

**NEURAL NETWORK COOPERATION
FOR HANDWRITTEN DIGIT RECOGNITION:
A COMPARISON OF FOUR METHODS**

Y. AUTRET (2), A. THEPAUT (1),

(1) TELECOM Bretagne, Technopole Brest Iroise
29285 Brest Cedex - FRANCE.

(2) Université de Bretagne Occidentale, Laboratoire LIM1
BP 809, 29285 Brest - FRANCE

ABSTRACT

In this paper we focus on "off-line digit recognition" with unknown scriptor. After presenting two neural recognisers, we evaluate four solutions to combine results obtained from the two systems.

The tests were performed by the French postal services (SRTP) on their secret data base containing more than 7000 digits taken from everyday mail. This allows us to evaluate our four cooperation methods and also to compare them to other methods developed by other research teams.

1. INTRODUCTION

It is well known that it may be useful to combine the results of several neural networks to improve classification performance. In this paper we propose several solutions based on the integration of two simple hybrid neural systems. Each network is fed with morphological features and takes its own decision. The final decision is taken by integrating individual results.

The first network is called PNN (Pixel Neural Network). It is a (576x100x10) fully connected Perceptron which receives nine (8x8) images. The first image is the raw image scaled to a (8x8) image. The other eight (8x8) images contain properties of the background of the raw image. Information indicating whether it is possible to direct a beam from a given point to a given direction without hitting the contour digit

is stored in the eight images [AUT 95]. Eight general directions were defined (North, North-East, East, South-East, South, South-West, West and North-West) and we obtained eight new images containing information of eight kinds of cavities.

The second network is called CNN (Contour Neural Network). It is a (256x70x10) fully connected Perceptron which receives contours. Contours are normalised and reduced to 256-value vectors [AUT 95].

These networks were individually tested by the French postal services on their secret base. Rates of 96.9% (PNN) and 93.5% (CNN) were obtained.

In the next section we will show that the accuracy of these results can be increased by cooperation and even by a very simple mechanism of cooperation. An overview of the cooperation mechanism is shown in figure 1. PNN and CNN are independent networks and produce their own solutions. PNN and CNN output vectors are merged in the last stage and four merging strategies will be discussed in the next section. The CNN system could have been replaced by a better system but CNN is very different from PNN and helps reduce the correlation between the two systems.

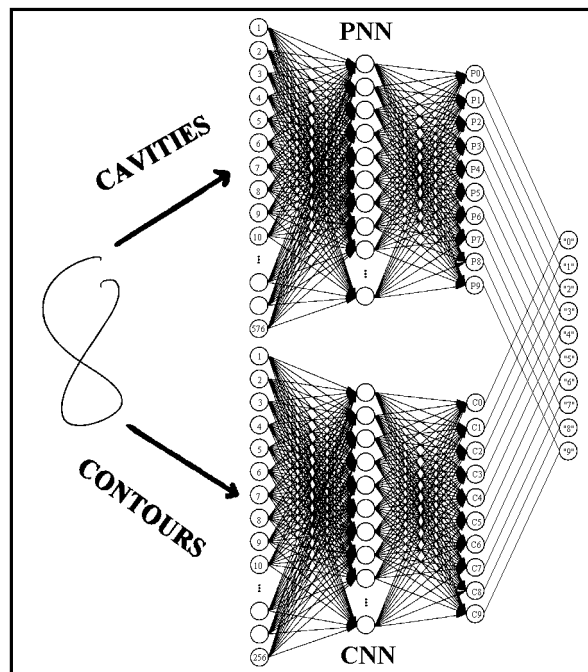


Figure 1: System overview

2. FOUR MERGING STRATEGIES

2.1 No weighting association

In this kind of cooperation, we do not focus on individual performances of neural networks. Each result is applied in the same manner. There are several strategies which belong to this class of cooperation. For example, a method proposed by [BAT

94] which consists of computing a mean output vector. Another way to cooperate may be to add or to multiply output vectors of the two neural nets.

Rates of 97.13 % and 96.26 % are respectively obtained by adding and multiplying output vectors. Despite its great simplicity, the addition of output vector produces accurate results. This fact has already been discussed by several authors [MON 95].

In this category of association of neural net we find voting methods. They consist of selecting the class which obtains the best score. For example, the *Borda count method* is a generalisation of voting methods. In this algorithm, we have to count the sum of class numbers which are in a given class. For example, if both PNN and CNN neural nets recognise the digit '3', then the value of *Borda count* will be : $1 + 1 = 2$. The selected class is the class which obtains minor value after voting.

formula : $For\ i = 0..9$

$$Borda[i] = range(PNN,i) + range(CNN,i)$$

$range()$ is a function which returns the range of the neuron « i » by the specified net.

A rate of 95.8 % is obtained by the Borda count method on secret data base. The implementation of this method is quite simple and does not require any *a priori* knowledge about classifiers. Each classifier is treated in the same manner, but we know that the CNN network exhibits lower performance than the PNN network.

2.2 Weighting association using fuzzy logic

In this kind of association, the main idea consists of computing, for each net before applying a combination rule, a coefficient which reflects its performance in regard to the others. Various strategies using weighting association were already proposed [SUG 77] [CHO 95].

We first implement a method for combining PNN and CNN neural networks based on fuzzy logic, especially the fuzzy integral [CHO 95a]. This non-linear method combines a network output according to the importance of the individual neural networks [CHO 95b]. We compute the degree of importance, named fuzzy densities (g^i), of each network in regard to the others (equ. 1).

$$g^i = \frac{p_i}{\sum_j p_j} \cdot dsum \quad (\text{equ. 1})$$

In equation (7) p_i is the performance of network i and $dsum$ is the desired sum of fuzzy densities. For each class ($i \in [0,9]$), we compute $g(A_i)$ as follows (equ. 2):

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}) \quad (\text{equ. 2})$$

where λ is obtained with (equ 3).

$$\lambda + 1 = \prod_{i=1}^n (1 + \lambda g^i) \quad \text{with } n \text{ the number of neural networks} \quad (\text{equ. 3})$$

For each class, we compute a fuzzy integral e (equ. 4)

$$e = \max_{i=1}^n [\min(h(y_i), g(A_i))] \quad (\text{equ. 4})$$

The class exhibiting the largest fuzzy integral e is chosen as the output class. Fuzzy integral method has been evaluated on secret base and has produced a rate of 97.1 %. This good result shows how important it is to look at the relative importance of PNN and CNN neural network.

2.3 Weighting association using Bayesian probabilities

We have also implemented a coupling method based on Bayesian probabilities. As described in [DEN 91], we transform neural-net output level to probability distributions. The confusion matrix, obtained on a public data base, is used to know the error rate of each net by class and to obtain conditional probabilities. Neural Network compute *a posteriori* Bayesian probabilities.

If $(X = x_1, x_2, \dots, x_T)$ is an unknown form and $(\Omega = \omega_1, \omega_2, \dots, \omega_T)$ a set of classes, then every neuron output estimates the probability $P(\omega_i / X)$ of appertenance to this class (equ. 5).

$$P(\omega_i / X) = \frac{P(X / \omega_i)P(\omega_i)}{P(\omega_i / X)} \approx f \left\{ \sum_{k=1}^H w_{ik} f \left(\sum_{j=1}^T w_{kj} x_j \right) \right\} \quad (\text{equ. 5})$$

where f is the sigmoid function, w_{kj} the weight between the input neuron and the neuron k which is in the hidden level, and w_{ik} is the weight between neuron k and the output neuron i . Then it is possible to compute *a posteriori* probabilities (equ. 6).

$$P(\omega_i / X) = \frac{P_{PNN}(\omega_i / X) + P_{CNN}(\omega_i / X)}{2} \quad (\text{equ. 6})$$

As in the previous method, the largest $P(\omega_i / X)$ value is chosen as the output class. This method produces a recognition rate of 96.24%.

2.4 Multi-stage coupling

Several authors have proposed to build systems composed of groups of classifiers. For example, Kimura [KIM 91] presented a complete study of different ways to combine two classifiers (in parallel and in series). As in [HO 94], we applied a combination rule to each group. The decisions of each group were then combined to form the final decision. We have chosen three different cooperation methods already presented: PNN and CNN vectors by addition and fuzzy logic method to obtain two new vectors which are then combined by the Borda count method. A rate of 97.24 % is obtained on our digit recognition problem.

3. COMPARISON OF THE RESULTS

In this section, we compare the results of different methods. They have all been tested on the same secret data base of the SRTP which contains 7388 digits. Results are shown in table 1.

Pottier et al. [POT 93]		95,1 %
Lemarié et al. [LEM 93]	RBF	96.1%
Schwenk et al. [SCH 93]	1-PPV	97.8%
Schwenk et al. [SCH 93]	DIAB EUCL	96.0%
Schwenk et al. [SCH 93]	DIAB BILAT	97.3%
Autret & Thépaut	PNN	96.9%
Autret & Thépaut	CNN	93.5%
Autret & Thépaut	PNN+CNN	97.1%
Autret & Thépaut	Fuzzy Logic	97.1%
Autret & Thépaut	Multi stage	97.2%

Table 1

Table 1 shows that the 1-PPV method proposed by Schwenk *et al.* [SCH 96] appears to be the best method (June 1996). But the authors specified that this performance, which is close to the performance of a human being, is obtained by a very complex method which requires two seconds per digit on a HP 715/50 workstation. A very fast method such as PNN+CNN produces similar results. So we see that a very simple coupling method like the addition of two output vectors of two independent neural networks (PNN+CNN) produces results that are not far from more complex methods such as our multi-stage coupling or the diabolo classifier of Schwenk *et al.*

4. CONCLUSION

Many works have shown that if individual classifiers are correctly optimised, combining large number of classifiers does not really improve recognition rates. Furthermore, the performances of these complex systems are often lower than the performances of a system which combines two or three nets. The biggest difficulty in this approach is to find models which produce decorelated errors. When such models are obtained (PNN and CNN for example), we have shown that the use of a simple coupling rule can provide as good a result as those provided by a more complex coupling rule.

ACKNOWLEDGEMENTS

The authors would like to thank B. Lemarie and M. Gilloux of SRTP/ Nantes (Post Office Research Center) for providing the public database of French zip codes and for performing the tests on the secret data base. They also wish to thank C. Ranaivo for his contribution to this work.

REFERENCES

- [AUT 95] AUTRET Y., THEPAUT A., "*Two simple cooperating neural systems for efficient handwritten digit recognition*" ICNN'95, pp. 2175-2178, Perth, Western Australia, dec.1995.
- [BAT 94] BATTITI R., COLLA A.M., "*Democracy in Neural Nets: Voting Schemes for Classification*" Neural Networks, vol. 7, n° 4, pp. 691-707, 1994.
- [BLA 63] BLACK D., "*The theory of committees and Elections*", 2nd ed. London: Cambridge University Press, 1958,1963.
- [CHO 95a] CHO.S.B. , KIM J.H., "*Multiple Network Fusion using Fuzzy Logic*", IEEE Transactions on Neural Networks, Vol. 6, N° 2, march 1995.
- [CHO 95b] CHO.S.B., KIM J.H., "*Combining multiple neural networks by fuzzy integral for robust classification*", IEEE Trans. Syst.,Man, Vol. 25, N° 1, 1995.
- [DEN 91] DENKER J.S., LE CUN Y., "*Transforming neural-net output levels to probability distributions*", In R.P. Lippmann, J. Moody, D.S. Touretzky (Eds.), Advances in Neural information processing systems, NIPS 3, pp. 853-859, San Mateo, Morgan Kaufmann.
- [HO 94] HO T. et al. "*Decision Combination in Multiple Classifiers Systems*", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 16, n°1, 1994.
- [KIM 91] KIMURA F., SHRIDHAR M., "*Handwritten numerical recognition based on multiple algorithms*", Pattern Recognition, 24, 10, pp. 969-983, 1991.
- [LEM 93] LEMARIE B. "*Réseaux de Régularisation pour la Reconnaissance de chiffres manuscrits*", JETPOSTE'93, pp. 541-549, Service de Recherche Technique de la Poste.
- [MON 95] MONTOLIU L., "*Architecture multi-agent et réseaux connexionnistes*", Thèse de doctorat de l'école polytechnique, 27 septembre 1995.
- [POT 93] POTTIER I., BUREL G., "*Evaluation of a Neural System for Handwritten Digit Recognition*", JETPOSTE'93, pp. 541-549, Service de Recherche Technique de la Poste.
- [SCH 96] SCHWENK H., MILGRAM M., "*Reconnaissance des codes postaux par Réseaux Diabolo*", CNED'96, 4ème Colloque National sur l'Ecrit et le Document, Nantes, juillet 1996.
- [SUG 77] SUGENO M., "*Fuzzy Measures and Fuzzy Integrals: a survey*", Fuzzy Automata and decisions processes, North Holland, pp. 89-102, 1977.
- [TIN 94] TIN KAM HO , "*Decision Combination in Multiple Classifier Systems*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 16, N° 1, January 1994.