A selection approach for scalable fuzzy integral combination

P. Bulacio, Serge Guillaume, E. Tapia, L. Magdalena

▶ **To cite this version:**

P. Bulacio, Serge Guillaume, E. Tapia, L. Magdalena. A selection approach for scalable fuzzy integral combination. Information Fusion, Elsevier, 2010, 11 (2), 6 p. hal-00490251

HAL Id: hal-00490251

https://hal.archives-ouvertes.fr/hal-00490251

Submitted on 8 Jun 2010

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.

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/

doi:10.1016/j.inffus.2009.06.003

A selection approach for scalable fuzzy integral combination

P. Bulacio [a], S. Guillaume [b], E. Tapia [c], L. Magdalena [d]

[a] *ETSI Telecomunicacion, Universidad Politecnica de Madrid, Spain*

[b] *Cemagref Montpellier, France*

[c] *Universidad Nacional de Rosario, Argentine*

[d] *European Centre of Soft Computing, Spain*

Abstract

We consider the problem of collective decision-making from an arbitrary set of classifiers under Sugeno fuzzy integral (S-FI). We assume that classifiers are given, i.e., they cannot be modified towards their effective combination. Under this baseline, we propose a selection-combination strategy, which separates the whole process into two stages: the classifiers selection, to discover a subset of cooperative classifiers under S-FI, and the typical S-FI combination of selected classifiers. The proposed selection is based on a greedy algorithm which through a heuristic allows an efficient search.

Key words: Multiclassifier scalability, Fuzzy integral, Greedy selection
PACS:

1 Introduction

Multiclassifier systems aim to enhance the performance of any single classifier. Although there are many ways to use more than one classifier, all of them requires the cooperation among classifiers, i.e., classifiers specifically combined do not propagate individual mistakes to collective results. Clearly, cooperation is only possible if classifiers make errors in different samples, which can be easily achieved with specialized classifiers. However, in the most general case, i.e., non specialized classifiers, the cooperation must be induced [1,5,8] or exploited [6].

The design of multiclassifier systems usually involves two steps [9]: the generation of classifiers, and their combination. In general, the first step creates a set

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

of classifiers including their cooperation ability towards a later combination. However, the set of classifiers may be given and just the combination stage can be done. In this latter case, the cooperation can be merely exploited without altering the classifier behavior. A typical example of this situation is focused in this paper: the requirement of a single decision-making from a population of classifiers that were not adapted to the collective work.

To guarantee effective results under the above condition, untrained combination rules are useless since collective generalization strength must be characterized. That means that the combination process should handle knowledge about collective skills of a possible numerous set of classifiers. Clearly, this knowledge induction can be extremely complex, i.e., it requires the behavior characterization of each classifier subset. Therefore, the treatment of a general population of classifiers entails an alternative design strategy.

From the above discussion, the characterization complexity of a given population of classifiers is separated into two trained and complementary processes: *selection and combination* (Fig. 1). The selection should reduce the initial set

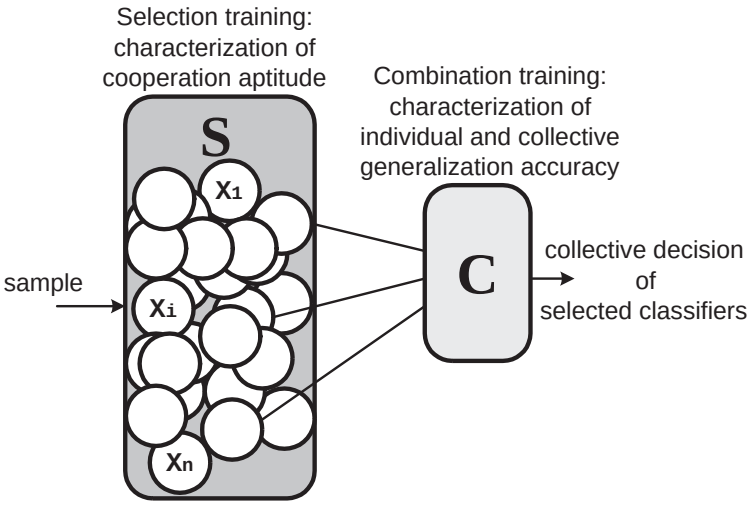


Fig. 1. Selection-Combination strategy.

to a tractable subset of cooperative classifiers. We desire both efficient and effective solutions. Regarding efficiency, exhausted searches are discarded. We suggest a heuristic search guided by a cooperation ability index. This index values the cooperation aptitude among classifiers under a specific combination rule by a rough analysis in the whole set of classifiers. After the selection, the combination takes place. Regarding effectiveness, the combination should be able to make a deeper characterization of collective behavior of the selected subset. The Sugeno integral allows such description thanks to a simple, but powerful, combination mechanism which takes into account, by the means of a fuzzy measure, the collective generalization strength.

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

The paper is organized as follows: Section 2 looks into the selection-combination strategy based on Sugeno fuzzy integral. Section 3 is dedicated to show the experiments on both benchmark data from the UCI repository and a new application. Finally, the main conclusions are highlighted in section 4.

2 Selection-combination strategy

The selection of cooperative classifiers should address the following questions: 1) Which are the matters associated with the effective work under the posterior and known combination?, and 2) How to reach efficiently an effective subset? Considering the Sugeno FI combination, its behavior must be analyzed to answer the first question. To answer the second one, a heuristic selection that exploits the information of the former step, based on greedy algorithms is suggested.

2.1 Sugeno FI combination

The fuzzy integral is a general trained combination methods. Its definition w.r.t. a *fuzzy measure* [7] provides a good framework to represent imprecise knowledge associated with the behavior of classifier subsets. See [4] for details. We focus on Sugeno FI assuming a given population of classifiers, $X = \{X_1, ..., X_i, ..., X_n\}$, which associate each input s with $\{w_1, ..., w_c\}$ possible classes. The classification function associated with i-th classifier is $f_i : s \rightarrow [0,1]^c$. The f_i components $(f_i^1, ..., f_i^c)$ can be interpreted as degrees of support of the i-th classifier to each class prediction.

Collective FI results are obtained by aggregating levels of decision where classifiers agree with collective abilities (g) of classifiers that support them. These classifier abilities represent the generalization strength which are characterized by fuzzy measures.

A set function $g : 2^X \rightarrow [0,1]$ is a *fuzzy measure* if it satisfies the following conditions:

(1) $g(\emptyset) = 0$, $g(X) = 1$ (boundary conditions).
(2) $A \subseteq B \quad \Rightarrow \quad g(A) \leq g(B)$ (monotonicity) for $A, B \in 2^X$.

The Sugeno integral [7] of a function $f : X \rightarrow [0,1]$ w.r.t. g on $(X, 2^X)$ is defined by

$$S_g(f) := \max_{i=1}^{n} \{min(f(X_{(i)}), g(A_{(i)}))\} \tag{1}$$

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

(i) indicates the permutation of indices, $0 \leq f(X_{(1)}) \leq \ldots \leq f(X_{(n)}) \leq 1$, $f(X_{(0)}) := 0$, and $A_{(i)} := \{X_{(n)}, \ldots, X_{(n)}\}$. The measure $g(A_{(i)})$ (or $g^{(i)}$ for short) quantifies the generalization ability of the subset $A_{(i)}$. In particular, the Sugeno integral for the class w_j is: $S_g(f^j) := \max_{i=1}^{n}\{min(f^j_{(i)}, g^{(i)}_j)\}$.

2.1.1 Behavior of Sugeno integral

When the first point of collective classification design is the construction of classifiers, their collective behavior can be induced. However, when they are externally given, the collective behavior must be carefully analyzed and characterized during the multiclassification procedure.

The collective behavior of classifiers under S-FI depend on f and g values, i.e., the relationship among classification decisions of the current sample and the characterization value of generalization ability, determines the final decision.

(1) The final decision is defined by just one classifier. This happens when there is a classifier with clear decisions (f_i) joint with strong ability measures (g^i), i.e., the minimum among a classifier decision and its fuzzy density is bigger than the rest of fuzzy densities. The corresponding classifier is named *predominant*.
(2) The final decision is collectively defined. This situation is presented when there is no classifier that prevails in its decision-ability relation for over the others. So, the final result depends on the collective generalization ability of classifiers that consents on different levels of decisions.

Clearly, both situations show that S-FI can be successful with a correct predominant or with a correct consensus. Under these conditions, the selection looks for those classifiers that maximize the number of well classified samples in the training dataset, taking into account the *f-g* relationship.

2.2 Selection process

The proposed selection is based on a heuristic search by means of a greedy algorithm. In other words, the selection of classifiers that can cooperate are computed based on a single criterion (selection rule), instead of having a recursive analysis over any of the alternative options or its effect on further steps. The selection process starts with an empty set and seeks to include it the most cooperative candidates. If the best candidate does not improves the already selected (stop rule), the selection ends. Otherwise, the process is repeated until there are no more classifiers to add.

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

being A a set of candidates from which a selection is done, its overall design involves iterations over the followings steps:

(1) The classifiers selection. It is performed by applying a selection rule, which chooses at each decision point the *best* candidate for working with the already selected classifiers.
(2) The selection end. It is evaluated through a stop rule, which determines the contribution of new candidates and so, the algorithm cut.

2.2.1 Selection rule

The set of selected classifiers (set O) starts as an empty set that is extended with the best candidate (X_r^*) at each selection step. X_r^* is determined from the analysis of each extended subset, $O_r = \{X_r \cup O\}$ with $r = 1, ..., n_r$, being n_r the amount of candidates. With this aim, the selection knowledge completes the S-FI behavior description: while S-FI knowledge handles a full description of collective behavior (2^X subsets) at class level, the selection knowledge characterizes a simplified view of collective behavior but at sample level. The selection picks the candidate that exploits the cooperation under S-FI applying the following selection rule.

X_r^* maximizes on the training dataset Z:

(1) The *coverage index* that evaluates the minimal condition to achieve correct collective results: at least one classifier of O_r must be correct.
(2) The *f-g relationship index* that evaluates the possibility of correct predominant or consensus by the relation among decisions-abilities of O_r classifiers.

In order to facilitate the *coverage* and *f-g relationship* study, the matrices of *decision pattern* F and *error pattern* E on $Z = \{z_k\}$ $(k = 1, ..., K)$ are analyzed.

$$F, E = \begin{array}{c} \\ z_1 \\ \vdots \\ z_k \\ \vdots \\ z_K \end{array} \begin{array}{c} \begin{array}{ccccccc} F_1, E_1 & \cdots & F_i, E_i & \cdots & F_n, E_n \end{array} \\ \begin{pmatrix} f_1(z_1), 0 & \cdots & f_i(z_1), 1 & \cdots & f_n(z_1), 1 \\ \vdots & & & \vdots & \\ & & f_{k,i}, e_{k,i} & & \\ \vdots & & & \vdots & \\ f_1(z_K), 1 & \cdots & f_i(z_K), 0 & \cdots & f_n(z_K), 0 \end{pmatrix} \end{array}$$

While $e_{k,i} = 0$ means correct classification and $e_{k,i} = 1$ implies error, $f_{k,i}$ is the decision vector of X_i on the sample z_k associated with the class w_j. The matrices E and F encloses a complete classifier generalization description;

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

several *diversity* and *accuracy* measures [2,3] can be computed from it. From their horizontal and vertical scanning, the collective behavior on the dataset can be examined: the vertical scanning shows the individual generalization strength on Z, and the horizontal scanning shows the collective behavior per sample.

Coverage index of X_r, B_r: It is computed as the average coverage on Z of classifiers of O_r, being the coverage per sample:

$$b_{k,r} = \begin{cases} 0, \text{ if classifiers of } O_r \text{ have a common error on } z_k; \\ 1, \text{ if at least a classifiers of } O_r \text{ is correct on } z_k. \end{cases}$$

$b_{k,r}$ values are initialized with the error pattern of the first selected classifier. B_r (with $r = 1, \ldots, n_r$) is the fraction of covered samples on Z, i.e., the proportion of "ones" of $b_{k,r}$, with $k = 1, \ldots, K$.

***f-g* relationship value of X_r, FG_r:** It is computed as the average on Z of the coverage strength or consensus of each candidate X_r, being:

- the coverage strength per sample z_k, the maximal correct decision (class w_j) of O_r members ponderated by the generalization ability.

$$s_{k,r} = \max_{q=1}^{Q_r} \{\min(f_{q,k}^j, g_j^q)\} \tag{2}$$

being Q_r the cardinality of O_r.

- the consensus per sample z_k, the average of correct decision values (class w_j) of O_r members ponderated by the generalization abilities.

$$c_{k,r} = \frac{1}{Q_r} \sum_{q=1}^{Q_r} f_{q,k}^j \times g_j^q \tag{3}$$

The selection of X_r^* gives priority to the classifier that maximizes values of decisions-abilities when it is correct, and positive consensus when it is wrong. Based on the above characterizations, a **vector of *f-g* characterization** fg_r of X_r is built.

$$fg_{r,k} = \begin{cases} s_{k,r} \text{ if } X_r \text{ is correct in } z_k; \\ c_{k,r} \text{ if } X_r \text{ is wrong in } z_k. \end{cases}$$

For each candidate, the *f-g relationship value* FG_r is valued as the mean of the $(fg_{r,k})$ components.

Selection rule: X_r^* is the candidate that achieves with the already selected ones, both the major coverage and f-g relation on Z, $X_r^* \leftrightarrow \max_{r=1}^{n_r} \{B_r + FG_r\}$

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/

doi:10.1016/j.inffus.2009.06.003

2.2.2 Selection end

A new X_r^* becomes a member of O whenever its selection contributes to the combination. To decide its inclusion, a collective performance ($P_r^*(Z)$) of $O_r^* = \{X_r^* \cup O\}$ is estimated. With this aim, the existence of predominant classifiers is determined. Samples store the *f-g* relationship, initially, the highest value (associated with w_m class) of minimum *f-g* relation, $min(f_t^m; g_m^t)$. This value is compared with the highest decision (f_r^{m*}) of the coming classifier. The following cases can be presented:

(1) If the $min(f_t^m; g_m^t) > f_r^{m*}$ then X_t continues predominating.
(2) If the $min(f_t^m; g_m^t) < f_r^{m*}$ we can have:
   - If $min(f_r^{m*}; g_{m*}^r) > f_t^m$ then X_r^* predominates
   - If $min(f_r^{m*}; g_{m*}^r) < f_t^m$ then there are no predominant. An estimation of the collective performance of O_r^*, such as weighted vote with f, g measures, is required .

If the candidate X_r^* predominates in z_k, the values of the sample characterization are updated by those of X_r^*; in addition, $P_r^*(z_k)$ is directly evaluated by comparing w_{m*} with the real class w_j. Otherwise, collective performance is estimated.

2.2.3 Selection algorithm

The main input are the matrices E and F of the given set of classifiers. They are evaluated using *ten*-fold cross-validation on Z, by appending the tenth parts of each fold.

Process beginning: Given are $E_{K\times n}$, $F_{K\times n}$.

(1) Evaluate the individual accuracy of classifiers and select the most accurate as the initial member of O, denoted by X_b.
(2) Evaluate the **coverage index** $B_r = \frac{1}{K}\sum_{k=1}^{K} b_{k,r}$ of each X_r with $r = 1, \ldots, n_r$ and with $k = 1, \ldots, K$.
(3) Evaluate the **f-g relationship index** FG_r of each X_r according to $fg_{r,k}$ sample values.
(4) Choose X_r^* applying the **selection rule**: $X_r^* \leftrightarrow \max_{r=1}^{n_r}\{B_r + FG_r\}$.
(5) Evaluate the **selection end rule**: P_r^* to decide the X_r^* inclusion:
   `IF` $(P_r^* < P(O)\cdot\alpha)$
   `THEN` stop selections.
   `ELSE` $O = \{O \cup X_r^*\}$ and GOTO 2.

In the first step, the most accurate classifier (X_b) is included in O. In this way,

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

an initial subset is defined and its greedy augmentation starts. In addition, coverage values $b_{k,r}$ are initialized with its error pattern.

The selection of X_r^* is done according to the potential cooperation among the already selected classifiers and the remaining ones. The cooperation is evaluated using measures of coverage and *f-g* relationship. These measures are computed using the given *error pattern* and *decision pattern* matrices. B_r is an optimist estimation of collective error distribution if the candidate X_r were included in O; a zero entry in $b_{k,r}$ means that at least one classifier of O_r classifies correctly the row sample. Additionally, the *f-g* relationship characterizes the strength of the candidate contribution depending on its levels of decisions and generalization abilities on Z dataset. A highest level of correct *f-g* relationship of some classifier of O as well as the high positive consensus may give a correct sample classification even if the new candidate is mistaken.

The selection process continues until the collective performance drops. The parameter α prevents the method for possible staking, especially at the beginning where the best classifier could reject further inclusions. We should note that the cooperation is sometimes impossible, e.g., when one classifier is much better than others. In that case the combination is not proper and the use of the best classifier is better.

3 Experiments

We evaluate selective multiclassfiers on benchmark UCI and real data. Without loss of generality, we considered population of 30 classifiers based on neural networks [1] (NN) and fuzzy inference systems [2] (FIS) trained in an automatic way over the whole output space, i.e., they were not adapted to any combination rule. The configurable parameters for their training are the following:

NN parameters

- *Neural Net structure:* Three layers, the first with a number of neurons equal to the number of input variables; the last two layers with a quantity of neurons equal to the number of classes.
- *Weight update:* The update algorithm is backpropagation, taking blocks of [1;10] examples for the updating.
- *Epochs:* The number of epochs is taken from the interval [50;500] in random form.

[1] http://fuzzy.cs.uni-magdeburg.de/~borgelt/software.html
[2] http://www.inra.fr/Internet/Departements/MIA/M/fispro/

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The official publication is available at http://www.sciencedirect.com/

doi:10.1016/j.inffus.2009.06.003

- *Input variables partitions:* In the range [1;5].
- *Partition induction from data:* The used algorithms are Hierarchical Fuzzy Partitioning (HFP); Regular Partitioning or Kmeans.
- *Rule induction:* The used algorithms are Fast Prototyping Algorithm (FPA), Wang and Mendel (W&M) and FDT (Fuzzy Decision Trees).

Regarding the training of the Sugeno FI, we use λ-measures. Fuzzy densities are determined (per classes) as the proportion of

$$g_j^i = P(z_k \in w_j / f_i^j = max\{f_i(z_k)\}) - P(z_k \notin w_j / f_i^j = max\{f_i(z_k)\}) \quad (4)$$

Being $P(z_k \in w_j / f_i^j = max\{f_i(z_k)\})$ the proportion of correct classification in the class, and $P(z_k \notin w_j / f_i^j = max\{f_i(z_k)\})$ the "false ones" in the others. The applied protocol to datasets uses random sampling techniques to generate 10 independent experiences. Each experience has the Z with 75% of the total samples and the validation set with the last 25%. Cross-validation technique is applied on Z to evaluate the error patterns of classifiers and the fuzzy densities.

Benchmark datasets: Table I shows the characteristics of six data sets from UCI [3] repository.

Dataset	Samples	#Attributes	#Classes
Car	1728	6	4
Glass	214	10	6
Iris	150	4	3
Pima	768	8	2
Wine	178	13	3
Yeast	1484	8	10

Table 1
Description of used datasets

Real dataset: The objective is to determine the *grape* variety from 8 input variables. These variables are expert selected wavelengths, due to their physical meaning, from a 512 wavelength near infrared spectrum.

The dataset consists of 50 examples for each grape variety. The output space is composed of 8 classes: carignan, grenache blanc, chardonnay, roussane, marselan, mourvèdre, grenache noir and clairette.

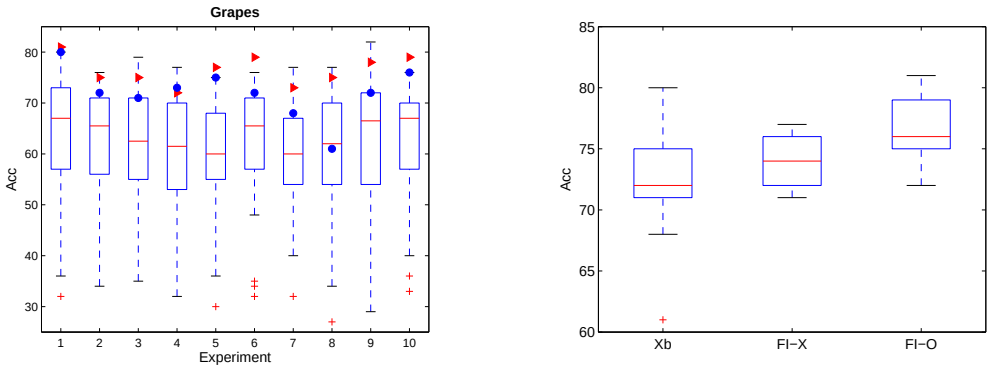
[3] http://www.ics.uci.edu/~mlearn/MLRepository.html

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

Dataset	X_b	SD_b	FI_X	SD_{FI_X}	$FI-S$	SD_{FI-S}
Car	94.28	1.50	92.36	2.32	96.09	1.09
Glass	90.00	4.80	84.07	4.72	90.93	3.08
Iris	95.00	2.90	96.05	2.24	96.05	2.34
Pima	76.90	2.34	78.18	1.91	77.29	2.10
Wine	97.33	2.29	97.77	2.10	97.56	1.95
Yeast	56.66	1.75	51.35	3.36	57.95	1.76

Table 2
X_b results, full combination, and selection-combination strategy



(a) **(b)**

Fig. 2. Ten experiences on Grapes data. (a) The accuracy of individual of X (boxplot), the X_b (circle) and the selection S-FI combination accuracy (triangle); (b) Comparison of X_b, total combination (FI-X), and selection S-FI combination (FI-O)

Fig. 2 (a) shows on each one of the ten experiment the performance diversity among classifiers of X, and the comparison among the best classifier and the selection-combination strategy by Sugeno FI. Fig. 2 (b) summarizes the X_b, the complete combination of X, and the selection-combination strategy performance.

Dataset	X_b	SD_b	FI_X	SD_{FI_X}	$S-FI$	SD_{S-FI}
Grapes	72.00	5.03	74.00	2.16	76.4	2.88

Table 3
X_b results, full combination, and selection-combination strategy for grapes dataset

Tables 2 and 3 summarize the validation results. First columns are the performance of X_b and its standard deviation (X_b, SD_b), and last ones are the performance with Sugeno fuzzy integral (FI-S, SD_{FI-S}).

Let us underline the two types of improvements due to the combination: the mean value of classification accuracy, and its concentration. As a result, the

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

combination is more robust than any single classifier.

Overall, the proposed method outperform the X_b. In addition, we should note that the generated population has classifiers which are near to the maximum rate of classification. As expected, improvements seem to be slighted when the best classifier is extremely good. However, when X_b is far to the maximum accuracy, the effectiveness of collective analyze increase its chances.

4 Conclusions

Aiming the practical FI combination, an efficient selection-combination strategy was proposed. We consider the problem of an efficient and effective decision making from a given population of classifiers. The efficiency was achieved by means of greedy algorithms which reduce the initial aggregation complexity to a selection of an effective subset for its posterior combination. The effective solution is attained thanks to a heuristic search that takes into account the fuzzy integral behavior.

The effectiveness was proved on benchamark and a real dataset. Experimental results suggest that this methodology is particular well-suited when best classifiers are far to the maximum accuracy, and so, the collective analyses have more chances of enhancement. Particularly, the improvement on the grapes dataset is achieved from individuals with a large proportion of errors, but its complementary errors per classes is well profited by the FI.

References

[1] G. Brown, J.L. Wyatt, R. Harris, and X. Yao, *Diversity creation methods: A survey and categorisation*, Information Fusion, Elsevier pub. **6(1)** (2005), 5–20.

[2] L.K. Hansen and P. Salamon, *Neural network ensembles*, IEEE Transactions on Pattern Analysis and Machine Intelligence **12** (1990), 993–1001.

[3] L.I. Kuncheva and C.J. Whitaker, *Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy*, Machine Learning **51(2)** (2002), 181–207.

[4] T. Murofushi and M. Sugeno, *Fuzzy measures and fuzzy integrals*, Fuzzy Measures and Integrals. Theory and Applications (Michel Grabisch, Toshiaki Murofushi, and Michio Sugeno, eds.), Physica Verlag, Heidelberg, 2000, pp. 3–41.

[5] D. Partridge and W. B. Yates, *Engineering multiversion neural-net systems*, Neural Comput. **8** (1996), no. 4, 869–893.

Author-produced version of the article published in
Information Fusion
Volume 11, Issue 2, April 2010, Pages 208-213.

The original publication is available at http://www.sciencedirect.com/
doi:10.1016/j.inffus.2009.06.003

[6] F. Roli and G. Giacinto, *Design of multiple classifier systems*, Hybrid Methods in Pattern Recognition (H. Bunke and A. Kandel, eds.), World Scientific Publ. Co., 2002, pp. 199–226.

[7] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. thesis, Tokio Institute of Technology, 1974.

[8] G. Valentini and F. Masulli, *Ensembles of learning machines*, in M. Marinaro and R. Tagliaferri, editors, Neural Nets WIRN Vietri-02, Series Lecture Notes in Computer Sciences, Springer-Verlag, Heidelberg (Germany) **2486** (2002).

[9] L. Xu, A. Krzyzak, and C. Y. Suen, *Methods of combining multiple classifiers and their applications to hand-written character recognition*, IEEE trans. on Systems, Man and Cybernetics **22(3)** (1992), 418–435.