Journal of Statistics: Advances in Theory and Applications
Volume 1, Number 1, 2009, Pages 1-17

AN EFFICIENT STATISTICAL APPROACH TO HANDWRITTEN DIGIT RECOGNITION

CRESCENZIO GALLO

Dipartimento Di Scienze Economiche
Matematiche E Statistiche
Centro Di Ricerca Interdipartimentale Bioagromed
Università di Foggia
Largo Papa Giovanni Paolo II n.1
71100 Foggia
Italy
e-mail: c.gallo@unifg.it

Abstract

Often a pattern recognition problem is too hard to solve with ordinary approaches involving too computationally “expensive” algorithms. Instead of this, it is sometimes better to combine the decisions coming from different simpler methods, working on the same data to produce a solution to the same problem, in order to give a global result better than that obtained by each single expert.

In this paper, this technique is applied to the well-known problem of handwritten digit recognition, where it is shown how the employment of a good combination applied to very simple, fast and with low single performances methods can catch up optimal reliability and performance comparable to those of “big” but slow and complex methods.

1. Introduction

This work investigates some techniques for the combination of decisions coming from different “expert” methods (algorithms), working

2000 Mathematics Subject Classification: 62-07, 68W20.

Keywords and phrases: pattern recognition, integration of experts.

Received January 4, 2009

on the same data and aiming at producing a solution to the same problem, in order to give a global result better than that obtained by each single expert. This tendency, taking advantage of the idea that “more heads together work better than a single one”, is commonly called *Combination of Multiple Experts* (C.M.E.), and in our case is applied to handwritten digit recognition. The “experts” (whose decisions are integrated) are some elementary algorithms of handwritten digit recognition, whose combined performances turn out to be clearly superior to those of each single method.

2. Combination of Experts

The adopted scheme considers an expert like a “black box”, equipped with an inferential mechanism, to which the data of a problem are given in input and that supplies a solution (the output). Such a solution can be single-output, multiple-outputs, or reject option depending on the nature of the problem and the characteristics, essentially nondeterministic, of the inferential mechanism. Inside the black box are embedded some “decision criteria” studying some important points of the problem considered relevant from particular points of view.

The aspects to consider in a decisional process that carries out a coherent synthesis of the results supplied by multiple classifiers (i.e., an *expert*) are several:

- some experts can systematically be more effective (*dominant* classifiers) and they can make useless or not convenient to use the poorer (or *dominated*) classifiers;
- some experts adopt equal or similar decisional criteria and create situations of “coalition”, that may unfavorably bias the behavior of all classifiers combination;
- some experts can supply a single solution, others instead produce a list of solutions, typically according to a probabilistic order.

The values representing the decisions are generally expressed in different scales by the various classifiers. Everyone of these problems leads to some questions, to be resolved before preparing a method able to effectively arrange the various decisions:

1. *Are the less effective experts to be discarded*?[1] If the less effective expert contributes to the final decision with a “point of view” different from all the others, then it, however, brings a positive contribution and it is therefore opportune to hold it into account.

2. *Are equal or similar experts to be discarded*?[2] For those equal the answer is affirmative, since their “point of view” is the same. For the similar ones there, however, exist some C.M.E. methods that do not demand classifiers' independence, such as the BKS (Behavior-Knowledge Space) method developed by the OCR researchers team of Concordia (see [8]).

3. *How are single-response and multiple-response experts to be considered*?[3] Currently two different ways are followed, according to the experts supplying one decision (the first way) or more (the second one). But, in theory, every list of decisions can become single extracting the most reliable one and evaluating it according to appropriate value scales.

4. *Experts respond with boolean*, *integer*, r*eal values or do not respond*? A normalization is required of the values representing the decision. As previously pointed out, the two possible approaches in the C.M.E. view are:

- the first, based upon classifiers providing a single decision;
- the second, based upon classifiers offering a list of measures.

---

[1] This also refers to “imbalance of expert accuracies”, addressed comprehensively by [7].

[2] This is the question of committee “diversity”, addressed comprehensively by many authors, notably Kuncheva [10, 11].

[3] This - and the next - question leads to a genuinely interesting research question, and not one often addressed by the classifier combination literature: *how do you fuse multi-modal outputs*, that is, classifier outputs in different forms.

In this work, we adopted both the first and the second approach, upon which a study has already been done by the Buffalo's Srihrari team [12] who analyzed several schemes of classifiers integration (*Bayesian Approach*, *Logistic Regression*, *Neural Network*, Fuzzy *Integral and Majority Vote*). Of these, here we analyze *Majority Vote* and *Bayesian Approach*, besides of the integration schemes known as *First Probability Driven* and *Majority Weighted Vote*.

2.1. Description of an integrator

A recognizer is an expert. It supplies the integrator with a list of *M* elements, each one representing the confidence-or reliability - level associated to every class, i.e., a measure of how much each class is near to the input pattern. Some integrators work on the entire table provided by the classifiers (*Majority Vote*, *Bayesian Approach*, *Majority Weighted Vote*), others instead use only the most reliable value (*First Probability Driven*).

In any case, the answer is unique (determined by a posterior probability) and lies within the set $\{0, \dots, M-1\}$ if the input is recognized (and therefore assigned to each one of the *M* classes), or *M* if it is rejected.

2.2. Majority vote

Given an input pattern $x \in X$ belonging to one of *M* classes C_0, C_1, $\dots$, C_{M-1}, every R_k recognizer can be considered like a mapping function from the patterns' domain *X* to an *M*-dimensional vector space V^M defined on the set of the real numbers.

The *j*-th member of *V* denotes the confidence level with the C_j class, that can be an Euclidean distance (measuring the simple “vicinity” between pattern and the C_j class) or a value obtained from a relaxed matching rather measuring the “similarity” degree. *V* is processed by S_1, $\dots$, S_K transformation functions (score functions), each assigning an integer positive score.

The scores vectors PTS_i are therefore obtained. Finally, an integrator, denoted with *I*, processes vectors $PTS_1, \dots, PTS_K$ and produces a new vector *T* of the totals.

Synthesizing:

$$\begin{aligned} V_i &= R_i(x) \quad (\text{recognizer } R_1, \ldots, R_K : \{\text{patterns' domain}\} \to \mathbf{R}^M), \\ PTS_i &= S_i(V_i) \quad (\text{score } S_1, \ldots, S_K : \mathbf{R}^M \to \mathbf{N}^M), \end{aligned}$$

$$T = I(PTS_1, \ldots, PTS_K) \quad (\text{combination } I : \mathbf{N}^{M \times K} \to \mathbf{N}^M).$$

The decision rule for accepting or rejecting the pattern is[4]:

$$R(x) = \begin{cases} j, & \text{if} \quad T(j) = \sum_{k=1}^{K} PTS_k(j) = \max_{\{i=0, \ldots, M-1\}} T(i) > \lambda \\ M, & \text{otherwise,} \end{cases}$$

where $j \in \{0, \ldots, M-1\}$ and λ is an appropriate threshold equal to $K \cdot M \cdot \alpha$, with K denoting the number of classifiers, M the number of classes and the maximum attributable score, and $\alpha \in [0, 1]$.

The integrator, after receiving the first list from the first method, executes a series of operations in order to assign the score to every class in relation to the values of such list. If the list supplied by a method contains the minimal distances of every class with reference to the input pattern, then a score is assigned, ranging from M to 1, starting from the class showing the minimal distance until that recording the maximum one. If, instead, the list (or vector) provided by the method contains similitude values or vicinity or matching scores, as greater as more coincides the input pattern with the class in the database, then the score, from M to 1, is assigned starting from the class that recorded the maximum list's value until that one showing the minimum value.

After receiving the lists from every method and having assigned scores to all classes in the list, the integrator must add all computed scores in order to obtain only one final list of total scores, each for every class. Then the input pattern is recognized as the class that, in the final

[4] See work by [10, 11, 13] on properties of majority voting.

list, obtained the maximum value and exceeds the threshold λ. In order to increase the accuracy of the integration method a second threshold has been introduced, to the aim of verifying the condition that the difference between the two highest scores is greater than the security value α[5].

2.3. Application of the Bayes method

Though we improperly call this an application of the "Bayes" method, it is actually a linear combination of posterior probabilities, a paradigm extensively covered by (for example) [18], [7], [22].

For a Bayes classifier e [21], the classification of an input pattern x is based upon a set of real values of a posteriori probability measures:

$$P(x \in C_i | x) \quad (i = 0, \dots, M-1), \tag{1}$$

(where $x \in C_i$ indicates that x belongs to class C_i), also simply denoted by:

$$P(C_i | x) \quad (i = 0, \dots, M-1).$$

They represent the probabilities that x belongs to every and each of the M classes under the condition given by the input pattern x. In principle, these probabilities are not concerning every classifier e_k. In practice, every e_k that classifies x is not really based on the values of (1) which are not available. Instead, for each x, e_k estimates a set of approximations for that real values. These approximations depend on the type of "feature" that e_k uses and on how e_k is "trained".

In order to clearly state what above said let us make use of the following notation:

$$P_k(C_i | x) \quad (i = 0, \dots, M-1;\ k = 1, \dots, K). \tag{2}$$

[5] In fact, if the difference between the two values is too much low, it means there that is an ambiguity in the pattern recognition and therefore this has to be rejected in order to avoid a too high substitution error probability.

For every e_k, a definitive decision is:

$$e_k(x) = j, \text{ with } P_k(C_j|x) = \max_{\{i=0,\, \dots,\, M-1\}} P_k(C_i|x). \tag{3}$$

An approximation is made for the values of (2), so as to be able to classify an input pattern x based on the result given by the K classifiers, making a combination of these results. The used approach [21] proposes the following use of the following mean value as a new estimate of the classifier E:

$$P_e(C_i|x) = \frac{1}{K}\sum_{k=1}^{K} P_k(C_i|x) \quad (i = 0, \dots, M-1). \tag{4}$$

The final decision made by E is given by:

$$E(x) = j, \text{ with } P_E(C_j|x) = \max_{\{i=0,\, \dots,\, M-1\}} P_E(C_i|x), \tag{5}$$

i.e., a Bayes decision is based upon this a posteriori newly estimated probabilities.

In order to make classifier's results more reliable one may use the formula (6) that replaces (5) in order to make an exchange between the substitution and rejection percentage. In (6) we have the threshold $\alpha \in [0, 1]$:

$$E(x) = \begin{cases} j, & \text{if } P_E(C_j|x) = \max_{\{i=0,\, \dots,\, M-1\}} P_E(C_j|x) \geq \alpha, \\ M, & \text{otherwise.} \end{cases} \tag{6}$$

As far as concerns the approximation of (2) a function must be found being able to provide values that satisfy the base axioms of the probability theory. The function here taken into account depends on the type of classifier used. In fact, for algorithms providing a list of $d_k(i)$ in which the lowest value is the one denoting the highest reliability, (e.g., Euclidean distances of every class with reference to the test pattern) it can be used:

$$p_k(i) = \frac{\frac{1}{d_k(i)}}{\sum_{j=0}^{M-1} \frac{1}{d_k(j)}} \quad (i = 0, \dots, M-1).$$

Instead, for the algorithms supplying tables of values $rel_k(i)$ directly proportional to the likeness of the input pattern as those present in the database, it can be used:

$$p_k(i) = \frac{rel_k(i)}{\sum_{j=0}^{M-1} rel_k(j)} \quad (i = 0, \ldots, M-1).$$

We can thus say that, for every classifier supplying whichever appearing a posteriori probability, (6) can be used for a possible comb-ination.

2.4. First probability driven

This technique, unlike the two previous, takes into account only one solution, and therefore does not work on the entire table of M elements provided by each classifier. It is based on an a priori study carried out on the a posteriori reliability of each of the experts, runnable through an analysis of the confusion tables of the experts themselves, exercised on of some training patterns.

The preliminary study serves to statistically calculate the probability that, if the expert's decision is "x", the input is really "x"; that is:

$$P(\text{input} = x|\text{output} = x), \text{ with } x = 0, 1, \ldots, M-1.$$

The base concept that governs the final decision is this***: if two statistically independent methods give the same solution, then the probability that they are both wrong is much lower than the probability that only one of them mistakes,*** and it is possible to compute it just starting from these last two.

Let therefore $E_1, E_2, \ldots, E_K$ the experts, everyone of which provides an answer C_i, with $i = 1, \ldots, K$ and C_i in $\{0, \ldots, M-1\}$. Supposing to know the a posteriori probability:

$$P_i(C_i|C_i) = P_i(\text{input} = C_i|\text{output} = C_i),$$

let alone the case that $P_i(C_i|C_i) = 0$ (so we may assume $P_i(C_i|C_i) > 0$), we can therefore define the error probability:

$$e_i = 1 - P_i(C_i|C_i) \in [0, 1) \quad (i = 1, \dots, K),$$

which we suppose being less than 50% (on the contrary, the expert would make more wrong than exact decisions).

Let now e_i and e_j (both < 0. 5) the two errors that the methods E_i and E_j can make in deciding respectively C_i and C_j, and let us assume the case $C_i = C_j$, i.e., both experts responding in the same manner. We therefore define a new combined error probability (the probability that "both" methods are wrong):

$$ec_{ij} = e_i e_j \quad (i, j = 1, \dots, K; i \neq j).$$

Obviously, being e_i and $e_j < 1$, will be $ec_{ij} < e_i, e_j$. In fact:

$$e_i = e_i \cdot 1 > e_i e_j = ec_{ij} \Rightarrow e_i > ec_{ij}$$

and so for e_j, too.

The criterion adopted for the final decision is the minimization of the *per class* error. Each one of the K experts produces an answer C_i, that has its associated e_i. The e_i can be re-associated per class, raising a number of combined errors K', which just depends on the number of different responses (classes) given by the experts, varying from 1 to K. Every class will have associated to it the new error given by the product of the a posteriori errors related to the methods answering with "that" class.

Let therefore $K' \leq K$ the number of different decisions supplied by the K experts; let ec_i the combined error associated to the answer C_i, with $i = 1, \dots, K'$. The final decision r of the combiner is:

$$r = C_j, \quad \text{if } ec_j = \min_{\{i=1, \dots,, K\}} ec_i.$$

In a real situation the values $P_i(C_i|C_i)$ (and therefore e_i and ec_i) are not known, but it is possible to approximate them with the estimated values coming from the preliminary studies on each method, therefore constituting the "training" of the combiner, which will always rely on them in order to calculate the reliability of an expert's answer. Finally, in order to add a refusal mechanism in case of low reliability, it is sufficient to set aside a maximum error threshold α and to make in way not to accept the decisions *r* having an $ec_i > \alpha$. Therefore:

$$r = \begin{cases} C_j, & \text{if } ec_j = \min_{\{i=1, \dots, K\}} ec_i \text{ and } ec_j < \alpha \\ M, & \text{otherwise,} \end{cases}$$

is the complete decision criterion, where $\alpha \in [0, 1]$.

2.5. Majority weighted vote

The experienced hybrid method is little more than a lead test in order to observe how the training of the First Probability Driven could influence the decision taken from the Majority Vote. The used hybrid - called *Majority Weighted Vote* - is in all equal to its parent (from which it takes its name), with the only difference that the score tables PTS_i are further transformed before being employed for the decision. Their internal values are in fact multiplied by the probability $P_i(C_i|C_i)$, where C_i is the digit with the highest score obtained with the *i*-th method.

With the same notations of the Majority Vote:

$$\begin{array}{rcll} V_i & = & R_i(x) & (\text{recognizer } R_1, \dots, R_K : \text{patterns' domain} \to \mathbf{R}^M) \\ PTS_i & = & S_i(V_i) & (\text{score } S_1, \dots, S_K : \mathbf{R}^M \to \mathbf{N}^M) \\ PTS'_i & = & PTS_i \cdot P_i(C_i|C_i) & (\text{weighted scores } PTS'_1, \dots, PTS'_K : \mathbf{N}^M \to \mathbf{R}^M, \text{ with } C_i \text{ such as } PTS_i(C_i) = \max_{\{j=0, \dots, M-1\}} PTS_i(j)) \\ T & = & I(PTS'_1, \dots, PTS'_K) & (\text{combination } I : \mathbf{R}^{M \times K} \to \mathbf{R}^M). \end{array}$$

The probability therefore acts as a "weight", giving a greater importance to the more reliable methods and making their decisions more “influential”. As happens for the Majority Vote, the decision rule in order to accept or to reject the pattern becomes:

$$R(x) = \begin{cases} j, & \text{when } T(j) = \sum_{k=1}^{K} PTS'_k(j) = \max_{\{i=0,\,\dots,\,M-1\}} T(i) > \lambda, \\ M, & \text{otherwise,} \end{cases}$$

where $j \in \{0, \dots, M-1\}$ and λ is an appropriate threshold equal to $W \cdot M \cdot \alpha$ being W the sum of the $P_i(C_i|C_i)$ and used as a weight for $\alpha \in [0, 1]$.

3. Test and Results

The combination schemes have been tried with four classifiers of handwritten digits: *Histograms method*, *Pattern Matching*, *Intersections method*, *Enhanced Modified Loci*; a short description is given in Appendix.

The recognition algorithms have been initially developed in *C* within an Unix environment. The aim in implementing these methods was to obtain efficient algorithms (simple and fast, also in spite of the effectiveness of the single ones) making so that the features extracted by each of these methods were complementary to those extracted by the others, therefore having experts with various “points of view”. The tests have been lead using two sets of 4000 pattern (sample digits), denote with TEST_0 and TEST_1. Each set is composed of 400 patterns in black and white, acquired at 150 dpi, for each of the ten digits. The low quality of some pattern has negatively conditioned the performances of the recognizers, in particular set TEST_1 contains various ambiguous exemplaries of which some example is brought back in Figure 1:

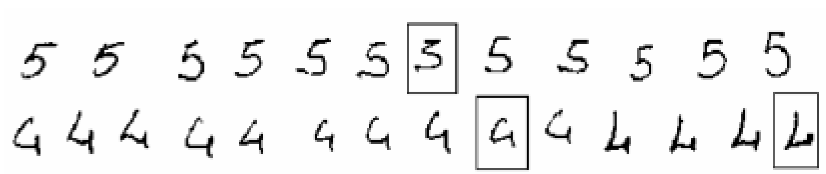

Figure 1. Some ambiguous patterns of the set TEST_1 (box dimension: 50×50 pixels).

The 8,000 patterns used are part of a much larger (~ 30,000) set. The authors of the patterns are for about 90% young students of age comprised between 20 and 25, of both sexes, asked to write at least ten samples for every decimal digit.

In the following tables four columns of percentages are illustrated. The percentage of acknowledgment has been calculated like:

$$\frac{number\ \ of\ \ recognized\ \ patterns}{number\ \ of\ \ exa\min ed\ \ patterns}$$

The percentages of refusal and error have been analogously calculated replacing the numerator with the number of rejected and erroneously recognized patterns, respectively. Aside the combiners a value between parenthesis is reported representing the threshold α, except for the First Probability Driven, for which it instead represents the accepted maximum error.

Recognitions-integration comparison tables

From the examination of Table 1 we can notice that the increment of the number of the used classifiers causes a sudden lowering of the substitution error percentage, therefore increasing the total reliability of the final decision.

Table 1. Use of the combination to increase the total reliability
H = Histograms EL = Enhanced Loci PM = Pattern Matching I = Intersections

TEST_0	4,000 patterns acceptable quality	Recognitions	Refusals	Erros	Reliability
Majority Vote (0.9) applied to:	H	80.02%	5.38%	14.60%	84.63%
	H + EL	79.36%	15.42%	5.22%	93.83%
	H + EL + PM	79.90%	15.90%	4.20%	95.00%
	H + EL + PM + I	85.95%	11.45%	2.60%	97.06%

Table 2. Combination schemes applied to the four experts (TEST_0 patterns) H = Histograms EL = Enhanced Loci PM = Pattern Matching I = Intersections MV = Major Vote B = Bayes 1PD = Ist Probab. Driven MWV = Major Weight Vote

TEST_0	4,000 patterns acceptable quality	Recognitions	Refusals	Erros	Reliability
Recognizers:	H	80.02%	5.38%	14.60%	84.63%
	PM	76.85%	9.73%	14.42%	84.20%
	I	74.25%	12.47%	13.28%	84.83%
	EL	76.30%	0.04%	23.66%	76.33%
Combination schemes:	MV (0.9)	85.95%	11.45%	2.60%	97.06%
	B (0.125)	85.00%	6.53%	8.47%	90.94%
	1PD (0.1)	91.00%	2.55%	6.45%	93.38%
	MWV (0.9)	85.90%	11.98%	2.12%	97.59%

Table 3. Combination schemes applied to the four experts (TEST_1 patterns) H = Histograms EL = Enhanced Loci PM = Pattern Matching I = Intersections MV = Major Vote B = Bayes 1PD = 1st Probab. Driven MWV = Major Weight Vote

TEST_1	4,000 patterns ambiguous quality	Recognitions	Refusals	Erros	Reliability
Recognizers:	H	78.90%	2.48%	18.62%	80.91%
	PM	72.37%	12.65%	14.98%	82.05%
	I	78.27%	0.98%	20.75%	79.04%
	EL	70.40%	4.40%	25.20%	73.64%
Combination schemes:	MV (0.9)	80.42%	16.03%	3.55%	95.77%
	B (0.125)	80.40%	8.85%	10.75%	88.21%
	1PD (0.05)	82.80%	11.35%	5.85%	93.40%
	MWV (0.875)	83.63%	12.80%	3.57%	95.91%

In Tables 2 and 3 are visible the improvements that the combinations bring to the percentages, both of acknowledgment and error.

For reliability we mean the "ability not to mistake", that is computable as:

$$1 - (\% \text{ of } \textit{substitution}).$$

Here we prefer to refer ourselves to the definition of reliability reported in [21], i.e.,

$$\textit{reliability} = \frac{\% \text{ of } \textit{acknowledge}}{1 - (\% \text{ of } \textit{refusal})}. \tag{7}$$

The fact that highest levels of acknowledgment are not reached is due to the however present error (that all-or nearly all-the recognizers do) when they must decide on a very ambiguous pattern: in other words, if all the single recognizers take a wrong decision, it cannot be expected that the combination of these decisions is right.

For completeness, let us observe that there are classifiers [4], [21] that-even if operating alone-are able to reduce the error percentage until 1%.

It was not an objective of the present work to develop such complex and powerful algorithms, therefore it has not been possible to try them with the already available 150-dpi patterns for being able to draw useful information from a comparison. We hope to be able, in future, to verify the combiners/classifiers with the same patterns used in [21].

4. Conclusions

This paper attempts to explore the heuristic that the probability that two classifiers make the same error is unlikely, assuming that the classifier outputs are independent.

The methods so far considered of handwritten digits recognition taken alone do not reach elevated performances, but they however satisfy the speed requirement, and their combined use in several combinations (in the integration phase) can increase the aforesaid performances allowing to reach a good reliability.

Among the combination techniques the *Majority Vote* turned out better, albeit the simplest. An improvement of the *Bayes* technique is surely possible, adopting a normalization method different from that one used here. *The First Probability Driven* has been used with a threshold α around 90% but we think that, raising it, it is possible to reduce the total error until reaching a reliability percentage near that of the Majority Vote. The *Majority Weighted Vote* always turned out near its predecessor; we foretold that the additional information given by the reliability estimates could very positively condition the final acknowledgment, which did not happen. We however found a greater solidity of the Majority Weighted Vote vs. the Majority Vote when we operated with patterns of slightly lower quality (TEST_1).

Further studies to carry out for future improvements concern:

● an analysis of the answers of the combiners and the recognizers, on varying of the parameters α and β;

● the application of the Information Theory to the acknowledgment methods, that can be seen like channels transporting a message; the pattern can be seen like a message written in a language L, which can be coded with a value from 0 to $M - 1$, (where M is the number of the classes) and is the input to the classifier-channel. The output (tables) is also the output of the channel. The channel matrix can be computed from the confusion tables already available, and the transmission error can be studied, let alone trying to minimize it;

● a revisiting of the combiners to take account also of the possible refusals (and therefore nondecisions taken from the experts), in order to observe if that leads to a lowering of the ambiguity in the decision;

● an improvement of the single recognition methods (obviously possible with an improvement of the various databases, considering the hypothesis of the insertion of new patterns), passing from the actual 400 (40 patterns for each digit) to at least 200 for each digit, even if that obviously involves a lowering of the execution speed. We are also examining the alternative posibilities like the recreation of the databases with criteria different from those up to now adopted (not set of features, but mean of features). Some first results on few samples indicate that, to a light lowering of the percentage of acknowledgment (which can be superseded, only due to the shortage of trainers in the database) corresponds a remarkable increment of the execution speed.

References

[1] G. Baptista and K. M. Kulkarni, A high accuracy algorithm for recognition of handwritten numerals, Pattern Recognition 21(4) (1988).

[2] Gavin Brown, Jeremy L. Wyatt, Rachel Harris and Xin Yao: Diversity creation methods: a survey and categorization, Information Fusion (INFFUS) 6(1) (2005), 5-20.

[3] A. d'Acierno, C. De Stefano and M. Vento, A structural character recognition method using neural networks, ICDAR-91, First International Conference on Document Analysis and Recognition, 30 Sep.-2 Oct. 1991, Saint Malo, France 2 (1991), 803-811.

[4] P. Dargenton, N. Vincent and H. Emptoz, Appariement de deux graphes structurels quelconques pour la reconnaissance de lettres manuscrites, Actes du 9e Congrès AFCET-RFIA, Paris, (1994), 461-471.

[5] V. Di Lecce, A. Guerriero, G. Dimauro, S. Impedovo, S. Pirlo and A. Salzo, A New Database of Confusing Characters for Testing Character Recognition Algorithms, 10th International Conference on Image Analysis and Processing (ICIAP'99) 939 (1999).

[6] A. C. Downton, R. W. S. Tregidgo, C. G. Leedham and Hendrawan, Recognition of handwritten British postal addresses, From Pixel to Features III: Frontiers in Handwriting Recognition, S. Impedovo and JC Simon (eds), Elsevier Science Publishers BV (1992), 129-143.

[7] Giorgio Fumera and Fabio Roli: A Theoretical and Experimental Analysis of Linear Combiners for Multiple Classifier Systems, IEEE Trans. Pattern Anal. Mach. Intell. (PAMI) 27(6) (2005), 942-956.

[8] Y. S. Huang, and C. Y. Suen, An Optimal Method of Combining Multiple Classifiers for Unconstrained Handwritten Numeral Recognition From Pixel to Features III: Frontiers in Handwriting Recognition, S. Impedovo and IC Simon (eds), Elsevier Science Publishers BV (1992).

[9] J. J. Hull, A. Commise and T. K. Ho, Multiple Algorithms for Handwritten Character Recognition, Frontiers in Handwriting Recognition, April 2-3, Montreal PQ, Canada (1990).

[10] Ludmila I. Kuncheva, Diversity in multiple classifier systems, Information Fusion (INFFUS) 6(1) (2005), 3-4.

[11] Ludmila I. Kuncheva, Christopher J. Whitaker, and Catherine A. Shipp, Limits on the majority vote accuracy in classifier fusion, Pattern Anal. Appl. 6(1) (2003), 22-31.

[12] D. S. Lee and S. N. Srihari, Handprinted Digit Recognition: A Comparison of Algorithms, From Pixel to Features III: Frontiers in Handwriting Recognition, S. Impedovo and IC Simon (eds), Elsevier Science Publishers BV (1992).

[13] Ofer Matan: On Voting Ensembles of Classifiers, In Proceedings of AAAI-96 workshop on Integrating Multiple Learned Models (1996).

[14] N. J. Naccache and R. Shingal, SPTA: A proposed algorithm for thinning binary patterns, IEEE Trans. on System Man and Cybernetics, SMC-14(3) (1984).

[15] T. Pamg-Ning and J. Rong, Research track posters: Ordering patterns by combining opinions from multiple sources, Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining KDD'04, ACM Press (2004).

[16] From Pixel to Features III: Frontiers in Handwriting Recognition, S. Impedovo and J. C. Simon (eds), Elsevier Science Publishers BV (1992).

[17] Konstantinos Sirlantzis, Sanaul Hoque and Michael C. Fairhurst, Input space transformations for multi-classifier systems Based on n-tuple classifiers with application to handwriting recognition, Multiple Classifier Systems (2003), 356-365.

[18] Kagan Tumer and Joydeep Ghosh, Error Correlation and error reduction in ensemble classifiers, Connect. Sci. 8(3) (1996), 385-404.

[19] P. S. P. Wang and A. Gupta, An Improved Structural Approach for Automated Recognition of Handprinted Characters. Proceedings in Fundamentals in Handwriting Recognition, Chateau de Bonas, France (1993).

[20] P. S. P. Wang, M. V. Nagendraprosad and A. Gupta, A neural net based 'hibrid' approach to handwritten numeral recognition, From Pixel to Features III: Frontiers in Handwriting Recognition, S. Impedovo and IC Simon (eds), Elsevier Science Publishers BV (1992).

[21] L. Xu, A. Kryzak, and C. Y. Suen, Methods of combining multiple classifiers and their applications to handwriting recognition, IEEE Transactions on Systems Man and Cybernetics 22(3) (1992).

[22] Manuela Zanda, Gavin Brown, Giorgio Fumera and Fabio Roli, Ensemble learning in linearly combined classifiers via negative correlation, MCS (2007), 440-449.

■