by the data whereas implicative rules are most utilized in the cognitive sciences field where reasoning is led by knowledge [17]. For this application, the conjunctive reasoning mechanism has been selected. Each rule is activated in parallel and a disjunction operator combines the intermediate results. This inference mechanism gives an interpretation and semantics, which differ from mechanisms using implication. In particular, it assures the consistency of the rule base [18]. If no information is processed that is the input space is not covered by rule set; the output gives an “unknown defect”. The chosen classifier is based on Ishibuchi’s algorithm which provides an automatic rule generation step [19]. There are many methods, which automatically provide fuzzy rules according to data set such as a genetic algorithm [20], but the Ishibuchi’s algorithm is quite simple and gives better results [10]. Moreover, its inference mechanism follows the Larsen model, which is better than the Mamdani model, because the Product is more adapted than the Minimum for the manipulation of several premises [20]. In fact, it allows non-linear cutting of the variables input space. The iterative version of the method [21] is used here because it supports the rule of having the maximum response.

The expert must prepare defective sample sets to generate the fuzzy rules via an automatic rule generation algorithm [18].

## 2.3 Rule Adjustment

The adjustment represents the iterative part of the algorithm. The following mechanism allows to adjust the decomposition of representation space according to achieved results [31]: From the training patterns, the algorithm generates a first model. If the classification rate is below a threshold defined by the user, we make the iterative part to adjust this rate. In fact, we regenerate the fuzzy rules by injecting the training patterns, by considering the new response of each rules and adjusting a confident coefficient. The algorithm proposes an additional refining step. This step allows to improve the membership degree of the maximum membership class by modifying the slope of its membership function. This way is not studied here because the graduality of the answers is needed. This vagueness improves the generalization capability.

### 3 Feature selection from the Choquet integral

#### 3.1 Basic notation

The Choquet integral was first introduced in capacity theory. Let us consider  $m$  classes,  $C_1, \dots, C_m$ , and  $n$  Decision Criteria, denoted DC,  $X = \{D_1, \dots, D_n\}$ . By Decision Criteria a feature description is considered and an associated similarity ratio is produced to ensure that any DC are in the same range, here  $[0, 1]$ . Let  $x_0$  be a pattern. The aim is to calculate for each DC, the confidence degree in the statement ‘‘According to  $D_j$ ,  $x_0$  belongs to the class  $C_i$ ’’. Let  $P$  be the power of  $X$ , a capacity or fuzzy measure  $\mu$ , defined on  $X$ ,  $\mu$  is a set function:

$$\mu: P(X) \rightarrow [0, 1] \quad (1)$$

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 avoiding the additivity axiom. In this application context of the Decision Criteria fusion,  $\mu(A)$  represents the weight of importance, or the degree of trust in the decision provided by the subset  $A$  of DC. The next step in building a final decision is to combine the Choquet integral with the partial confidence degree according to each DC into a global confidence degree. Let  $\mu$  be a fuzzy measure on  $X$ . The discrete Choquet integral of  $\varphi = [\varphi_1, \dots, \varphi_n]^t$  with respect to  $\mu$ , noted  $C_\mu(x)$ , is defined by:

$$C_\mu(\varphi) = \sum_{j=1, n} \varphi(j) \left[ \mu(A_{(j)}) - \mu(A_{(j+1)}) \right] \quad (2)$$

where  $\varphi(1) \leq \dots \leq \varphi(n)$ . Also  $A(j) = \{(j), \dots, (n)\}$  represents the  $[j..n]$  associated criteria in increasing order and  $A(n+1) = \emptyset$ .

#### 3.2 Learning Data

The purpose of this step is to determine the more suitable learning data taking into account the existing confusion between decision criteria. A training pattern yields  $m$  training samples  $\Phi_1, \dots, \Phi_m$ , with  $\Phi_i = (\varphi_{i1}, \dots, \varphi_{im})$  where  $\varphi_{ij}$  represents the confidence in the fact that the sample belongs to class  $i$ , according to DC  $j$ . For each of these samples, a target value must be assigned. For techniques using a different fuzzy measure per class, the optimal target value that minimizes the quadratic error is known [24]. The confusion between classes is estimated by first building the confusion matrix for each DC. Then an average confusion matrix is built by averaging these matrices. Following the global confusion between classes, a decreasing function is defined to take it into account. The more important the confusion is the closer to 0 the value is. Thus, the target value for a sample which is 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. This implies the biggest modification possible.

### 3.3 Learning step

The calculation of the Choquet integral requires the definition of the fuzzy measure, i.e. the assessment of any set of  $P(X)$  which by definition is  $\mu(\emptyset) = 0$ ,  $\mu(X) = 1$ . Several ways to automatically set the  $2^n - 2$  remaining values [22] exist. The main problem is giving a value to the sets having more than three elements while keeping the monotonicity property of the integral. The goal is to find an approximation of the fuzzy measure that minimizes the error criterion. Generally the problem is translated to another minimization problem which is usually solved using the Lemke method. M. Grabisch [9] has shown that such an approach may be inconsistent when using a low number of samples. In this instance - ill-conditioned matrices – the constraint matrix becomes parsed when the set of learning data grows causing undesired behavior of the algorithm. To overcome, these problems, an optimal approach based on gradient algorithm with constraints, which is an extension of Muroshi and Sugeno's method [23], has been proposed in [24]. It assumes that in the absence of any information, the most reasonable way of aggregation is the arithmetic mean, i.e the Choquet integral with respect to an additive equally distributed fuzzy measure. 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 determined path 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. This means that coefficients of the fuzzy measure which are not related to the data are kept as near as possible to the equilibrium point. Thus, this algorithm is still efficient when training data set is limited. It also has a low computing time and a low memory cost.

### 3.4 Indexes

Once the fuzzy measure is learned, it is possible to interpret the contribution of each decision criterion in the final decision. Several indexes can be extracted from the fuzzy measure, helping to analyze the behavior of DC [9].

The importance of each criterion, also called the Shapley index, is based on the definition proposed by Shapley in game theory [25] and is put back into fuzzy measure context by Murofushi and Soneda [9][26]. Let a fuzzy measure  $\mu$  and a criterion  $i$  be considered:

$$\sigma(\mu, i) = \frac{1}{n} \sum_{t=0}^{n-1} \frac{1}{\binom{n-1}{t}} \sum_{\substack{T \subseteq X \setminus i \\ |T|=t}} [\mu(T \cup i) - \mu(T)] \quad (3)$$

The Shapley value can be interpreted as a weighted average value of the marginal contribution  $\mu(T \cup i) - \mu(T)$  of criteria  $i$  alone in all combinations. A property worthy to be noted is that  $\sum_{i=1, n} \sigma(\mu, i) = 1$ . Hence, a DC with an importance index value less than  $1/n$  can be interpreted as a low impact in the final decision. Otherwise an importance index greater than  $1/n$  describes an attribute more important than the average.

The interaction index, also called the *Murofushi and Soneda* index [9][26] represents the positive or negative degree of interaction between two Decision Criteria. 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 combination  $T \subseteq X \setminus ij$  is given by:

$$(\Delta_{ij}\mu)(T) = \mu(T \cup ij) + \mu(T) - \mu(T \cup i) - \mu(T \cup j) \quad (4)$$

After averaging this criterion over all the subsets of  $T \subseteq X \setminus ij$  the assessment of the interaction index of Decision Criteria  $i$  and  $j$ , is defined by (values in  $[-1, 1]$ ):

$$I(\mu, ij) = \sum_{T \subseteq X \setminus ij} \frac{(n-t-2)!t!}{(n-1)!} (\Delta_{ij}\mu)(T) \quad (5)$$

This continues with any pair  $(i, j)$  with  $i \neq j$ . Obviously the index are symmetric, i.e  $I(\mu, ij) = I(\mu, ji)$ . A positive interaction index for two DC  $i$  and  $j$  means that the importance of one DC is reinforced by the second one. In other words, both DC are complementary and their combined use betters the final decision. The magnitude of this complementarily is given by the value of the index. A negative interaction index indicates that the sources are antagonist.

### 3.5 Automatic extraction of subsets of Decision Criteria.

Once the lattice is known, we analyze the individual performance of each DC in the produced fuzzy measure [11]. This analysis is performed using the importance and interaction indexes. The DC having the least influence in the final decision, and interacting the least with the other criteria are assumed blurs the final decision. A two step selection scheme has been implemented to discard such DC. First, the Shapley value is scaled by the number of DC,  $n$ . A DC with a scaled importance index greater than 1 describes a DC more important than the average. The set of low significant criteria  $S_L$  having an importance index lower than 1 is selected:

$$S_L = \{k / n \cdot \sigma(\mu, k) < 1\} \quad (6)$$

Then, the subset of decision criteria having the least positive synergy with the others is extracted from  $S_L$ . For each criterion  $S_{L_i}$ , the values of its interaction with others

are averaged to estimate its global interaction. Finally the subset of criteria to be removed  $MS_L$  is composed of the criteria from  $S_L$  that have an interaction index lower than the mean of the interaction indexes of all criteria of  $S_L$ :

$$MS_L = \left\{ k / \sum_{j=1,n} I(\mu, k_j) < m \right\}, k \in S_L \quad (7)$$

with the global mean interaction index:

$$m = I / |S_L| \sum_{k \in S_L} \sum_{j=1,n} I(\mu, k_j) \quad (8)$$

#### 4 Putting the FRC and the Feature Selection scheme together

The Fuzzy Rule Classifier and the Suitable Feature Selection are embedded in a pattern recognition system. In such a system, a large set of features provide a large amount of fuzzy rules which are hardly exploitable. The selection process aims at decreasing the number of rules by discarding weak parameters while keeping the interesting recognition rates. First, the inference engine is run using initial features. From this first set of features and associated Decision Criteria, a set of learning samples is determined.

Then the fuzzy measure is obtained with respect to the Choquet integral. Indexes are extracted to determine the least representative Decision Criteria. The recognition model is generated without the first least representative features and tested. The process is iterated while it remains less representative features (Fig. 2).

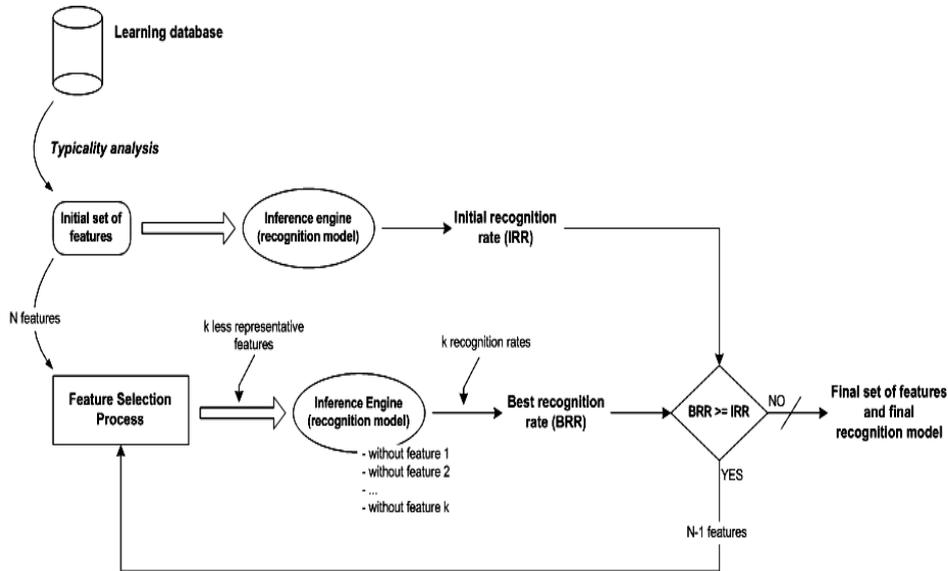


Fig. 2: Fuzzy Rule Iterative Feature Selection Method (FRIFS).

## 5 Experimental applications

The two data sets, which have been tested, correspond to two different fabrics. For confidential reasons, the used features cannot be explicitly named. But, it can be noted that industrialist uses only attribute forms, sizes, and colors which have a signification, i.e. interpretable for a human who must tune the system. The first data set, named Fibre1, contains six output classes (from C1 to C6) with 570 training samples. This set is decomposed in the following way: C1: 5, C2: 118, C3: 274, C4: 82, C5: 34, C6: 57. It can be noted that the class C6 corresponds to invalid defects, i.e. detected defects which are not validated like defects from the customer. Thus this class is strongly heterogeneous. The second data set, named Fibre2, consists of a more extended version of the first data set. It contains the same six output classes with 618 training samples: C1: 12, C2: 188, C3: 230, C4: 131, C5: 2, C6: 55. In both data sets, the classes C2, C3 and C4 allow to consideration for a consistent training step which is not the case for the other classes with, in the worst case, 2 or 5 representative samples. The constraints of the system require the use of simple features which can be quickly calculated. The counterpart of this simplicity is less of a discriminating aspect for some parameters or a redundancy for others. The number of features was initially set to eleven. Local industrial expertise made it possible to delete two parameters which did not bring anything to the system and strongly decreased the results. The remaining parameters are equal to nine. The use of these nine parameters generates a consequent number of rules. In addition, some of them are strongly correlated with others. Table 1 summarizes the results that were obtained for the two data sets (Fibre1 and Fibre2). For these tests, the learning database consists of 33% of data sets. The remaining part (66% of data sets) is used to do the generalization step.

The columns of the Fibre1 data set present the recognition rates obtained by using this data set for the learning step and by applying the generated model to the Fibre2 data set, and inversely for the columns of the Fibre2 data set. Thus, these rates correspond to the generalization of the model. The proposed approach is iterative. It uses the recognition rate as an ending criterion. The value of this criterion has been empirically given to be 90%.

The FRIFS method gives the best results and a better feature selection in comparison with the other methods. The other algorithms stop more quickly (except for the Fibre1 data set with the BFS and FFS methods) than the proposed approach by providing a less satisfactory parameter set. It can be noted that the recognition rates are similar in Fibre 1 for the SBFS, SFFS and FRIFS methods because these methods select the same parameters in the same order.

It can also be noted that the suppression order of the non-relevant parameters is different, not only among the four methods, but also according to the data set chosen for the training step (Fibre 1 or 2). On the other hand, the same four parameters are selected by the FRIFS approach (even 5, if an ending criterion is chosen lower than 90%). It is not the case for the other methods.

Thus, the FRIFS method seems to be more efficient and provides a more stable characteristic vector composed of more significant features

Table 1. Comparison among the different methods of feature selection – Generalization rates obtained with Fibre1 and Fibre2 data sets.

Data Sets		Fibre1				Fibre2			
Methods		SVM	SBFS	SFFS	FRIFS	SVM	SBFS	SFFS	FRIFS
9 features	rate	90.13				92.11			
8 features	with-out rate	P5 89.48	P3 90.94	P3 90.94	P3 90.94	P3 92.63	P3 92.63	P3 92.63	P3 92.63
7 features	with-out rate	P2 86.89	P0 90.29	P0 90.29	P0 90.29	P7 92.46	P5 92.63	P5 92.63	P5 92.63
6 features	with-out rate	P8 88.67	P5 89.48	P5 89.48	P5 89.48	P6 90.00	P8 92.63	P8 92.63	P8 92.63
5 features	with-out rate	P7 81.88	P8 86.73	P8 86.73	P8 86.73	P8 83.51	P2 90.00	P2 90.00	P0 92.81
4 features	with-out rate	P3 81.23	P7 70.07	P7 70.07	P7 70.07	P2 81.05	P8 81.05	P8 81.05	P7 92.81
3 features	with-out rate					P1 74.74	P1 74.74	P1 74.74	P2 90.88
SVM = support vector machine, SBFS = backward feature selection									
SFFS = forward feature selection, FRIFS = Proposed Method									

## 6. Conclusion and future works

The Fuzzy Rule Iterative Feature Selection (FRIFS) method proposed in this article is based on the analysis of a training data set in three steps. The first step, representing the initialization of the method, allows for the choice of a first subset of parameters starting from an analysis of the data typicality. The second and third steps are the iterative parts of the method and reduce the dimension problem while keeping a high recognition rate. The FRIFS approach seems to be more robust because it keeps homogeneity and better stability according to the recognition rates and to the number of selection rules. The experimental results have shown that the proposed method allows the choice of an optimal subset of parameters, increasing the recognition rate in comparison with the choice carried out by expertise and keeping a certain degree of interpretability in the model. Thus, further investigations aim to reduce the number of generated rules. An extension of the proposed FRIFS method aims to analyze each class and not all the training data set, as currently carried out.

## References

1. Kudo, M., Sklansky, J.: Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition* 33 (2000) 25–41.
2. Zhang, H., Sun, G.: Feature selection using Tabu Search method. *Pattern Recognition* 35 (2002) 701–711.
3. Melnik, O., Vardi, Y., Zhang C.H.: Mixed group ranks: preference and confidence in classifier combination. *IEEE Trans. PAMI* 26 (2004) 973-981.
4. Littewood, B., Miller, D.: Conceptual Modeling of Coincident Failures in Multiversion Software. *IEEE Trans. Software Engineering* 15 (1989) 1596-1614.
5. Ho, T.K.: Multiple Classifier Combination: Lessons and Next Steps, *Hybrid Methods in Pattern Recognition*. Kandel, A., Bunke, H., (eds.): World Scientific, (2002).
6. Breiman, L.: Bagging Predictors. *Machine Learning* 24 (1996) 123-140.
7. Schapire, R.E., Freund, Y., Bartlett, P., Lee, W.: Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods. *Annal. Statistics* 26 (1998) 1651-1689.
8. Hadjitodorova, S.T., Kuncheva, L.I., Todorova, L.P.: Moderate diversity for better cluster ensembles. *Information Fusion* 7 (2006) 264-264.
9. Grabisch, M.: The application of fuzzy integral in multicriteria decision making. *Europ. journal of operational research* 89 (1995) 445-456.
10. Schmitt, E., Bombardier V., Charpentier, P.: Self-Fuzzification Method according to Typicality Correlation for Classification on tiny Data Sets. *Proc. 16th Int. Conf. on Fuzzy Systems, FUZZIEEE'07.* (2007) 1072-1077.
11. Rendek, J., Wendling, L.: Extraction of Consistent Subsets of Descriptors using Choquet Integral. in *Proc. 18th Int. Conf. on Pattern Recognition, Hong Kong, Vol. 3* (2006) 208-211.
12. Schmitt, E., Mazaud, C., Bombardier V., Lhoste, P.: A Fuzzy Reasoning Classification Method for Pattern Recognition. *Proc. 15th Int. Conf. on Fuzzy Systems, FUZZIEEE'06.* Vancouver, Canada, (2006) 5998-6005.
13. Zadeh, L.A.: The concept of linguistic variable and its application to approximate reasoning. *Information sciences* 8 (1975) 199-249 ; 301-357.
14. Cordon, O., Gomide, F., Herrera, F., Hoffmann, F., Magdalena, L.: Ten years of genetic fuzzy systems: current framework and new trends. *Fuzzy Sets and Systems* 141 (2004) 5-31.
15. De Carvalho, F.A.T.: Fuzzy c-means clustering methods for symbolic interval data. *Pattern Recognition Letters* 28 (2007) 423-437.
16. Forest, J., Rifqi, M., Bouchon-Meunier, B.: Class Segmentation to Improve Fuzzy Prototype Construction: Visualization and Characterization of Non Homogeneous Classes. *IEEE World Congress on Computational Intelligence, Vancouver, Canada, (2006)* 555-559.
17. Dubois, D., Prade, H.: What are Fuzzy rules and how to use them?. *Fuzzy Sets and Systems* 84 (1996) 169-185.
18. Dubois, D., Prade, H., Ughetto, L.: Checking the coherence and redundancy of fuzzy knowledge bases. *IEEE Trans. Fuzzy Systems* 5 (1997) 398-417.
19. Ishibuchi, H., Nozaki, K., Tanaka, H.: Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets and Systems* 52 (1992) 21-32.
20. Berthold, M.R.: Mixed fuzzy rule formation. *Int. Journal. of Fuzzy Sets and Systems* 32 (2003) 67-84.
21. Ishibuchi, H., Nozaki, K., Tanaka, H.: A Simple but powerful heuristic method for generating fuzzy rules from numeric data. *Fuzzy sets and systems* 86 (1997) 251-270.
22. Grabisch, M., Nicolas, J.M.: Classification by fuzzy integral - performance and tests. *Fuzzy Sets and Systems, Special Issue on Pattern Recognition* 65 (1994) 255-271.

23. Murofushi T., Sugeno M.: A theory of fuzzy measures: representations, the Choquet integral, and null sets. *J. Mathematical Analysis and Applications* 159 (1991) 532-549.
24. Grabisch, M.: A new algorithm for identifying fuzzy measures and its application to pattern recognition. in *Int. Joint Conf. 7th IEEE Int. Conf. on Fuzzy Systems*, Yokohama, Japan, (1995) 145-150.
25. Shapley, L.: A value for n-person games. *Contributions to the Theory of Games*, *Annals of Mathematics Studies*. Khun, H., Tucker, A., (ed.). Princeton University Press (1953) 307-317.
26. Murofushi, T., Soneda, S.: Techniques for reading fuzzy measures(iii): interaction index. in *proc. 9th Fuzzy System Symposium*, Sapporo, Japan (1993) 693-696.
27. Cho C.-S., Chung B.-M, Park, M.-J.: Development of real time vision based fabric inspection system, *IEEE Transactions on Industrial Electronics*, vol.52, no.4, (2005) 1073–1079.
28. Abdulhady M., Abbas H., Nassar S.: Performance of neural classifiers for fabric faults classification, in *proc. IEEE International Joint Conference on Neural Networks (IJCNN '05)*, Montreal, Canada, (2005) 1995-2000.
29. Conci A. and Proena C.: A system for real-time fabric inspection and industrial decision. In *Proc. ACM Int. Conf. on Software Engineering & Knowledge Engineering*, Ischia, Italy, (2002) 707–714.
30. Yang X., Pang G., Yung N.: Fabric defect classification using wavelet frames and minimum classification error training, *37<sup>th</sup> IAS Industry Application Conference*, Pittsburgh, PA, USA, vol 1 (2002) 290–296.
31. Murino V., Bicego M., Rossi I.A.: Statistical classification of raw textile defects in *Proc. of the 17th Int. Conf. on Pattern Recognition (ICPR'04)*, Cambridge, UK, vol 4 (2004) 311-314.