CACIC 2005 - XI Congreso Argentino de
Ciencias de la Computación
Workshop de Agentes y Sistemas Inteligentes (WASI)

Adaptive Decision Making Systems

Pilar Bulacio, Luis Magdalena
ETSI Telecomunicación
Universidad Politécnica de Madrid
Spain
{pb,llayos}@mat.upm.es

Elizabeth Tapia
Escuela de Ingeniería Electrónica
Universidad Nacional de Rosario
Argentine
etapia@eie.fceia.unr.edu.ar

Abstract

Given a population of classifiers, we consider the problem of designing highly compact and error adaptive decision making systems. A selection approach based on misclassification diversity and potential cooperation among classifiers is proposed. The compactness constraint allows us the efficient implementation of fuzzy integral combination rules regarding both the interpretability of fuzzy measures and low complexity of fuzzy integral operator. Experimental results show the feasibility of our approach.

Keywords: Scalability, multiclassifier system, fuzzy integral.

1 Introduction

Current decision making systems are mainly designed in a sequential approach. Usually, two stages are involved [26]. The first one, which is typically highly problem specific, is concerned with the design of a population of base classifiers. Assuming that diversity is a useful property in the design of decision making systems [14, 15, 25], a highly diverse population of base classifiers is constructed in this stage. In addition, the second stage is concerned with the design of combination rules over individual classification results, which is generally application independent. At this stage, combining rules enforcing complementarity are implemented. It should be noted, however, that a rich population of classifiers in terms of diversity might not be entirely errors complementary, i.e., there is no explicit measure of diversity involved in the sequential design [14].

Intuitively, when the number of diverse classifiers is increased, the quantity of joint misclassifications should diminish. However, many combinational methods might yield to inconsistent results or increase greatly their computational complexity. Inconsistencies might be due to the lack of fulfilment of combination rules hypotheses (i.e., independence assumption in simple rules). In trained combination rules [7] complexity problems arise due to the estimation of a huge number of parameters, and estimation simplifications might also imply inconsistent results. Therefore, with a growing population of classifiers an alternative design strategy should be considered.

In this paper, a proposal for the design of highly scalable multiclassifier systems is presented. The proposal is based on the selection of a reduced subset of potentially cooperating classifiers. The target subset is constructed in a greedy fashion using a heuristic method, which takes into account both the performance and errors distribution of selected classifiers according to the intended combination rule. At a first glance, our proposal might be considered as an instance of previous solutions of large multiclassifier systems, like *choose the best* [18] or *test and select* [22] methods. We recall that the *choose the best* method considers only the individual accuracy of classifiers under the assumption of a uniform distributed diversity. In our work, the uniform distributed assumption is not mandatory. In addition, *test and select* methods are basically brute force approaches because a target subset is searched in the set of all available subsets of classifiers. In the present work, a target subset is selected using a single scanning over the set of error patterns of classifiers. At this scanning process, classifiers selection is performed taking into account the accuracy and complementarities of misclassifications per class. Computational complexity of this selection process *scales linearly* w.r.t. the number of available classifiers.

In this paper, we focus on Fuzzy Integral (FI) combination rules. Due to the selection process, FI parameters can be easily computed. As a fact, the estimation of FI parameters might be computationally unaffordable if the whole population of classifiers were used. Furthermore, even in the case those FI parameters were simplified by the use of decomposable measures, like λ-measures, precision problems might arise instead. For the sake of simplicity, we constrain ourselves to minimalist FI rules, i.e., target subsets of three classifiers. The proposed subset size is tightly linked to our recent work about the design of scalable decision making systems based on the selection of multiple, not necessarily disjoint but still small, subsets of classifiers [3]. We expect the present work enlightens the computational cost vs. robustness of joint decision. This paper is organized as follows. In section 2, we present the problem. In section 3, we propose the design cycle of scalable multiclassifiers systems. We also give a briefly review of FI combination rules before going into details about the core selection process. In Section 4, we present experimental results. In Section 5, we present the conclusions and further work.

2 Problem Statement

Multiclassifier systems are effective if individual misclassifications can be overcame by a cooperative work among classifiers. The joint decision is achieved by combining individual results. The use of trained combination rules, like fuzzy integrals, allows us to overcome hard classification scenarios, where few classifiers in the given population are strongly fitted to the target classification problem [21] and consequently, simple combination rules, like majority voting, might fail. In addition, combinations of a growing population of classifiers might yield to inconsistent results or increase their computational complexity. Therefore, a proper selection of a small subset of complete classifiers, complete enough to cover the problem, joint with the use of a powerful combination rule may provide a framework for solving a general scalability problem.

The aim of this work is to propose a systematic process to make possible scalable joint decision makings when a general combination rule is used. To do that, we propose a simple selection of the most error complementary and task redundant subset of classifiers (figure 1) regarding FI combination rules.

Fuzzy integrals combine individual decisions taking into account their reliability. This approach

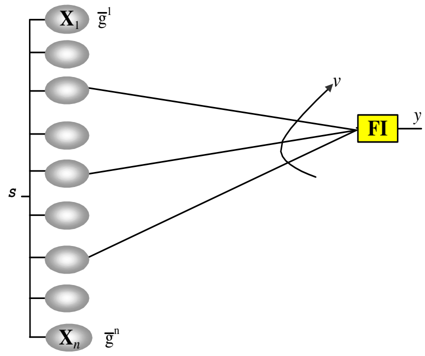

Figure 1: Scalable decision making system

has sense if there is redundancy in the analysis and if classifiers are different. Previous studies on joint decision making agree with the need of diversity in base classifiers [6, 14, 21].

Let us assume that a population of *n* classifiers, $\{X_1, ..., X_n\}$, non specialized on the set of possible alternatives, $W = \{w_1, ..., w_c\}$, is given. The selection process computes the individual accuracy and error distribution of classifier.

We assume that available data comes into train and test data sets. In addition, we presume that enough training data is available, so that, two training data sets can be generated: one for training the base classifiers and the other for training the combination rule, e.g., for fuzzy measures estimations. Hereafter, we will refer to training data sets as *classifiers training data set* or T_1, and *combination rule training data set* or T_2; and the *testing data set* is called T_3.

The distributions of errors related to classifiers are estimated during the training of the combination rule. To do that, each classifier will analyse the samples of T_2. When individual classifiers disagree their decisions are piled up in one matrix called *matrix of error pattern* or E. Assuming that classifiers disagree in *k* samples, E will be a $k \times n$ binary matrix with each column composed of a classifier performance, where 0 implies error and 1 implies non error in the current sample.

The selection of a suitable subset, hereafter named O and *v* its cardinality, takes into account the constraint related to the estimation of FI parameters, and the individual and joint behaviour of classifiers related to misclassifications.

3 Adaptive Design Cycle of Decision Making Systems

Concerning the adaptability issues in the design of scalable decision making systems, we will consider the selection of a subset of classifiers based on both their accuracy and their distribution of errors. The overall proposed design involves the following stages:

- *Base Classifiers Generation:* Aiming the generation of a highly diverse population of classifiers, different artificial intelligent techniques are implemented and different parameters in the generation of knowledge bases are used.

- *Combination Rule Formulation:* Aiming the induction of highly descriptive combination rules, FI rules involving a number *v* $<<$ *n* are trained. For the sake of simplicity *v*=3 is adopted.

- *Joint and Individual Behaviour Characterization:* Two kind of measures are considered for the performance description: errors density and distribution. In addition, the behaviour characterization process also takes into account the selected combination rule constraints, e.g., scalable fuzzy densities and measures computation.

- *Target Subset Selection:* This stage simultaneously takes into account the two previous stages.

The above mentioned design steps are summarized in the diagram 2.

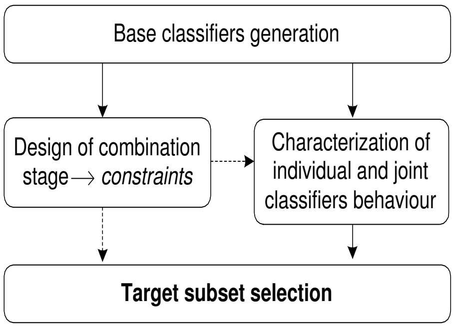

Figure 2: Multiclassifiers design cycle based on misclassifications complementariness

We will concentrate in the last three points, the first one is a huge line of research [2], which involves methods of manipulation of training parameters such as *training data sampling* [1, 8, 12], *data preprocessing* [20], and *modification of feature space* [10, 17].

In what follows, we briefly revise FI combination rules regarding their use in the error adaptive design of decision making systems. After that, we will describe the subset selection procedure.

3.1 Decision Integration with FI

Fuzzy integrals have been shown to be a useful method for combining results of multiple sources of information. Its definition with respect to a *fuzzy measure* [23] or *capacity* [5] provides a good framework to represent the imprecise knowledge associated with classifiers. In the literature, practical implementations [4, 13] only combine 2 or 3 classifiers due to the constraint of the parameter estimation with a large number of classifiers.

We will consider the discrete Choquet and Sugeno FIs. See [9, 16] for theoretical details.

Assuming a multiclassifier system composed of *n* classifiers, and C possible alternatives for input samples (*s*), that belong to a *p*-dimension feature space, $s \in \Re^p$, we call $f(X_i(s))$ or $f_i : \Re^p \to [0,1]^C$, the classification function that supplies the individual decision through a vector of *C* components. Each component of this vector represents the degree of support given by the classifier X_i to the hypothesis that *s* comes from w_j, with $j = 1, ..., C$. The joint decision is obtained by aggregating all partial classifiers evidences weighed by ability degrees (*g*).

Let *g* be a fuzzy measure on *X*, whose elements are denoted $X_1, ..., X_n$. The Sugeno integral [23] of a function f:X $\to [0,1]$ w.r.t. *g* is defined by

$$S_g(f) := \max_{i=1}^{n}\{min(f(X_{(i)}), g(A_{(i)}))\} \tag{3.1.0}$$

The (discrete) Choquet integral [5] of a function $f : X \to [0,1]$ w.r.t. *g* is defined by

$$C_g(f) := \sum_{i=1}^{n}(f(X_{(i)}) - f(X_{(i-1)}))g(A_{(i)}) \tag{3.1.0}$$

Where $\cdot_{(i)}$ indicates the permutation of indices 0$\leq$ $f(X_{(1)}) \leq \ldots \leq f(X_{(n)})$ $\leq$1, $f(X_{(0)})$:=0, and $A_{(i)}$:=$\{X_{(i)}, \ldots, X_{(n)}\}$.

A fuzzy measure or capacity *g* on *X* is a function $g : 2^X \rightarrow$ [0, 1] if

1. g(∅) = 0,
2. g(X)= 1,
3. A $\subset$ B $\subset$ X $\Rightarrow$ g(A) $\leq$ g(B).

In our case, $g(A_{(i)})$ quantifies the ability of $A_{(i)}$ to classify the input on the *W* space. In particular, when *g* is related to a single element, X_i with $i \in \{1, ..., n\}$, is called *fuzzy density* of the *i*th source or g^i.

The estimation of fuzzy measures can be done in different ways, e.g., using probabilities of misclassification [9, 19], through iterative algorithms to diminish the quadratic error [9, 11] or using genetic algorithms [24]. In general, fuzzy measures cannot be evaluated from fuzzy densities.

Though FI operators provide a useful framework for the design of fuzzy adaptive decision making systems, they require the estimation of $2^n - 1$ parameters for *n* classifiers, which even for small *n* values might be computationally unaffordable.

In an attempt to overcome the complexity estimation of fuzzy measures, Sugeno introduced λ-fuzzy measures, which are decomposable measures related to fuzzy densities, that fulfil the further property:

$$g_\lambda^{A,B} = g_\lambda(A \bigcup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$$

for every disjoint subsets A, B of *X* and for some $\lambda \in$ (-1,∞). λ can be uniquely determined for a finite set by means of the second property of fuzzy measure ($g(X) = 1$), which leads to solving a polynomial of ($n - 1$)th degree. Consequently, dealing with large groups of classifiers would imply a difficulty in the polynomial root determination, but we use small coalitions to attend the precision constraint: λ value propagates the estimation errors of all fuzzy densities.

The use of λ-measure simplification reduces to $n - 1$ the number of parameters to be estimated, but also limits the flexibility in fuzzy measures relationship.

3.2 Systematic Selection of Complementary Classifiers

In this section, we describe the core selection process. Main inputs to the process are fuzzy densities and joint distribution of errors over the training data of combination rules.

In the first step, we seek the classifier exhibiting the lowest amount of misclassifications ($seek\_X_1^*$ method of algorithm 1). In this way, an initial O subset is defined and its greedy augmentation starts. We compute the performance value, in each class, related to the current subset of classifiers ($O = \{X_1^*\}$).

Regarding this purpose, the rows of the E matrix are separated in *C* sub-matrices depending on the class associated with samples ($split\_w$). In addition, a new vector of pattern of error associated with O is generated per class ($E_o[w_j]$) and its density of 1s is evaluated. At first, E_o is initialised with the error pattern related to X_1^*.

3.2.1 Example:

In order to illustrate the selection methodology, we shall see a toy example. Let us consider the following matrix of the error pattern of 12 samples, table 3.2.1, associated with 4 classifiers and 3 classes. The first five samples are related to w_1, the next three are related to w_2, and the last fourth are from w_3.

$$
E= \begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 \\
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0
\end{array}
$$

Table 1: Matrix of the error pattern

The first selected classifier is X_4 (X_1^*) due to its coverage (density of 1s) on the 12 samples. Next, E is divided into $E[w_1]$, $E[w_2]$, and $E[w_3]$ associated with w_1, w_2 and w_3 respectively.

$$
E[w_1]= \begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 \\
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 \\
0 & 1 & 1 & 0 \\
1 & 1 & 1 & 0
\end{array}
\quad
E[w_2]= \begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 \\
1 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 \\
0 & 0 & 1 & 0
\end{array}
\quad
E[w_3]= \begin{array}{cccc}
X_1 & X_2 & X_3 & X_4 \\
1 & 0 & 0 & 1 \\
1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 0
\end{array}
$$

So, the coverage of classifiers of $O = \{X_4\}$ (first time only X_1^*) on the subset of samples computed in each class $(comput\_perf)$ is: $WPerf[1] = \frac{2}{5}$, $WPerf[2] = \frac{2}{3}$, $WPerf[3] = \frac{2}{4}$. Therefore, the next classifier to be included is in order to improve w_1 (the lowest performance class).

The class yielding the lowest performance, $w_{(-)}$, is used in order to guide the addition of the next classifier (X^*). As a fact, we seek the best classifier in terms of complementarity and quantity of errors w.r.t. to the poorest class using its error pattern, $E[w_{(-)}]$. Regarding this purpose, a new matrix of errors ($E_{||}$) is computed by a simple *or* operation (denoted by $||$) between the current vector of the subset error pattern ($E_o[w_{(-)}]$) and each column of $E[w_{(-)}]$. Non zero entries in this new matrix points out that the sample might be correctly classified when adding the column classifier to the current subset. Hence, we should look for columns (the error pattern of each classifier) with highest density of ones (max_1X). Tie breaking is handled by choosing the classifier with less misclassifications w.r.t. the current subset ($max\_XRedundance$).

3.2.2 Example:

Following the previous example, the computation of the error pattern of the current subset w.r.t. the worst class ($w_{(-)} = w_1$) and the evaluation of complementarity of remaining classifiers is shown bellow.

$$E[w_1]= \begin{array}{ccc} X_1 & X_2 & X_3 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array} \quad \| \quad \begin{array}{c} E_o[w_1] \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{array} \quad \longrightarrow E_{\|}[w_1] \begin{array}{ccc} X_1 & X_2 & X_3 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{array}$$

As a result, $X^* = X_3$ is included to O. Both X_2 or X_3, working jointly with the classifiers of O, could achieve an equal coverage on samples associated with w_1, but X_3 has less individual misclassifications (or higher coverage on w_1).

When a new classifier is included to O, the $E_o[w_j]$ vector is upgraded computing or operation between the current $E_o[w_j]$ and the *i*-th column of $E[w_j]$ matrix. The selection process continues until the target size of the preselected subset is achieved ($Q = v$).

3.2.3 Example:

Hence, the new coverage with $O = \{X_4, X_3\}$ is:

$$\begin{array}{c} E_o[w_1] \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{array} \quad WPerf[1] = \frac{4}{5} \qquad \begin{array}{c} E_o[w_2] \\ 1 \\ 1 \\ 1 \end{array} \quad WPerf[2] = \frac{3}{3} \qquad \begin{array}{c} E_o[w_3] \\ 1 \\ 0 \\ 1 \\ 0 \end{array} \quad WPerf[3] = \frac{2}{4}$$

In the same way, X_2 is the next classifier to be included in order to improve the classification of w_3.

It should be noted that the or operation among the error pattern of classifiers allow us to estimate the classifiers ability (computing the coverage) for analysing the set of validating samples.

Next, the whole selection process is presented in an algorithmic form (algorithm 1).

4 Illustrative Example

We will illustrate an application of scalable multiclassifiers architecture with *wine*[1] data sets.

A. Data Processing

The *wine* data set contains 178 samples, 13 continuous inputs variables, and 3 output classes. The protocol applied to data set does samplings to generate 10 random sets of 75% samples to train (half for classifiers and half to estimate FI parameters), and 25% for system testing.

B. FI parameters

Columns: n = 10, given classifiers with 20% of mean errors. Their diversity is achieved implementing different inference methods (5 classifiers are neural networks[2], and 5 classifiers implement fuzzy inference[3]).

[1] http://www.ics.uci.edu/˜mlearn/MLRepository.html

[2] http://fuzzy.cs.uni-magdeburg.de/˜borgelt/software.html

[3] http://www.inra.fr/bia/M/fispro/

Algorithm 1 Error adaptive classifier selection

Input: The error pattern of *n* classifiers on $T_2 = \{z_1, ..., z_k\}$ samples, $E_{k\times n}$
    v: Upper bound of classifiers subset size
Output: Subset of *v* classifiers, O, with the highest misclassification complementarity
$O \leftarrow seek\_X_1^*(E)$
$Q \leftarrow 1$
for $1 \leq j \leq C$ **do**
    $E[w_j] \leftarrow split\_w(E)$
end for
for $1 \leq j \leq C$ **do**
    $E_o[w_j] \leftarrow E[w_j][X_1^*]$
end for
while $Q \leq v$ **do**
    for $1 \leq j \leq C$ **do**
        $WPerf[j] \leftarrow comput\_perf(E_o[w_j])$
    end for
    $w_{(-)} \leftarrow min(WPerf[j])$
    for $1 \leq i \leq n$ **do**
        $E_{\|}[i][w_{(-)}] \leftarrow E_o[w_{(-)}] \| E[i][w_{(-)}]$
        $X^* \leftarrow max\_1X(E_{\|}[i][w_j])$
        if *tie* = *true* **then**
            $X^* \leftarrow max\_XRedundance(E_{\|}[w_j])$
    end for
    $O \leftarrow include(X^*)$
    $Q \leftarrow Q + 1$
end while

The dimension of O subset: v=3.
The results are summarized in the table 2.

n=10 and v=3	**Sugeno**	**Choquet**
Median Error %	0	0
Mean Error %	2.7	2.5
n=10 and v=10 Mean Error	$\sim$13	$\sim$12

Table 2: Scalable multiclassification performance

We noted that combining the whole set of classifiers the mean error was more than 12%. With the selection proposed the mean error using Sugeno FI was 2.7% and 2.5% with Choquet FI, better than the best classifier and better than the combination of the complete ensemble.

5 Conclusions

We have presented a method to design compact multiclassifier systems when a large base of classifiers is given. As we could note, the improvement of overall accuracy can be done when more information is taken from classifiers. The joint decision making using fuzzy integral combination rules can be done under an appropriate selection of classifiers. In fact, an adaptive classifiers selection, to their error pattern, may provide a framework for solving a general scalability problem.

We have utilized a fixed size of classifiers subset. In further works, this value (v) could be optimised by means of cross-validation to each data set. Intuitively, we expect that highly compact and monolithic decision making systems, like the ones we are proposing, might exhibit less adaptation capabilities than those based on a completely distributed design. However, computational costs might be also higher.

Experimental results show that a given group of diverse classifiers can improve their individual performance and the global performance with a cautious splitting and selection.

Acknowledgements

This research was supported by Spain (TIC2001-2741) and Argentine (PICT /2003, project 15132).

References

[1] L. Breiman, *Bagging predictors*, Machine Learning **24** (1996), no. 2, 123–140.

[2] 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.

[3] P. Bulacio, L. Magdalena, and E. Tapia, *Scalable fuzzy multiclassifier system*, Fourth Conference of the European Society for Fuzzy Logic and Technology and Eleventh Rencontres Francophones sur la Logique Floue et ses Applications, In Press (2005), 1–6.

[4] S. Cho and J. Kim., *Multiple network fusion using fuzzy logic*, Neural Networks, IEEE Transactions on **6(2)** (1995), 497–501.

[5] G. Choquet, *Theory of capacities*, Annales de l'Institut Fourier **5** (1953), 131–295.

[6] Padraig Cunningham and John Carney, *Diversity versus quality in classification ensembles based on feature selection*, Machine Learning: ECML 2000, 11th European Conference on Machine Learning, Barcelona, Catalonia, Spain, May 31 - June 2, 2000, Proceedings, vol. 1810, Springer, Berlin, 2000, pp. 109–116.

[7] R.P.W. Duin, *The combining classifier: to train or not to train?*, Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 2, 2002, pp. 765–770.

[8] Yoav Freund and R.E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, European Conference on Computational Learning Theory, 1995, pp. 23–37.

[9] M. Grabisch, *Fuzzy measures and integrals for decision making and pattern recognition*, TATRA MOUNTAINS Mathematical Publications **13** (1997), 7–34.

[10] Tin Kam Ho, *The random subspace method for constructing decision forests*, IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998), no. 8, 832–844.

[11] J.M. Keller and J. Osborn, *Training the fuzzy integral*, Int. J. Approx. Reasoning **15** (1996), no. 1, 1–24.

[12] Anders Krogh and Jesper Vedelsby, *Neural network ensembles, cross validation, and active learning*, Advances in Neural Information Processing Systems (G. Tesauro, D. Touretzky, and T. Leen, eds.), vol. 7, The MIT Press, 1995, pp. 231–238.

[13] L.I. Kuncheva, *'fuzzy' vs 'non-fuzzy' in combining classifiers: an experimental study*, Proc LFA'01, Belgium, 2001, pp. 11–22.

[14] L.I. Kuncheva and C.J. Whitaker, *Ten measures of diversity in classifier ensembles: limits for two classifiers*, In Proc. IEE Workshop on Intelligent Sensor Processing, Birmingham, IEEE., 2001, pp. 10/1–10/6.

[15] F. Masulli and G. Valentini, *Quantitative evaluation of dependence among outputs in ecoc classifiers using mutual information based measures*, Proceedings of the IJCNN'01, K. Marko and P. Webos (eds.), IEEE, Piscataway, NJ, USA **2** (2001).

[16] T. Murofushi and M. Sugeno, *Fuzzy measures and integrals: Theory and applications*, ch. Fuzzy Measures and Fuzzy Integrals, pp. 3–41, Physica-Verlag, 2000.

[17] N. C. Oza and K. Tumer, *Input decimation ensembles: Decorrelation through dimensionality reduction*, In Proc. of the 2 nd International Workshop on Multiple Classifier Systems, Cambridge, UK (2001), 238–247.

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

[19] Tuan D. Pham and Hong Yan, *Information fusion by fuzzy integral*, Intelligent Information Systems (Narasimhan and Jain, eds.), Australian and New Zealand Conference on, Narasimhan and Jain, 1996, pp. 191–194.

[20] Y. Raviv and N. Intrator, *Boostrapping with noise: An effective regularization technique*, Connection Science, Special issue on Combining Estimators **8(3-4)** (1996), 355–372.

[21] A. Sharkey and N. Sharkey, *Combining diverse neural net*, The Knowledge Engineering Review **Vol 12, Issue 3** (1997), 231–247.

[22] Amanda J. C. Sharkey, Noel E. Sharkey, Uwe Gerecke, and G. O. Chandroth, *The "test and select" approach to ensemble combination*, Lecture Notes in Computer Science **1857** (2000), 30–44.

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

[24] D. Wang, X. Wang, and J.M. Keller, *Determining fuzzy integral densities using a genetic algorithm for pattern recognition*, Fuzzy Information Processing Society (1997), 263–267.

[25] David H. Wolpert, *Stacked generalization*, Neural Networks **5** (1992), no. 2, 241–259.

[26] 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.