International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 17, No. 8 (2003) 1303–1324

COMBINATION, COOPERATION AND SELECTION OF CLASSIFIERS: A STATE OF THE ART

V. GUNES*, M. MÉNARD and P. LOONIS

Laboratoire d'Informatique, d'Image et d'Interaction, Université de La Rochelle, Avenue Michel Crépeau, F - 17 042 La Rochelle Cédex 1, France
**v.gunes@laposte.net*

S. PETIT-RENAUD

Laboratoire d'Informatique de l'Université du Maine, Université du Maine, Avenue René Laënnec, F-72 085 Le Mans Cédex 9, France

When several classifiers are brought to contribute to the same task of recognition, various strategies of decisions, implying these classifiers in different ways, are possible. A first strategy consists in deciding using different opinions: it corresponds to the combination of classifiers. A second strategy consists in using one or more opinions for better guiding other classifiers in their training stages, and/or to improve the decision-making of other classifiers in the classification stage: it corresponds to the cooperation of classifiers. The third and last strategy consists in giving more importance to one or more classifiers according to various criteria or situations: it corresponds to the selection of classifiers. The temporal aspect of Pattern Recognition (PR), i.e. the possible evolution of the classes to be recognized, can be treated by the strategy of selection.

Keywords: Multiple classifier systems; combination; cooperation; adaptive selection; dynamic selection; ensemble of classifiers.

1. Introduction

The first allusions to *systems of classifiers* or *multiple classifier systems* (also called *ensemble* of classifiers, especially by the neural networks community), quoted by Ho,[30] go back to Nilsson[46] and Haralick[29] for the error analysis relating to a set of Bayesian classifiers. However, the systems of classifiers were put into practice concretely only from the early 1990s, in particular within the framework of the recognition of words[30] and handwritten words.[59,60,66] It is significant to specify that the expression *classifier system* is used by the community of genetic algorithm (see in particular, Ref. 24) in the sense of an expert system whose search engine is founded on a genetic algorithm to generate some behavioral rules, in interaction with its environment.

In *Pattern Recognition*, or PR, the objective is to obtain a *system of classifiers*,[a] i.e. an association of classifiers, highly powerful (in the sense of a high recognition rate) and based on the decisions of a set of classifiers (Ref. 66; more recently, Refs. 2 and 21). Another motivation for this type of system lies in the complexity of the forms of the classes involved; when the classes are multi-mode or when they overlap, it becomes difficult to model them, because their representation (for example, by centers, probability distributions or fuzzy sets) becomes complicated. Methods in which mixture models are used (see in particular the work of McLachlan and Basford[44]) are suited to such classes, but whenever the size of training set is limited, this modeling leads to an over-fitting with the available data, which do not necessarily represent accurately the real classes. By associating various classifiers (different algorithms and/or different constructions), one hopes to obtain better performances. Lastly, within the framework of the recognition of evolutionary objects, it is useful to determine the strategies of associations which are best adapted to the temporal aspect of dynamic classes.

Let $F = \{f_1, f_2, \ldots, f_j, \ldots, f_m\}$ be the feature space (as different from the feature representation space, often contracted by "feature space" and defined in the case of statistical pattern recognition by $\mathbb{R}^n$), where the f_i are different features and m is the number of available features. And let $\Omega = \{C_1, C_2, \ldots, C_i, \ldots, C_l\}$ be the decision space (also called frame of discernment or space of discernment), where the C_i are the different classes or assumptions and l is the number of classes or assumptions. According to the situation, the systems of classifiers can have the following advantages:

- In the same way as an additional feature f_i makes it possible to a classifier to improve its decision (provided that this feature is complementary to the other features), an additional opinion (defined on the space of decision Ω) coming from another classifier makes it possible for a system of classifiers to better decide (provided that this opinion is complementary to the other opinions).
- The association of classifiers allows a modular and dispatched/shared approach of PR.
- When, for a given PR problem, some features of F are real and the others are discrete or symbolic, it is often unavoidable to use two different kinds of classifiers. Thanks to a common space of discernment Ω, the opinions of the two classifiers can be easily combined.
- In certain cases, algorithmic complexity (in training stage and/or in classification stage) can be reduced if the data processing is distributed on several levels.
- Lastly, for the evolutionary classes, it can be necessary to specialize several classifiers on different temporal intervals, and to use these classifiers according to time and/or according to the pattern to be classified.

[a]Other used terms are: multi-experts combination, committee of classifiers, fusion of classifiers and others.

Currently, the most commonly used operation to associate the various classifiers of a system is the "combination" of outputs of classifiers.[21] Recently, some researchers studied another kind of operation, called "adaptive selection of classifiers"[19] or "dynamic selection of classifiers".[31,65] Other authors proposed to use the outputs of individual classifiers as features for the input of another classifier (of "decision"), carrying out the training on the decisions of these classifiers.[1,62] However, it should be noted that the very large majority of current work relate to the combination only. This state of the art includes also related work within the framework of other classifiers associations strategies.

Finally, very recently, a method of automatic design of "multiple classifier systems" was proposed.[22] The proposal is to choose a set of classifiers of different methods and/or parameters among a greater set of classifiers and to select (with a validation set, different from the test set) the subset of the classifiers whose performances are high. Then, the candidate classifiers who do not improve significantly recognition rates (whose errors are correlated with the errors of one or more other classifiers) are eliminated. Lastly, in classification stage, the outputs of the selected classifiers are combined with the majority vote rule.

This article is organized as follows. We begin with some explanations about the taxonomy of the systems of classifiers (Sec. 2). Then, the operations such as combination, cooperation and selection of classifiers are respectively detailed in Secs. 3–5. This is completed by a section about hybrid systems and the topology of such systems. Finally, Sec. 7 includes a conclusion and some prospects.

2. Taxonomy of Systems of Classifiers

Recently, Giacinto *et al.*[21] insisted on the possibility of utilizing various types of operations in the systems of classifiers. In addition to the *combination* of the outputs of classifiers, researchers recently introduced other types of operations. Gosselin[25] used the terms of *cooperation* and *combination* in the same way. Franke and Mandler[18] also used the terms in the same way, while proposing two different types of constructions; "combined outputs of cooperating classifiers being used as features for a classifier of higher level" or "combination of the outputs of classifiers being interpreted like votes of automatic specialists with each one having a certain point of view on the input pattern". We call these two types of associations *cooperation* of classifiers and *combination* or *fusion* of classifiers, respectively. We propose to distinguish the systems of classifiers in three categories, according to the type of operation between the classifiers:

- combination of classifiers,
- cooperation of classifiers,
- selection of classifiers.

If a system operates with several types of associations, the system is known as hybrid or mixed. The first two types of associations also seem to be major concerns in the neural networks community. According to Sharkey,[54] this community uses mainly two terms: *combination of an ensemble* (of classifiers) and *combination of modules* (of classification). Thus, the term *combination* is used here in the way of an association. According to this author, "*in an* ensemble, *component nets are redundant in that they each provide a solution to the same task ... By contrast, under a* modular *approach, a task is decomposed into a number of subtasks, and the complete task solution requires the contribution of all the several modules* (*even though individual inputs may be dealt with by only one of the modules*)". It is clear that the *ensemble* approach is an expression of the combination of classifiers while the *modular* approach is an expression of the cooperation of classifiers. The contents of the last brackets implicitly make an allusion to the selection (adaptive, see the definition of this term in Sec. 5) of classifiers. The author also specifies that the concept of selection of the members of an *ensemble* was approached in a certain number of publications, and proposes to make use of it explicitly. The proposal is to determine the performances of different *sets* of classifiers (among a fixed higher number of classifiers), by testing them (method "test and select") on a validation set, different from the training set and the test set. But the disadvantages of this approach are that, on the one hand, the step of validation/constitution of an ensemble requires a high computing time and on the other hand, the obtained *ensemble* is fixed and cannot change according to a pattern to be classified.

Some authors, while also being interested in the complexity of the algorithms, proposed to carry out conditional combinations. Thus, Gosselin[25] proposed to classify the classifiers according to their performances and to treat an unknown pattern by a first classifier. Then, the proposal is to accept its decision, if the pattern is not rejected. In the opposite case, the decision will be made using the combination of the outputs of the first and the second classifiers. The same reasoning can apply, until the pattern is classified or the outputs of all the classifiers are combined. This conditional combination makes it possible to reduce the execution times effectively. The disadvantage is the need for fixing multiple rejection thresholds associated with the various decisions.

Table 1 at the end of this article, presents the related work in a synthetic way. This work is presented according to the above taxonomy as well as according to the types of specializations of the classifiers. The most common related works deal with the combination of classifiers using different features (in the first column, first line), whereas the most recent related works involve rather the adaptive and dynamic selection.

3. Combination of Classifiers

The combination of classifiers[b] takes place within the more general framework of multi-source combination. The encountered problems are those which can also be

[b]In the literature, terms such as fusion of classifiers and data fusion are also used.

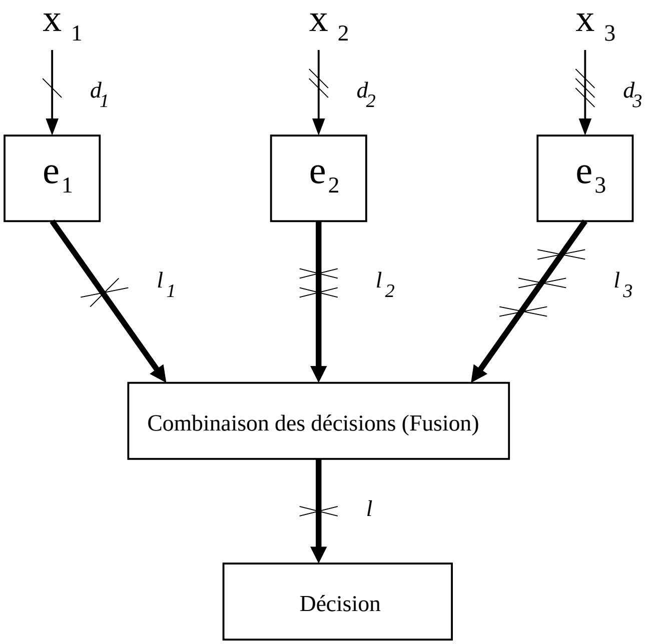


Fig. 1. System of classifiers carrying out the combination of classifiers. In most related work, $l_1 = l_2 = l_3 = \{C_1, C_2, \ldots, C_l\}$. The available features are $F = \{\{x_1\}, \{x_2\}, \{x_3\}\}$ and d_1, d_2 and d_3 are the dimensions of x_1, x_2 and x_3, respectively. The features of a classifier e_i can be different from the features of a classifier e_k but the tasks of the two classifiers can still be the same (e.g. these classifiers can use different features to recognize the same faces). This property is useful in the case of a dispatched/shared system (all the features may not be available at the same location).

found in the following fields:

- combination of sensors,
- combination of expert opinions, and
- combination of data bases (of their answers).

Data fusion techniques are employed in many applications: image processing (Refs. 5, 71, and many others), handwritten digit and word recognition,[34,59] face recognition,[1] speech recognition[68] and others. The information provided can come from various nature and type of sources (sensors, human experts, data bases, classifiers). Indeed, they can constitute a signal, a sound or an image, and be of numerical and/or symbolic type. Unfortunately, the sources are often sullied with inaccuracy and/or uncertainty and also with incompleteness and/or ambiguity. Figure 1 illustrates the possible interactions between the various modules of a system of classifiers putting combination into practice. In this configuration, the chosen feature vectors for the various classifiers can be different, but they should concern the same pattern to be recognized. The figure shows a problem with l classes and d available features (thus, we have $d_1 \leq d$, $d_2 \leq d$ and $d_3 \leq d$). Generally, the classifiers e_1, e_2 and e_3 have a common decision space, i.e. the l classes of the various classifiers have the same significances.

The theoretical and experimental results which are seen in the literature show clearly that the combination of classifiers can be effective only if the individual classifiers are "accurate" and "diverse",[53,61] i.e. if they have low error rates and if they induce different errors (this is also called complementarity). Two classifiers induce different errors when the patterns badly classified by the first are different from the patterns badly classified by the second. This latter condition means that the decisions or the outputs of the classifiers are decorrelated (each classifier specializes on a certain number of classes). It can be satisfied by using different feature spaces, different training sets or different (intrinsically) classifiers (various parameter settings or different types of classifiers). A mixed approach, combining these various means, is also possible. However, basically, three types of combinations are possible. They come from the specialization of the classifiers according to:

- different features,
- different methods, or
- different learning sets.

By using various features, one specializes the classifiers in the feature space Ω. By using various methods (and/or parameter settings), one specializes the classifiers in the decision space (frame of discernment). Indeed, to find the advantage of a classification method over another or a parameter setting against another, means to consider a decision against another or certain decisions against others. Lastly, by using various training sets, one specializes the classifiers in the space of representation. The majority of work already completed relate to specializations of the first type. However, the two other types should be considered as complementary.

The techniques of *Bagging*[6,47] and *Boosting*[52] implicitly specialize the various produced classifiers in the space of representation. The technique of *Bagging* redefines the training set for each classifier, by independent and random selection (uniform probability distribution) of the patterns of an available training set, whereas that of *Boosting*, assigns to the patterns of the training set various weights for the various classifiers. In the latter technique, weights are also assigned to the classifiers, at the stage of combination of the decisions.

Kuncheva *et al.*[39] showed that improvement is also possible if the classifiers are dependent (i.e. whose decisions are correlated), but provided that the errors made by the classifiers concern very different patterns (the authors call this the "negative dependence").

In this type of system of classifiers, each classifier is applied in parallel and their outputs are combined to reach a consensus; these approaches use mainly the framework of the theory of vote,[41,66] the unanimous consensus,[31,66] methods of vote using heuristic rules of decision,[38] the Bayesian theory,[66] theories of uncertainty,[51,66] the theory of possibilities[17] and the theory of fuzzy sets.[69] The choice of the combination type depends, in particular, on the classifiers outputs type.

According to their type of outputs, the classifiers can be divided into three sets. Type 1 is defined by crisp decisions, type 2 is defined by ranked (and crisp) decisions and type 3 is defined by a similarity measure to each class (not crisp). Owing to the fact that the outputs of the type 3 are richer in information, our interests lie more in fusion for this type of outputs.

3.1. *Common combinations*

To generalize, let us define $v_{is}(x_k)$ as a numerical value calculated by the classifier e_s for the class C_i during the classification of a pattern x_k and w_i as a numerical value calculated by the system of classifiers for the class C_i. These values can be, for example, probabilities or degrees of memberships. Let l be the number of possible assumptions (classes) and S the number of classifiers whose outputs enter into combination. Let $l+1$ be the class of rejection and T (with $0 < T \leq 1$) be a chosen reject threshold. Usually the most usually employed rules are:

- The Maximum rule:

$$\forall\, i \in [1, l], \quad w_i(x_k) = \max_{s=1}^{S} v_{is}(x_k)\,. \tag{1}$$

- The Minimum rule:

$$\forall\, i \in [1, l], \quad w_i(x_k) = \min_{s=1}^{S} v_{is}(x_k)\,. \tag{2}$$

- The Sum rule:

$$\forall\, i \in [1, l], \quad w_i(x_k) = \sum_{s=1}^{S} v_{is}(x_k)\,. \tag{3}$$

- The Mean rule:

$$\forall\, i \in [1, l], \quad w_i(x_k) = \frac{1}{K}\sum_{s=1}^{S} v_{is}(x_k)\,. \tag{4}$$

- The Median rule :

$$\forall\, i \in [1, l], \quad w_i(x_k) = \mathrm{median}_{s=1}^{S} v_{is}(x_k)\,. \tag{5}$$

Usually, the decision rule is defined by a function $SC(x_k)$ such as:

$$SC(x_k) = \begin{cases} C_j, & \text{if } w_j(x_k) = \max_{i=1}^{l} w_i(x_k) \\ w & \text{and} \quad w_j(x_k) \geq T \\ C_{l+1}, & \text{else}\,. \end{cases} \tag{6}$$

In the following sections, we will present the majority vote rule followed by an introduction to the *product* rule within the framework of the Bayesian theory and within the framework of the Belief theory.

3.2. *Combination by vote*

Let us suppose that each classifier $e_i \in E_i$ carries out a "crisp" classification, assigning each vector of features (or pattern) to one of the classes. A simple method of combination of the outputs of these classifiers is to interpret each output as "a vote" for one of the classes. The class which obtains a number of votes higher than a preset threshold is retained as a final decision. If this threshold is selected as being half of the number of voting classifiers, the rule is called combination with majority vote.

The advantage of this type of combination is that it can be used for any type of classifier, whatever is the type of outputs of these classifiers (i.e. "crisp" classification or not).

3.3. *Combination and decision by the probability theory*

Let $\{y_1, y_2, \ldots, y_s, \ldots, y_S\}$ be the decision vectors provided by the classifiers $e_1, e_2, \ldots, e_s, \ldots, e_S$ respectively for the pattern x_k. In the probability model, the value v_{is} calculated by e_s is represented by a probability, denoted as conditional probability according to class C_i:

$$v_{is}(x_k) = P(x_k/C_i), \quad i \in [1, l]\,. \tag{7}$$

The information combination concerning the class of x_k is represented by an *a posteriori* probability obtained with the rule of Bayes:

$$P(C_i/y_1, y_2, \ldots, y_S) = \frac{P(y_1, y_2, \ldots, y_S/C_i)P(C_i)}{P(y_1, y_2, \ldots, y_S)}\,. \tag{8}$$

where $P(C_i)$ is the *a priori* probability of the class C_i. Under independence assumptions (i.e. independence of classifiers $e_1, e_2, \ldots, e_S$ and conditional independence of $e_1, e_2, \ldots, e_S$ to C_i), Eq. (8) can be simplified as follows:

$$P(C_i/y_1, y_2, \ldots, y_S) = \frac{P(C_i)\prod_{s=1}^{S} p(y_s/C_i)}{\sum_{n=1}^{l} P(C_n)\prod_{s=1}^{S} p(y_s/C_n)}\,. \tag{9}$$

Generally, the conditional probabilities are parameterized by some family distributions and all the parameters have to be estimated in a learning stage using methods such as the "maximum likelihood" and the so-called "expectation maximization" algorithms.[12] The determination of the *a priori* probabilities is not easy. In some related work, they are considered equiprobable and in others, they are estimated by the rate of the patterns belonging to the class in the learning set (i.e. ratio of the number of patterns of a class over the total number of patterns for all the classes).

Several decision methods are available. They are all based on the *a posteriori* probability $P(C_i/y_1, y_2, \ldots, y_S)$. The most popular one is the MAP (maximum *a posteriori* probability):

$$SC(x_k) = \begin{cases} C_j, & \text{if } P(C_j/y_1, y_2, \ldots, y_S) = \max\limits_{i=1}^{l} P(C_i/y_1, y_2, \ldots, y_S) \\ & \text{and} \quad P(C_j/y_1, y_2, \ldots, y_S) \geq T \\ C_{l+1}, & \text{else}\,. \end{cases} \tag{10}$$

where $0 < T \leq 1$ is a chosen reject threshold. Among other methods, one can state the "minimum expected risk",[10] which generalizes the MAP algorithm, or the "maximum entropy".

3.4. *Combination and decision by the Belief theory*

In some cases, the use of this rule of Bayes can be unsuited to the combination of decisions of several classifiers. Indeed, the rule of Bayes requires that the measurements behave like probabilities. In the case of the decisions of experts, this condition is often impossible to respect,[3] because an expert can decide to give a certain credit to two classes (mixed assumption), which cannot be expressed simply with the probability theory. The Belief theory[51] is one of the most known tools of reasoning for the representation of uncertainty and inaccuracy in knowledge based systems. Compared to the statistical and fuzzy approaches, the Belief theory has the advantage of including a modeling of uncertainty and inaccuracy during the combination of classifiers and during the assignment of a pattern to a class. Masses are assigned to all the elements of the power set 2^Ω, rather than to the elements of Ω only (Ω being the space of decision). Bayesian fusion is considered as a particular case of the fusion in the framework of Belief theory. The examples of cases where the Belief theory can be made profitable are:

- when the decision of a classifier is ambiguous, the theory takes this into account by giving masses to the union of classes,
- when the classifiers have different reliabilities; it is possible to take them into account by reducing or reinforcing their masses with a weighting,
- when the knowledge on the reliabilities of the classifiers depends on classes, this can be taken into account by modifying the masses according to this knowledge.

Let Ω, a limited set of l mutually exclusive assumptions $\omega_i, i \in \{1, \ldots, l\}$ and 2^Ω, the power set of Ω. Let A, be a subset of Ω and thus element of 2^Ω ($A \subseteq \Omega, A \in 2^\Omega$)). The fundamental concept representing uncertainty is that of a distribution of belief by a function called function of mass allocation, defined as an application m, of the power set 2^Ω in the interval $[0, 1]$, checking:

$$\begin{aligned} m(A) > 0 \quad &\text{if and only if } A \in 2^\Omega \\ \text{and} \quad &\sum_{A \subseteq \Omega} m(A) = 1 \,. \end{aligned} \tag{11}$$

Any subset of Ω such as $m(A) > 0$ is called the focal element of m. The $m(A)$ quantity can be interpreted as a measurement of the belief assigned exactly to A and none of its subsets, in the absence of other information. In this theory uncertainty related to an event $A \in 2^\Omega$ is measured using the Belief (Bel) and Plausibility (Pl) functions.

Bel(A) is understood as a total belief assigned to A:

$$\begin{aligned} \mathrm{Bel} : 2^\Omega &\longrightarrow [0, 1] \\ A &\longmapsto \mathrm{Bel}(A) = \sum_{B \subseteq A} m(B) \,. \end{aligned} \tag{12}$$

Pl(A) is understood as a belief assigned to A. Indeed, $1 - \mathrm{Pl}(A)$ is a measure of doubt in A. It is defined as follows:

$$\begin{aligned} &\mathrm{Pl} : 2^{\Omega} \longrightarrow [0,1] \\ &A \longmapsto \mathrm{Pl}\ (A) = \sum_{B \cap A = \emptyset} m(B) = 1 - \mathrm{Bel}\ (\bar{A})\,. \end{aligned} \tag{13}$$

The most significant operation concerning the distributions of belief is the orthogonal sum (or rule) of Dempster.[11] Two distributions of Belief m_1 and m_2 are known as combinable if there are at least two nondisjoined subsets B and C such as $m_1(B) > 0$ and $m_2(C) > 0$. The orthogonal sum of m_1 and m_2, noted $m_{1,2} = m_1 \oplus m_2$ is then defined by:

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) m_2(C)}{1 - \sum_{B \cap C = \emptyset} m_1(B) m_2(C)}\,. \tag{14}$$

For S sources, owing to the fact that the rule of Dempster is associative and commutative, the combination is done by successive combinations (in an unspecified order) and makes it possible to obtain the final distribution of belief m which is noted:

$$m = \bigoplus_{i=1}^{S} m_i\,. \tag{15}$$

Several other combination rules have been proposed in the literature, particularly through a disjunctive combination.[16,56]

Finally, a rule of decision must be chosen and applied, in order to retain, usually, the most appropriate class or assumption C_i: a first method is to transform the mass in a convenient probability distribution and apply the MAP algorithm to this probability. In the Transferable Belief Model,[55,58] the authors suggest the use of what they call the *pignistic*[c] probability, denoted as P_{bet} and defined as:

$$\forall\, C_i \in \Omega, \quad P_{\mathrm{bet}}(C_i) = \sum_{A \ni C_i} \frac{m(A)}{\mid A \mid} \tag{16}$$

where $\mid A \mid$ is the cardinality of A. Then, the decision rule is defined as:

$$SC(x_k) = \begin{cases} C_j, & \text{if} P_{\mathrm{bet}}(C_j) = \max_{i=1}^{l} P_{\mathrm{bet}}(C_i) \quad \text{and} \quad P_{\mathrm{bet}}(C_j) \geq T \\ C_{l+1}, & \text{else} \end{cases} \tag{17}$$

where $0 < T \leq 1$ is a chosen reject threshold.

Other decision rules have been proposed, generally over simple (as opposed to compound ones) assumptions[13,14]: for example, the maximum of plausibility, the maximum of credibility (Belief) and the rules using expected utility (see also Ref. 57).

[c]This term comes from "pignus": bet in Latin.

3.5. *Other theoretical frameworks for the combination*

Other theoretical frameworks also make it possible to carry out combinations, often with equivalent results. Just like it is difficult to prove that such a classifier is better than a similar one for any PR problem, it is not easy to prove that such theoretical framework is more appropriate than for any PR problem. In particular, the Fuzzy Sets theory and the Possibility theory[17,70] constitute powerful theoretical frameworks. Verikas *et al.*,[63] among others, carried out combinations using the fuzzy integrals of Sugeno and Choquet. In these frameworks $v_{is}(x_k)$ represents the membership degree of a fuzzy set or a possibility distribution.

A global evaluation of the pattern x_k membership of the class C_i is then provided by the combination of $v_{is}(x_k)$, with $s \in [1, S]$. The advantage of these frameworks rely on the variety of combination operators[4,15]: some of them are disjunctive (sum, maximum, and more generally triangular co-norm operators). Others are conjunctive (product, minimum and more generally triangular norm operators) and some others are neutral, for example, the ordered weighted averaging operators.[67] After the combination step, as in the common cases, decision is usually done with the maximum of membership values.

4. Cooperation of Classifiers

These methods use the decisions and/or the results of a classifier for better guiding one or more other classifiers in their decision-making. Generally, it is the decision vector which is transmitted between classifiers (in particular for Refs. 28, 34 and 35), because this type of information has less differences (in its form) from a type of classifier to another one (this leads to a more general method). This advantage leads to easier cooperation of various types of classifiers. However, other types of information can also be transmitted or received (e.g. a classifier could communicate the rejected classes).

The cooperation operation can induce either a reduction of the decision space,[7] or a reduction of the representation space, consequently, of the learning set,[27] or a reduction of the features space.[37,42] Thus, the classifiers entering into cooperation can be specialized in three different ways:

- in the feature space,
- in the decision space, and/or
- in the (features) representation space.

Figure 2 illustrates the possible interactions between the various classifiers of a system carrying out the cooperation operation. In this configuration, the vectors of features x_1, x_2 and x_3 concern a same pattern to be classified, but they can be different from/to each other. The classifiers e_1, e_2 and e_3 do not have necessarily a common decision space (it is the case, in particular, of the hierarchical classifiers).

A particular case of a system of multiple classifier carrying out cooperation operation is that of the multilevel (or hierarchical) classifiers put into practice

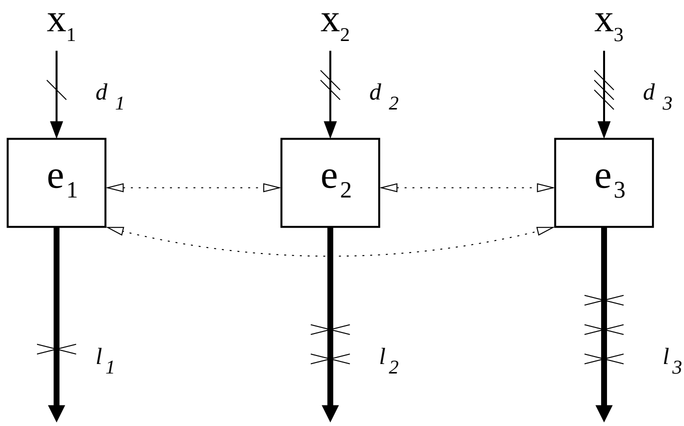


Fig. 2. System of classifiers carrying out the cooperation of classifiers. The vectors of features x_1, x_2 and x_3 concern the same pattern to be classified (but not necessarily as opposed to the combination, can we have $l_1 \cap l_2 \cap l_3 = \emptyset$) and they can be different from one classifier to another. Furthermore, the results of a classifier can be used in another classifier (in the learning and/or classification stage) in different ways. For example, e_1 can supply the learning set of e_3.

primarily by the neural networks community. This type of system is a particular case of cooperation, since from Fig. 2, when one eliminates the connection between e_2 and e_3, and changes into one-way the connections between e_1 and e_2 on the one hand and between e_1 and e_3 on the other hand (directed towards the right-hand side), a system is obtained of hierarchical classifiers where e_1 is in the higher level and e_2 and e_3 are in the lower level. Generally, they make it possible to reduce the decision space, a classifier of the higher level supplying its outputs to a classifier of the lower level which is able to classify the pattern introduced in the system's input. This can also be seen as a serial architecture (as opposed to the combination or parallel architecture). The system suggested by Cao *et al.*[7] is of this type. It also makes it possible to carry out an adaptive selection (see this term further, in the following section).

The reader can find in Table 1 the references of some work according to the system type of classifiers and the specialization type of the implied classifiers.

5. Selection of Classifiers

In the *Machine Learning* community, the majority of the current studies relates to the comparison of the various algorithms.[36] The qualification criterion generally adopted is that of the performance of these algorithms on standard sets of data. As specified by Salzberg,[50] since the performances of the classifiers strongly depend on the specific field in which they are applied, such comparisons can lead to errors. This is why the researchers of this community recommend the empirical evaluation of the algorithms in their application field. This, quite naturally, led to the static selection of classifiers. However, within the framework of the multiple classifier systems, the best classifier (or the best subset of classifiers) can be different from one pattern (to be classified) to another. This problem can be treated with adaptive selection and dynamic selection of classifiers.

According to Giacinto *et al.*,[23] the approaches used by the neural networks community can be classified into two strategies of design: the "direct" strategy and, referring to classifiers, the "over-produce and choose" strategy (see also Ref. 49). The first strategy consists in generating decorrelated neural networks directly (whose outputs can then be combined), while the second consists in creating a consequent initial set of networks and to choose the subsets of the decorrelated networks. This second strategy is in fact a strategy of *static* selection of classifiers.

Within the framework of the adaptive selection, several articles present methods of selection of a classifier using an estimate of the local performances in small areas of the representation space, surrounding an unknown test sample.[65] However, it is only the output of the best classifier for this data which is used for the final decision-making. These algorithms are often founded on methods of optimization, such as, for example, the genetic algorithms and the neural networks.

Some authors consider that adaptive selection and dynamic selection have the same meaning.[19,31] We make a distinction between these two concepts. In the case of the adaptive selection, if it is true that various classifiers or subsets of classifiers are selected according to the pattern introduced in input of the system, several patterns, identical but resulting from different moments, lead to the selection of the same classifier(s). The temporal aspect, i.e. the fact that a pattern is associated with a given moment t (and its features vector could be different at another moment), is thus not taken into account. A dynamic selection must, on the contrary, lead to different choices according to the moment considered. This can be useful when the classes are evolutionary (e.g. when the class centers move as a course of time).

In the cases of dynamic selection, for a given pattern x, the system retains the outputs of one or more classifier(s) adapted to the pattern (i.e. with the numerical values of the features of the pattern) to be treated, at the pattern evolution moment. The behavior of the system can be periodical: for two patterns x_{t1} and x_{t2} respectively taken at the moments t_1 and $t_2 = T + t_1$, where T is the periodicity of the behavior of the process, the system must select, in both cases, the same classifier or the same subset of classifiers. To our knowledge, the method that we proposed[26] constitutes the only example verifying this condition. In such cases, if the learning sets are constituted as a sum of all patterns, over a period of time, the static or the adaptive approaches can still be used. Thus, the dynamic selection can also be seen as a special case of the static or the adaptive ones. Therefore, we propose the following taxonomy:

Static selection: For any pattern x_k, at any moment t, the classifier or the subset of classifiers retained by the (multiple classifier) system is that defined in the learning stage. The strategy "over-produce and choose" of the neural networks community

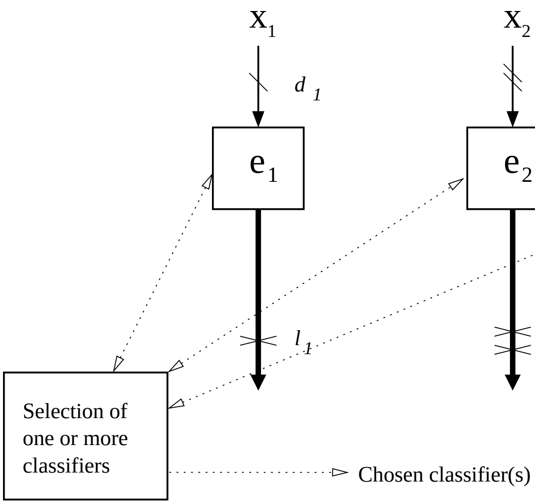


Fig. 3. System of classifiers carrying out the selection of classifiers. The selection module activates one or more classifiers according to the input pattern and/or according to time. The classifiers e_1, e_2 and e_3 do not have necessarily a common decision space. Thus, if only one classifier is selected at any time, we can have $l_1 \cap l_2 \cap l_3 = \emptyset$.

uses this type of selection. Giacinto *et al.*[22] and Sharkey *et al.*[54] proposed this type of systems.

Adaptive selection: For any pattern x, at any moment t, the system retains the outputs of one or more classifier(s) adapted to the numerical values of the features of the pattern to be classified. The 3C algorithm which we proposed[27] is an example of this type of selection. Woods *et al.*[65] and Cao *et al.*[7] also proposed this type of systems.

Figure 3 illustrates the possible interactions between the various classifiers of a system carrying out the selection operation. In this configuration, the vectors of features x_1, x_2 and x_3 concern the same pattern to be classified, but they can be different from/to each other. The classifiers e_1, e_2 and e_3 do not have necessarily a common space of discernment. On the other hand, if one wants to retain the results of several classifiers, it is necessary to respect the constraints of the combination. Thus, one can associate with this selection, a combination and/or a cooperation of classifiers. There is then a *mixed* or *hybrid approach* of multiple classifier systems.

6. Hybrid Systems of Classifiers and Their Topology

Some systems of classifiers carry out, at the same time, the combination, the cooperation and/or the selection (static, adaptive or dynamic) operations. The system suggested by Jaimes and Chang[37] is of this type. It allows the combination and the cooperation of classifiers. The cooperation is done between hierarchical classifiers, i.e. the decisions of the classifiers of low levels are used as inputs for classifiers for higher level who make decisions of higher levels.

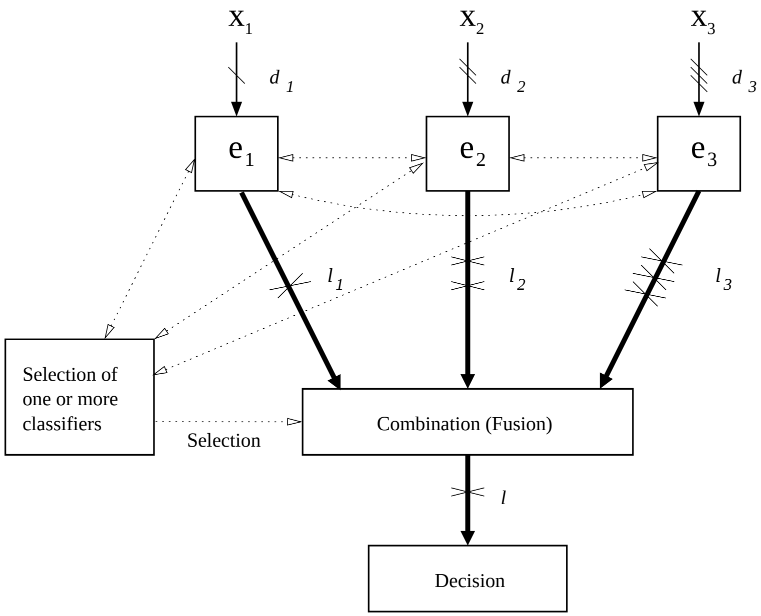


Fig. 4. System of classifiers carrying out operations such as combination, cooperation and selection of classifiers. The available features are $F = \{\{x_1\}, \{x_2\}, \{x_3\}\}$ and d_1, d_2 and d_3 are the dimensions of x_1, x_2 and x_3, respectively. Each of these classifiers can have a different decision space. In the example, the decision spaces of the classifiers e_1, e_2 and e_3 are l_1, l_2 and l_3, respectively.

Such systems are called hybrid systems of classifiers and Fig. 4 illustrates the possible interactions between the different classifiers of such a system. The connections in dotted lines, between the various "modules" represent various information. This information can be communicated during the training and the classification steps. In this way, the module of selection is an intelligent module which retains some information resulting in the training step and uses it in the classification step. A classifier can transmit its decisions to one or more other classifiers.
Examples of information exchange:

- The classifier e_1 transmits its training set to e_3 (step of system training).
- The classifier e_1 informs the module of selection that it is necessary or not to take into consideration the outputs of classifier e_3 (in classification step).
- The module of selection "requests an answer" from classifiers e_1 and e_2, concerning an unknown pattern available in input.

The connections in full line represent the features (with bars) or the decisions (with crosses). In this configuration, the vectors of features x_1, x_2 and x_3 concern the same pattern to be classified, but they can be different from/to each other. If we have d features, then we have:

$$d_1 \leq d, \quad d_2 \leq d \text{ and } d_3 \leq d\,.$$

Table 1. Examples of systems of classifiers. To our knowledge, some solutions are not treated yet (empty cases).

Specialization of Classifiers		Combination of Classifiers	Cooperation of Classifiers	Selection	
⇓ Spaces ⇓	⇓ Methods ⇓			Static	Adaptive
Features	Heuristics and various	Van Breukelen *et al.*[62] Chen *et al.*[8] Xu *et al.*,[66] Ho [32]	Jaimes and Chang[37]	Sharkey *et al.*[54] "Test and Select"	Gosselin[25]
Decision	Different principals	Giacinto *et al.*[22] Ho *et al.*[30]	Giacinto and Roli[20]	Giacinto *et al.*[22]	Woods *et al.*[65]
	Different parameters or initializations	Ng and Singh[45] Rogova[48] Cho and Kim[9]	Giacinto and Roli[20] Cao *et al.*[7]	Giacinto *et al.*[22] Sharkey *et al.*[54] "Test and Select"	Cao *et al.*[7] "Multiple features and multistage classifiers"
Representation	Random selection	Wolpert[64] "Stacked generalization" Ho[33] "Random subspace"			Giacinto and Roli[19] "Adaptive selection"
	Clustering	Gunes *et al.*,[27] "3C"	Gunes *et al.*,[27] "3C"		Gunes *et al.*,[27] "3C"
	Temporal selection	Gunes *et al.*[26]			

The classifiers e_1, e_2 and e_3 do not have necessarily a common decision space. The dimension of Ω being l (combination output), we have $\dim(l_1) \leq l$, $\dim(l_2) \leq l$ and $\dim(l_3) \leq l$. When only one classifier is selected, there is no operation of combination. On the other hand, if we want to keep the results of several classifiers in order to combine them, it will then be preferable that the selected classifiers have a common decision space. If this latter constraint is not respected, for example, when a classifier decision is a compound class (two or more classes among the decision space), the combination is treated, generally, through the Belief theory of Dempster–Shafer.[51] This theory is well suited when the outputs of the classifiers are the richest in terms of information (type 3 outputs: they are not crisp). When the outputs are crisp, other methods, based on an ordering of all the classes are also available.

Example: If the selection recommends the choice of the results of the classifiers e_1 and e_2, it is preferable that the l classes or assumptions are the same ones from one classifier to another

$$\begin{cases} \dim(l_1) = \dim(l_2) = l \\ \text{and } \Omega_1 = \Omega_2 = \Omega = \{C_1, C_2, \ldots, C_l\}\,. \end{cases} \tag{18}$$

The module of selection can be based on an algorithm of clustering/classification (it is the case for the 3C system, suggested in Ref. 27), while a classifier e_i can itself consist of several classifiers (i.e. of a system of classifiers). In the *conditional topology* approach,[40] the selection is based only on the reject option, i.e. a selection is carried out only if the first classifier(s) reject(s) the input pattern. However, according to the results of a first classifier (selection classifier), many choices are possible.

Some authors[27] have proposed a system which carries out the three types of operations (combination, cooperation and adaptive selection). Table 1 summarizes existing work and all the possibilities offered within the framework of the systems of classifiers. The hybrid systems are easily identifiable by the fact that one finds them or not in several columns (in this way, the two columns of *selection* are merged into one column). The table shows that the systems of classifiers can carry out different types of operations and that, in a complementary way, the individual classifiers can be specialized on the features, on the classes (in the decision space), or in the representation space.

7. Conclusion

One can notice that in the literature, the combination, the cooperation and the selection (static or adaptive) are already used in the *Systems of Classifiers* to recognize static classes. When the classes have complex forms, these operations make it possible, generally, to obtain better rates of recognition, compared to one classifier only. The adaptive selection is sometimes confused with the dynamic selection of classifiers. This is due to the fact that the temporal aspect was neglected, whereas it should be taken into account explicitly.

Furthermore, it appears clearly that a formalization of the systems of classifiers is needed. Indeed, a wide variety of systems are proposed with various architectures, various terminologies and various validations. Therefore, the advantages and the drawbacks of the proposed approaches cannot be proved easily. Such a formalization, defined as a set of classifiers and a set of relations between these classifiers, would permit comparative studies. In the field of dynamic selection of classifiers, such a proposal would clarify the selection procedure and the contribution of each classifier to the global task.

The *Systems of Classifiers* could be useful in the applications with distributed data because they are predisposed with the implementation according to this type of architecture. Indeed, when the distant data are obtained with a high flow/low distance, the strategy of the combination between local classifiers and the distant classifier can be the most effective because of its precision. When they are obtained according to a medium flow/medium distance, the cooperation with distant classifiers can be the strategy which is the best adapted (possibility of reducing the quantity of exchanged information). Lastly, when the data are obtained with a low flow/long distance or when they can undergo temporal evolutions, the selection of distant classifiers can be the best strategy.

References

1. B. Achermann and H. Bunke, "Combination of classifiers on the decision level for face recognition," Report IAM-96-002, Institut fü Informatik, University of Bern, Switzerland, January 1996.
2. F. M. Alkoot and J. Kittler, "Improving the performance of the product fusion strategy," *Proc. ICPR2000, 15th Int. Conf. Pattern Recognition*, Barcelona, Spain, September 2000, pp. 164–167.
3. D. Bahler and L. Navarro, "Methods for combining heterogeneous sets of classifiers," *17th Natl. Conf. Artificial Intelligence (AAAI 2000), Workshop on New Research Problems for Machine Learning*, 2000.
4. I. Bloch, "Information combination operators for data fusion: a comparative review with classification," *IEEE Trans. Syst. Man Cybern.* **26**, 1 (1996) 52–67.
5. I. Bloch and H. Maître, "Fusion of Image Information under Imprecision," *Aggregation and Fusion of Imperfect Information*, ed. B. Bouchon-Meunier, Series Studies in Fuzziness, Physica Verlag, Springer, 1997, pp. 189–213.
6. L. Breiman, "Bagging predictors," *Mach. Learn.* **24**, 2 (1996) 123–140.
7. J. Cao, M. Ahmadi and M. Shridhar, "Handwritten numerals with multiple features and multistage classifiers," *IEEE Int. J. Circuits Syst.* **6** (1996) 323–326.
8. K. Chen, L. Wang and H. Chi, "Methods of combining multiple classifiers with different features and their applications to text-independant speaker identification," *Int. J. Pattern Recognition and Artificial Intelligence* **11**, 3 (1997) 417–445.
9. S. B. Cho and J. H. Kim, "Combining multiple neural networks by fuzzy integral for robust classification," *IEEE Trans. Syst. Man Cybern.* **SMC-25**, 2 (1995) 380–384.
10. V. M. Dang, "Classification de données spatiales: modèles probabilistes et critères de partitionnement," Ph.D. Thesis, Université de Compiègne, 1998.
11. A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.* **38** (1967) 325–339.

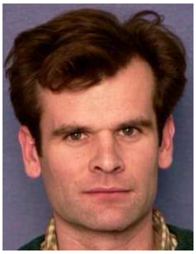
12. A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. Roy. Statist. Soc.* **B39** (1977) 1–38.
13. T. Denœux, "A k-nearest neighbor classification rule based on Dempster–Shafer theory," *IEEE Trans. Syst. Man Cybern.* **25**, 5 (1995) 804–813.
14. T. Denœux, "Analysis of evidence-theoretic decision rules for pattern classification," *Patt. Recogn.* **30**, 7 (1997) 1095–1107.
15. D. Dubois and H. Prade, "A review of fuzzy sets aggregation connectives," *Inform. Sci.* **36** (1985) 85–121.
16. D. Dubois and H. Prade, "On the unicity of dempster rule of combination," *Int. J. Intell. Syst.* **1** (1986) 133–142.
17. D. Dubois and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, NY, 1988.
18. J. Franke and E. Mandler, "A comparison of two approaches for combining the votes of cooperating classifiers," *Proc. 11th Int. Conf. Pattern Recognition*, Vol. 2, The Hague, Netherlands, 1992, pp. 611–614.
19. G. Giacinto and F. Roli, "Adaptive selection of image classifiers," *ICIAP'97*, *9th Int. Conf. Image Analysis and Processing*, Florence, Italy, September 1997, Lecture Notes in Computer Science 1310, Springer, pp. 38–45.
20. G. Giacinto and F. Roli, "Ensembles of neural networks for soft classification of remote-sensing images," *Proc. Eur. Symp. Intelligent Techniques*, Bari, Italy, March 1997, pp. 166–170.
21. G. Giacinto and F. Roli, "Methods for dynamic classifier selection," *ICIAP'99*, *10th Int. Conf. Image Analysis and Processing*, Venice, Italy, September 1999, pp. 659–664.
22. G. Giacinto and F. Roli, "An approach to the automatic design of multiple classifier systems," *Patt. Recogn. Lett.* **22** (2001) 25–33.
23. G. Giacinto, F. Roli and G. Fumera, "Design of effective multiple classifier systems by clustering of classifiers," *Proc. ICPR2000*, *15th Int. Conf Pattern Recognition*, Barcelone, Espagne, September 2000, pp. 160–163.
24. A. Giani, "Parallel cooperative classifier systems," Ph.D. Thesis, Université de Pise, Italie, March 1999.
25. B. Gosselin, "Cooperation of multilayer perceptron classifiers," *Proc. 8th Works Circuits Systems and Signal Processing*, Mierlo, Pays-Bas, 1997, pp. 187–190.
26. V. Gunes, P. Loonis and M. Ménard, "A fuzzy petri net for pattern recognition: application to dynamic classes," *Knowl. Inform. Syst.* **4**, 1 (2002) 112–128.
27. V. Gunes, M. Ménard and P. Loonis, "A multiple classifier system using ambiguity rejection for clustering-classification cooperation," *Int. J. Uncert. Fuzz. Knowl. Based Syst.* **8**, 6 (2000) 747–762.
28. D. Happel and P. Bock, "Overriding the experts: a stacking method for combining marginal classifiers," *Proc. 13th Int. FLAIRS Conf.*, Menlo Park, CA, USA, 2000.
29. R. M. Haralick, "The table look-up rule," *Commun. Stat. — Th. Meth.* **A5**, 12 (1976) 1163–1191.
30. T. K. Ho, "A theory of multiple classifier systems and its application to visual word recognition," Ph.D. Thesis, State University of New York at Buffalo, May 1992.
31. T. K. Ho, J. J. Hull and S. N. Srihari, "Decision combination in multiple classifier systems," *IEEE Trans. Patt. Anal. Mach. Intell.* **16** (1994) 66–75.
32. T. K. Ho, "Random decision forests," *Proc. 3rd Int. Conf. Document Analysis and Recognition*, Montreal, Canada, August 1995, pp. 278–282.
33. T. K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. Patt. Anal. Mach. Intell.* **20**, 8 (1998) 832–844.
34. T. S. Huang and C. Y. Suen, "Combination of multiple experts for the recognition of

unconstrained handwritten numerals," *IEEE Trans. Patt. Anal. Mach. Intell.* **17**, 1 (1995) 90–94.
35. Y. S. Huang and C. Y. Suen, "A method of combining multiple classifiers — a neural network approach," *Proc. 12th Int. Conf. Pattern Recognition and Computer Vision*, Jerusalem, 1994, pp. 473–475.
36. A. Jaimes and S. F. Chang, "Automatic selection of visual features and classifiers," *IS&T/SPIE Storage and Retrieval for Image and Video Databases VIII*, San Jose, CA, USA, January 2000.
37. A. Jaimes and S. F. Chang, "Integrating multiple classifiers in visual objects detectors learned from user input," *4th Asian Conf. Computer Vision* (*ACCV2000*), Taipei, Taiwan, January 2000.
38. F. Kimura and M. Shridar, "Handwritten numerical recognition based on multiple algorithms," *Patt. Recogn.* **24**, 10 (1991) 969–983.
39. L. I. Kuncheva, C. A. Whitaker, C. A. Shipp and R. P. W. Duin, "Is independence good for combining classifiers?" *Proc. ICPR2000, 15th Int. Conf. Pattern Recognition*, Barcelone, Espagne, September 2000, pp. 168–171.
40. L. Lam, "Classifier combinations: implementations and theoretical issues," *Proc. First Int. Work. Multiple Class. Syst.*, Lecture Notes in Computer Science, 2000, pp. 77–86.
41. L. Lam and C. Y. Suen, "A theoritical analysis of the application of majority voting to pattern recognition," *Int. Conf. Pattern Recognition*, Jerusalem, 1994, pp. 418–420.
42. P. Loonis and M. Ménard, "Fusion connexionniste et algorithme gntique pour adapter les systémes de reconnaissance des formes á leur environnement," *Traitement du Signal*, November 1996.
43. S. Mathevet, L. Trassoudaine, P. Checchin and J. Alizon, "Combinaison de segmentations en régions," *Traitement du Signal* **16**, 2 (1999) 93–104.
44. G. J. McLachlan and K. E. Basford, *Mixture Models*: *Inference and Applications to Clustering*, Marcel Dekker, NY, July 1988.
45. G. S. Ng and H. Singh, "Data equalisation with evidence combination for pattern recognition," *Patt. Recogn. Lett.* **19** (1998) 227–235.
46. N. J. Nilsson, *Learning Machines*, McGraw-Hill, 1965.
47. J. R. Quinlan, "Bagging, boosting, and c4.5," *Proc. 13th Nat. Conf. Artificial Intelligence*, Portland, Oregon, USA, August 1996, pp. 725–730.
48. G. Rogova, "Combining the results of several neural network classifiers," *Neural Networks* **7**, 5 (1994) 777–781.
49. F. Roli and G. Giacinto, "Design of multiple classifier systems," *Hybrid Methods in Pattern Recognition*, eds. H. Bunke and A. Kandel, Series on Machine Perception and Artificial Intelligence, World Scientific, 2002.
50. S. L. Salzberg, "On comparing classifiers: a critique of current research and methods," *Data Min. Knowl. Discovery* **1** (1999) 1–12.
51. G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
52. R. E. Shapire, Y. Freund, P. Bartlett and W. S. Lee, "Boosting the margin: a new explanation for the effectiveness of voting methods," *Proc. 14th Int. Conf. Machine Learning*, 1997.
53. A. J. C. Sharkey, "Multi-net systems," *Combining Artificial Neural Nets*, *Ensemble and Modular Multi-Net Systems*, 1999, pp. 1–27.
54. A. J. C. Sharkey and N. E. Sharkey, "The 'test and select' approach to ensemble combination," *Proc. First Int. Works. Multiple Classifier Systems*, Lecture Notes in Computer Science, 2000, pp. 30–44.
55. P. Smets, "The combination of evidence in the transferable belief model," *IEEE Trans.*

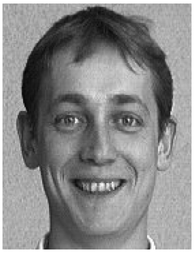
Patt. Anal. Mach. Intell. **12**, 5 (1980) 447–458.
56. P. Smets, "Belief functions: the disjuntive rule of combination and the generalized Bayesian theorem," *Int. J. Approx. Reason.* **9** (1993) 1–35.
57. P. Smets, "Decision analysis using belief functions," *Int. J. Approx. Reason.* **4** (1994) 391–418.
58. P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.* **66** (1994) 191–243.
59. C. Y. Suen, C. Nadal, R. Legault, T. A. Mai and L. Lam, "Computer recognition of unconstrained handwritten numerals," *Proc. IEEE* **80** (1992) 1162–1180.
60. C. Y. Suen, T. A. Nadal, T. A. Mai, R. Legault and L. Lam, "Recognition of totally unconstrained handwritten numerals based on the concept of multiple experts," *Frontiers in Handwriting Recognition*, Montreal Corcordia University, 1990, pp. 131–143.
61. K. Tumer and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," *Connect. Sci.* **8** (1996) 385–404.
62. M. Van Breukelen, R. P. W. Duin and D. M. J. Tax, "Combining classifiers for the recognition of handwritten digits," *1st IAPR TC1 Statistical Technniques in Pattern Recognition*, Prague, Tchéquie, June 1997, pp. 13–18.
63. A. Verikas, A. Lipnickas, K. Malmqvist, M. Bacauskiene and A. Gelzinis, "Soft combination of neural classifiers: a comparative study," *Patt. Recogn. Lett.* **20** (1999) 429–444.
64. D. H. Wolpert, "Stacked generalization," *Neural Networks* **5**, 2 (1992) 241–259.
65. K. Woods, W. P. Kegelmeyer and J. R. K. Bowyer, "Combination of multiple classifiers using local accuracy estimates," *IEEE Trans. Patt. Anal. Mach. Intell.* **19**, 4 (1997).
66. L. Xu, A. Krzyzak and C. Y. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," *IEEE Trans. Syst. Man Cybern.* **22**, 3 (1992) 418–435.
67. R. R. Yager, "On ordered weighted averaging aggregation operators in multi-criteria decision making," *IEEE Trans. Syst. Man Cybern.* **18** (1988) 183–190.
68. K. Yu, X. Jiang and H. Bunke, "Combining acoustic and visual classifiers for the recognition of spoken sentences," *Proc. ICPR2000, 15th Int. Conf. Pattern Recognition*, Barcelona, Spain, September 2000, pp. 491–494.
69. L. A. Zadeh, "Fuzzy sets," *Inform. Contr.* **8** (1965) 338–353.
70. L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst. Man Cybern.* **3** (1973) 28–44.
71. E. Zahzah, "Contribution à la représentation des connaissances et à leur utilisation pour l'interprétation automatique des images satellites," Ph.D. Thesis, Université Paul Sabatier, Toulouse, September 1992.

V. Gunes graduated from the University of Haute Alsace (France) in electrical engineering in 1994. In 2001, he received the Ph.D. degree from the University of La Rochelle, France. He is currently a research fellow at the Laboratory of Computing, Images and Interactions at the University of La Rochelle.

His current research interests include pattern recognition, computer vision and image processing.



P. Loonis received a Ph.D. in image and signal processing from the University of La Rochelle (France) in 1996. His work deals with the analysis of dynamic systems in relation to pattern recognition and artificial intelligence — tracking the drift of clusters, web profiling, classification of temporal series, ... as soon as the time index appears; combining various classifiers can bring a solution to a high semantic level; a contrario, other approaches try to express the temporal data in a specific space where common pattern recognition tools can be relevant.



M. Menard is currently an assistant professor at the University of La Rochelle, France. He holds a Ph.D. degree from the University of Poitiers, France (1993), and a HDR from the University of La Rochelle (2002).

His research interests are fuzzy pattern recognition, image processing and data fusion with particular applications to medical imaging.

S. Petitrenaud graduated in 1995 as a statistician from the Ecole Nationale de la Statistique et de l'Administration Economique in Paris, France and obtained a Ph.D. in 1999 from the Université de Compiégne, France. After a postdoctorate at the Ecole Nationale Supérieure des Télécommunications in Paris, he joined the Laboratoire d'Informatique de l'Université du Maine as an Assistant Professor in 2000.

His current research interests include statistical pattern recognition, speech recognition, image processing, data fusion and uncertainty modeling in statistical inference.

This article has been cited by:

1. Abdellah Lamallem, Lionel Valet, Didier Coquin. 2010. Local Evaluation of a Fusion System for 3-D Tomographic Image Interpretation. *International Journal of Optomechatronics* **4**:4, 362-378. [CrossRef]
2. A. Ibarguren, I. Maurtua, B. Sierra. 2010. Layered architecture for real time sign recognition: Hand gesture and movement. *Engineering Applications of Artificial Intelligence* **23**:7, 1216-1228. [CrossRef]
3. P. Soda, G. Iannello. 2009. Aggregation of Classifiers for Staining Pattern Recognition in Antinuclear Autoantibodies Analysis. *IEEE Transactions on Information Technology in Biomedicine* **13**:3, 322-329. [CrossRef]
4. Lei Chen, Mohamed S. Kamel. 2009. A generalized adaptive ensemble generation and aggregation approach for multiple classifier systems. *Pattern Recognition* **42**:5, 629-644. [CrossRef]
5. Albert H.R. Ko, Robert Sabourin, Alceu Souza Britto, Jr.. 2008. From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition* **41**:5, 1718-1731. [CrossRef]
6. C. L. He, C. Y. Suen. 2007. A hybrid multiple classifier system of unconstrained handwritten numeral recognition. *Pattern Recognition and Image Analysis* **17**:4, 608-611. [CrossRef]
7. J.M. Martínez-Otzeta, B. Sierra, E. Lazkano, A. Astigarraga. 2006. Classifier hierarchy learning by means of genetic algorithms. *Pattern Recognition Letters* **27**:16, 1998-2004. [CrossRef]
8. Paolo Lombardi, Bertrand Zavidovique, Michael Talbert. 2006. On the Importance of Being Contextual. *Computer* **39**:12, 57-61. [CrossRef]
9. HEE-JOONG KANG, SEONG-WHAN LEE. 2005. COMBINATION OF MULTIPLE CLASSIFIERS BY MINIMIZING THE UPPER BOUND OF BAYES ERROR RATE FOR UNCONSTRAINED HANDWRITTEN NUMERAL RECOGNITION. *International Journal of Pattern Recognition and Artificial Intelligence* **19**:03, 395-413. [Abstract] [References] [PDF] [PDF Plus]