FEATURE SELECTION BASED ON IMPORTANCE AND INTERACTION INDEXES
Hierarchical Fuzzy Rule Classifier Application

Vincent Bombardier[1], Laurent Wendling[2] and Emmanuel Schmitt[1]
[1] *CRAN, CNRS UMR 7039, Université Henri Poincaré, BP 239, 54506 Vandoeuvre-lès-Nancy Cedex, France*

[2] *LIPADE, Laboratoire Informatique Paris Descartes, 75270 Paris Cedex 06, France*

Keywords: Feature Selection, Pattern Recognition, Fuzzy Rules.

Abstract: This paper proposed an extension of an iterative method to select suitable features for pattern recognition context. The main improvement is to replace its iterative step with another criterion based on importance and interaction indexes, providing suitable feature reduced set. This new scheme is embedded on a hierarchical fuzzy rule classification system. At last, each node gathers a set of classes having a similar aspect. The aim of the proposed method is to automatically extract an efficient subset of suitable features for each node. A selection of features is given. The associated criterion is directly based on importance index and assessment of positive and negative interaction between features. An experimental study, made in a wood defect recognition industrial context, shows the proposed method is efficient to producing significantly fewer rules.

1 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 extracted 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 translated as being an optimization problem where a subset of features is searched in order to maximize the classification rate of the recognition system.

Because of 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 specific class because of its rareness). Another difficulty is to respect the real time constraint in the industrial production system. So, low complexity must be kept for the recognition model. Such a classification problem has been relatively poorly investigated in the early years (Abdulhady, 2005), (Yang, 2002), (Murino, 2004). Thus, this work takes place on a “small scale” domain according to (Kudo 2000), (Zhang, 2002) definition because of the weak number of used features.

The Fuzzy Rule Iterative Feature Selection (FRIFS) method proposed in (Schmitt, 2008), is based on the analysis of a training data set in three steps. It combines an original Fuzzy Rule Classifier (Bombardier, 2010) and feature selection associated to capacity learning has been proposed. This approach allows reducing the dimensionality of the problem while keeping a high recognition rate and improving the system interpretability by discarding weak parameters.

First a reference set of features is set and the associated average recognition rate is kept to check the next training. Then an iterative global feature selection process is performed and can be roughly split into two steps:

- Step 1. From the previous set of features, an interactivity process is applied to determine the less representative features.
- Step 2. Generate the recognition model without the first less representative features and test it. The reached recognition rate is stored to be compared to the previous step one.

Bombardier V., Wendling L. and Schmitt E..
FEATURE SELECTION BASED ON IMPORTANCE AND INTERACTION INDEXES - Hierarchical Fuzzy Rule Classifier Application.
DOI: 10.5220/0003672704930496
In *Proceedings of the International Conference on Evolutionary Computation Theory and Applications* (FCTA-2011), pages 493-496
ISBN: 978-989-8425-83-6

And so on while minimization is on the way. A rollback step provides the remaining set of suitable features. This model has shown its ability to efficiently detect fibre defects in industrial application (Schmitt, 2008).

Nonetheless in many industrial applications, it is interesting to group together classes or products following their specificities. A new selection scheme based on hierarchical description is proposed. New sets of features are assessed by a new criterion calculated from both importance and interaction indexes. In this paper, we propose to restrict the iterative part of FRIFS Method (Steps 1 and 2) by directly selecting a pertinent subset of features.

This approach is applied to an industrial pattern recognition problem. The aim is to identify wood singularities. Comparisons with well-known selection methods like Sequential Backward Feature Selection (SBFS) or Sequential Forward Feature Selection (SBFS) methods, attest of the good behaviour of our method.

2 SELECTION OF SUITABLE SUBSET OF FEATURES

2.1 Hierarchical Description

Feature selection based on Choquet integral (Grabisch, 1994) provided a suitable set of features to the Fuzzy Rules Classifier (Bombardier, 2010). As said previously, the rules are then obtained by learning. This principle is extended here to process with a hierarchical description associated to classes.

For each node, a set of suitable features is extracted from an analysis of their degree of importance combined with their level of positive and negative interactions (see next section). The selected subset directly depends on the recognition rates achieved by considering independently each node with the FRC and so on in order to process with the whole description.

2.2 Capacity 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 (Grabisch, 1995). The importance of each criterion is based on the definition proposed by Shapley in game theory (Shapley, 1953). Let a fuzzy measure μ and a criterion D_i be considered:

$$\sigma(\mu, D_i) = \frac{1}{n}\sum_{t=0,n-1} \frac{1}{\binom{n-1}{t}} \sum_{\substack{T\subseteq N\setminus D_i \\ |T|=t}} \left[\mu(T\cup D_i) - \mu(T)\right] \qquad (1)$$

The Shapley value can be interpreted as a weighted average value of the marginal contribution $\mu(T\cup D_i) - \mu(T)$ of criterion Di alone in all combinations.

The interaction index, also called the Murofushi and Soneda index (Murofushi, 1993), (Rendek, 2006) 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 Di and Dj, conditioned to the presence of elements of combination $T \subseteq X\setminus D_iD_j$ is given by:

$$I(\mu, D_iD_j) = \sum_{T\subseteq N\setminus D_iD_j} \frac{(n-t-2)!\,t!}{(n-1)!}\left(\Delta_{D_iD_j}\mu\right)(T) \qquad (2)$$

With:

$$\left(\Delta_{D_iD_j}\mu\right)(T) = \mu\left(T\cup D_iD_j\right) + \mu(T) - \mu(T-D_i) - \mu(T-D_j)$$

And so on, considering any pair (D_i,D_j) with $i \neq j$. Obviously the index are symmetric, i.e $I(\mu,D_iD_j)=I(\mu,D_jD_i)$. A positive interaction index for two DC D_i and D_j means that the second one reinforces the importance of one decision criterion.

In other words, both DC are complementary and their combined use betters the final decision. The behaviour is given by the value of the index. A negative interaction index indicates that the sources are antagonist.

2.3 Node Features

The aim is to find a suitable subset of features for all the nodes. Each node is assumed to be a cluster of classes having similar characteristics. Considering the whole set of features it is obvious that the set of suitable parameters for one node could be not the same for another one. The processing time relies on the cardinality of the considered hierarchical description. A selection scheme is introduced to decrease processing complexity while quickly focusing suitable sets of features. The method relies on a combination of both indexes previously described.

A Shapley value property is $\Sigma_{i=1,n}\ \sigma(\mu, D_i)= 1$. Generally values are multiplied by a number of decision criteria $n=|N|$. Hence, a DC with an importance index value less that 1 can be interpreted as a low impact in the final decision. Otherwise an importance index greater than 1 describes an attribute more important than the average that is $F = \{D_i \in N / n \times \sigma(\mu, D_i) \geq 1\}$, the weaker: $\overline{F} = \{N \setminus F\}$

Among them, some decision criteria may impact the recognition by having positive interactions with selected important criteria. Then, the DCs having an interaction of order 2 greater to the mean are selected:

$$G = \left\{ D_i \subseteq \overline{F} / \frac{1}{|\overline{F}|} \sum_{D_j / i \neq j} \Delta(\mu, D_i D_j) \geq \frac{1}{|F|^2} \sum_{\substack{D_t, D_v \subseteq F^2 \\ t \neq v}} \Delta(\mu, D_t D_v) \right\}$$

At last final set $N' = F \cup G$ is sent to the Fuzzy Rule Classifier to determine if it improves or not the recognition using the same learning samples. If the recognition rate is better than previous epoch one, the process is run again with N' instead of N. Otherwise the previous set of features is kept (rollback step). This process is run until all the nodes are studied.

3 EXPERIMENTAL STUDY

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 singularities to be recognized are intrinsically fuzzy (gradual transition between clear wood and knots for instance). The features extracted from the images are thus uncertain (but precisely calculated) and the use of fuzzy logic allows taking this 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 classes representing a small knot and the class representing a large knot). Finally, the customer needs and the human operator experience are subjective and mainly expressed in natural language.

3.1 Wood Defect Recognition Case

The results presented in this section are based on a real set of samples collected from a wood industrial case. The objective of this application is to develop a vision system for singularity identification on wooden boards used to estimate the quality of the final products.

During the image segmentation, a set of features is calculated on the achieved regions to provide a characteristic vector to the recognition step. This set is composed with geometrical (SURF, MIN_AXIS, MAJ_AXIS…) and topological (C1, C2, C3…) features but we cannot make them explicit because of confidentiality. We can note those calculated features are rather basic and simple due to the real time industrial constraints. So, associated values are redundant and often contradictory bringing noise to the final decision.

3.2 Results and Discussion

The classification is done with the features issued from the segmentation stage performed by the industrialist. The database is composed of 877 samples divided in nine classes of wood singularities (called nuodo muerto, grieta, medula, resina …).

The learning database, used to compute the learning recognition rates, is composed of 250 samples. The generalisation rates are obtained with the generalisation database constituted by the 627 remaining samples. These databases are relatively heterogeneous, for instance, the 250 samples of the learning database are composed of 8, 56 7, 7, 18, 47, 5, 93, 9 samples of the nine classes. 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. Let us consider the hierarchical description provided in Figure 1.

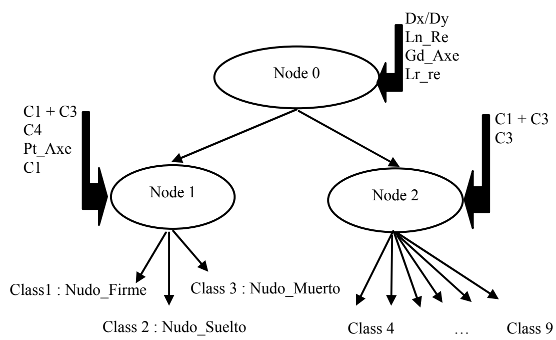

Figure 1: Example of hierarchical description.

Tests aim to reduce the dimensionality by removing non-efficient features per node until reaching an interpretable model while keeping a “good” recognition rate. Three selection methods are used: SBFS, SFFS and EXP. Two of them are automatic feature selection methods used as references (Pudil, 1994). The third one is an expert method where the feature sets have been manually determined.

Each selected feature set is provided for input to the FRC. This classifier is used in a hierarchical version rather than a single node version. Here, the structure is composed of three nodes as shown in figure 1. The main advantages are to reduce the total number of rules of the system, as the rule set of each node is smallest and also to enhance the recognition rate. Table 1 shows the learning rates (LR), the generalisation rates (GR), the number of features (#Param.) and the number of generated rules (#Rules) obtained for five methods.

Despite the proposed method provided better results considering separately each node, the recognition rates obtained by SFFS, SBFS and the proposed method (PM) are comparable. Scores between 97% and 98% are reached for the generalization step considering the whole database. The expert selection (EXP) gives rise to a rate around 96%. The difference is insignificant.

However, the number of parameters extracted by our method is the lowest assuming the better interpretability of the model. The numbers of rules are carried out using expert (EXP: 3775 rules) and the Proposed Method (PM: 1375 rules). Other methods lead to a rule number higher than 350 000.

Table 1: Recognition rates and Generated rules.

		EXP	SFFS	SBSF	PM
Node 0	LR	96.4%	96.%	96,00%	96.4%
	GR	90.59%	91.38%	91.39%	90.59%
	#Param	4	2	2	4
	#Rules	625	**25**	**25**	625
Node 1	LR	78.771%	94.97%	90.50%	78.77%
	GR	72.236%	77.64%	74.69%	72.24%
	#Param	**5**	9	8	**5**
	#Rules	3125	1953125	390625	3125
Node 2	LR	100.%	100.%	100.%	100.%
	GR	96.818%	**97.27%**	**97.27%**	96.82%
	#Param	2	3	3	2
	#Rules	**25**	125	125	**25**

4 CONCLUSIONS

An enhancement of a Fuzzy Rule Iterative Feature Selection method has been presented. It allows the decreasing of learning time processing while focusing on relevant samples. A suitable set of features is obtained. Then, FRIFS method is adapted to handle with a hierarchical fuzzy inference system. Industrial real-data tests show the efficiency of the proposed method. The recognition rate is similar to other methods but the number of features significantly decreases and thus the number of rules too. Actually, the extension of our model to provide selection of parameters per class is under consideration.

REFERENCES

Abdulhady, M., Abbas, H., Nassar, S., 2005. Performance of neural classifiers for fabric faults classification. In *proc. IEEE International Joint Conference on Neural Networks (IJCNN '05)*, Montreal, Canada, 1995-2000.

Bombardier, V., Schmitt, E., 2010. Fuzzy rule classifier: Capability for generalization in wood color recognition. In *Eng. Appli. of Artificial Intelligence*, v23, 978-988.

Grabisch, M., Nicolas, J. M., 1994. Classification by fuzzy integral - performance and tests. In *Fuzzy Sets and Systems*, v65, 255-271.

Grabisch, M., 1995. The application of fuzzy integral in multicriteria decision making. In *Europ. journal of operational research,* v89, 445-456.

Kudo, M., Sklansky, J., 2000. Comparison of algorithms that select features for pattern classifiers. In *Pattern Recognition,* v 33, 25–41.

Murino, V., Bicego, M., Rossi, I. A., 2004. Statistical classification of raw textile defects. *In Proc. of the 17th Int. Conf. on Pattern Recognition (ICPR'04),* Cambridge, UK, 311- 314.

Murofushi, T., Soneda, S. 1993. Techniques for reading fuzzy measures(iii): interaction index. In *proc. 9th Fuzzy System Symposium*, Sapporo, Japan, 693-696.

Pudil, P., Novovicova, J., Kittler, J., 1994. Floating search methods in feature selection. In *Pattern Recognition Letters*, v15, 1119–1125.

Rendek, J., Wendling, L. 2006. Extraction of Consistent Subsets of Descriptors using Choquet Integral. In *Proc. 18th Int. Conf. on Pattern Recognition*, Hong Kong, 208-211.

Shapley, L., 1953. A value for n-person games. Contributions to the Theory of Games. In *Annals of Mathematics Studies*. Khun, H., Tucker A., Princeton University Press 307-317.

Schmitt, E., Bombardier, V., Wendling, L., 2008. Improving Fuzzy Rule Classifier by Extracting Suitable Features from Capacities with Respect to the Choquet Integral. In *IEEE trans. On System, man and cybernetics* v38-5 1195-1206.

Yang, X., Pang, G., Yung, N., 2002. Fabric defect classification using wavelet frames and minimum classification error training. *In 37th IAS Industry Application Conference,* Pittsburgh, USA, 290–296.

Zhang, H., Sun, G. 2002. Feature selection using Tabu Search method. In *Pattern Recognition,* v 35 701–711.