A deterministic biologically plausible classifier

Thierry Viéville and Sylvie Crahay

INRIA, BP 93, INRIA, 06902 Sophia, France

Abstract

Regarding biological visual classification, recent series of experiments have enlighten that data classification can be realized in the human visual cortex with latencies of about 100 ms, which, considering the visual pathways latencies, is only compatible with a very specific processing architecture, described by the so-called Thorpe model.

Surprisingly enough, this experimental evidence is in coherence with algorithms derived from the statistical learning theory, following the work of Vapnik. More precisely, there is a double link: on one hand, the Vapnik theory offers tools to evaluate and analyze the Thorpe model performances and on the other hand, this model is an interesting front-end for algorithms derived from the Vapnik theory.

The present contribution develops this idea and experiments its performances using a tiny sign language recognition experiment.

Key words: Neuronal classifier, Supervised learning, Vapnik dimension, Biological model

Biological visual classification is a well-known and very common, but still intriguing fact. Here data classification simply means being able to put a *unique label* on a given data input (e.g. "oh, there is a dog").

Recent series of experiments have enlighten this biological mechanism: data classification can be realized in the human visual cortex with latencies of about 100 ms [10] which, considering the visual pathways latencies [8], may only be compatible with a very specific processing architecture and mechanism, the so-called *Thorpe model*. Even "high level" visual data classification such as face recognition [3] as pointed out recently by [9] can be realized at such a very fast rate.

As carrefully discussed in [13], the selection mechanism of the data category may be viewed as a *simple nearest-neighbor mechanism*: i.e. for a data, *the maximal proximity to the prototypes* is selected, the data category being the proximal prototype's category.

Regarding fast data classification, the Thorpe model [9] is based on the fact that very short observed latencies are only compatible with an information flow much more related to the *occurrence* of a neuronal signal than to e.g. the spike frequency; since the latency of the first spike of a given neuron is a direct decreasing function of the neuron input value, only neurons with the *highest values* generate spikes fast enough to be taken into account. As a consequence:
Quantification of the neuronal information: the quantitative value is thus directly related to the spike delay and is thus a bounded value (with *min* and *max* values) with a *finite precision*.
Sparseness of the neuronal information: among the rather huge number of input neurons (typically the dimension n of the neuronal "vector" has an order of magnitude of 10^5) only a rather small number ($\simeq 10^3$) is taken into account.

Training capability and learning performances

The Vapnik learning theory [12] allows to formalize the idea that efficient models have a limited complexity. As such, it is a formalization and in fact an improvement of the well-known Occam's Razor principle.

Let us review this piece of theory . For a given classifier C, it relates:

- the *expected risk* $R(C)$ (i.e. the "average" probability for the classifier to provide a wrong answer) for a set of inputs, randomly chosen according to an unknown probability distribution with
- the *empirical risk* $R^M_{emp}(C)$ (i.e. the "average" probability for the classifier to provide a wrong answer) for the calibration set of size M;
  this quantity being equal to zero for a deterministic classifier, as discussedpreviously.

More precisely, for a chosen probability δ, the *expected risk* can be bounded with a probability at least $1-\delta$ as follows:

$$R(C) \leq \underbrace{R^M_{emp}(C) + \underbrace{\epsilon(M,\delta)}_{\text{bias}}}_{\text{"guaranteed risk"}} \tag{1}$$

where the *bias* (also called confidence bound) $\epsilon(M,\delta)$ is a function of the chosen probability δ, the calibration set size M and the *class of the classifiers*, i.e. the set of classifiers which is used during the learning phase.

Indeed, we expect this bias to decrease with the calibration set size M. The

learning mechanism is *consistent* iff $\lim_{M\to\infty}\epsilon(M,\delta)=0$. Better than that, if the classifier functions are bounded, at the convergence, the smallest/optimal value of the expected risk [12] is obtained.

It appears that if the classifiers class is too large, the process is *not* consistent: with a very large class of classifiers, we can classify everything, but what does everything does anything. Here, inconsistency means that the bound does not decrease to zero when M increases.

This is clearly a problem for neuronal networks used as classifiers, because their V_c dimension is higher than the order of magnitude of the number of neurons. More precisely [1], for an arbitrary feedforward neuronal network with a binary activation function the V_c dimension is of $o(W\ \log(W))$ where W is the number of weigths free parameters in the network, while [6] for a multilayer feedforward neuronal network with a sigmoid activation function, the V_c dimension is of $o(W^2)$.

More precisely, this bias is related in the Vapnik theory to the "margin" between prototypes and the "number of degree of freedom" of the model, these complex notions being formalized in the V_c dimension. In the case of the Thorpe model this correspond to a small V_c of about 10^3 for the brain (less than 10^2 for the SpikeNet simulator) which is an impressive result and is a theoretical justificatiion of the observed performances.

1 Experimental results

Interactive 2D demonstration

In order to validate the previous method, a statistically optimal version of the Spikenet Thorpe model [2] has been implemented as detailed in [13] using an Hebbian learning rule as in usual biologically plausible learning mechanisms. Please refer to the on-line [1] software documentation for details.

Sign-language recognition

We consider a tiny experiment related to the recognition of the Quebecian Sign Language Alphabet. This can not be considered as a real experiment of sign-language recognition, whereas it is only used to evaluate the present method.

[1] In `http://www-sop.inria.fr/odyssee/imp` the `imp.math.Classifier` classes.

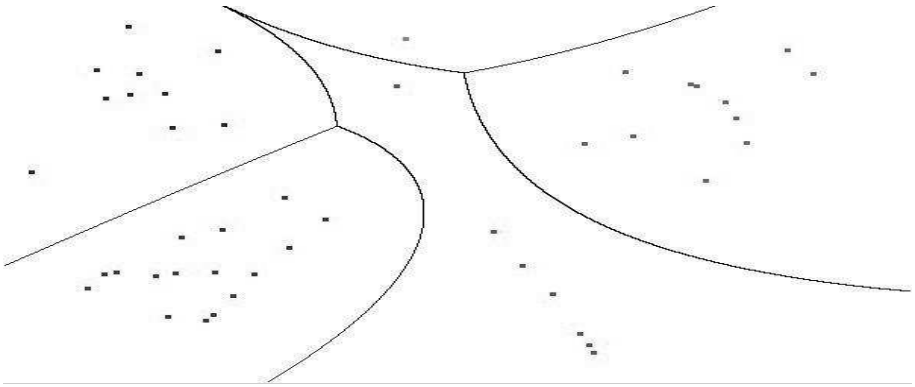

Fig. 1. An example of result : the classifier is optimized with a second order polynomial model, with 5 prototypes: $Y_g = 28$. In this case, we obtain a solution with 5 prototypes regardless the fact that prototypes have been edited or not, before the optimization. With this data set, second order polynomial model yields an optimal Y_g dimension.

The static (one image) spelling of four subjects have been recorded using a standard video system with a resolution of 384×288, as follows:

Subject		data
Sy	Experimented	2 series of 9 letters (1 particularly good)
Th	Beginner	3 series of 9 letters, 2 without shadow, 1 with hand shadow

examples of such images being given in Fig. 2.

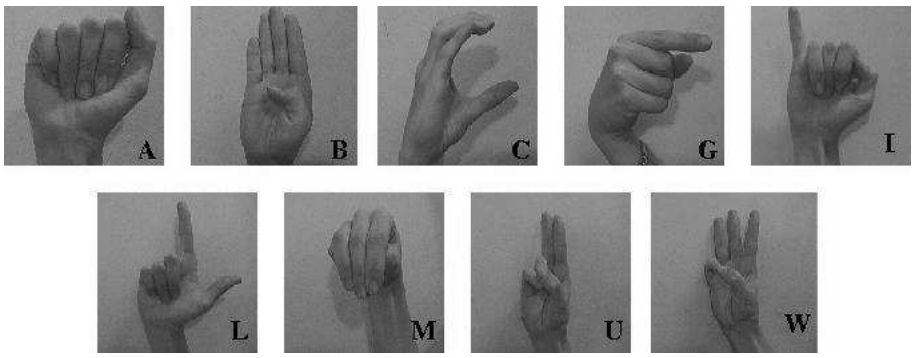

Fig. 2. An example of the nine letters taken into account in this experiment, here sub-images containing containing the hand have been automatically cropped as discussed in the text.

The following experimental configurations have been chosen in order to evaluate the method with respect to combinations of data:

Experiment label	learning set	test set
Experiment 1	Sy, 9 letters from all samples except 1 series	Sy, 9 letters from the last sample
Experiment 11	Th, 9 letters from all samples except 1 series	Th, 9 letters from the last sample
Experiment 2	Sy, all samples of 9 letters	Th, 3 series of 9 letters

the learning and test sets intersection being always empty.

In order to extract relevant parameters, using standard image analysis methods, edges are detected and binarized using a fixed threshold. This simple paradigm is sufficient, with the lighting conditions of the data acquisition, to pre-process the image. Signs are compared using edge maps, as detailled in [13]

In order to compare the performances of the present method with a well established mechanism, we use a standard 1-to-1 SVM method, as discussed elsewhere [13] .

A subset of the obtained results is summarized in the following table:

Experiment label	Raw classifier	standard 1-to-1 SVM	optimized NN classifier
Experiment 1	11 %	11 %	11 %
Experiment 11	30 %	22 %	27 %
Experiment 2	37 %	33 %	28 %

where the percentage of errors have been reported. This clearly demonstrates, that up to the 1st order, both methods have similar performances. This was not entirely obvious since we have implemented the optimization using a biologically plausible minimization method. A step further, we clearly observed that the deterministic method have better performances when the calibration data set has no mistakes, while performances are degraded when it contains errors.

One important fact here is that Hebbian-like optimization does not induce huge computation times, but only a little overhead.

2 Conclusion

We also have made explicit and experimented that SVM like mechanisms can easily be implemented using Hebbian-like correction rules. Such optimization mechanism is not as fast as the standard method, but its biological plausibility is much better, while final performances are similar.

This point of view is in deep relation with fast visual recognition in the brain. It may, for instance, explain why biological classifiers have better generalization performances, although the input data has a huge dimension.

More precisely, it has been possible to estimate a bound of the V_c dimension of the Thorpe model, showing that its generalization performances are expected

to be much better than standard models based on neuronal networks.

Please refer to [13] for a development.

Acknowledgments: *Simon Thorpe*, *Yann Guermeur* and *Olivier Faugeras* are gratefully acknowledged for some powerful ideas at the origin of this work.
This work has been realized within the scope of the French Amaria/Robea project.

References

[1] E.B. Baum and D. Haussler. What size net gives valid generalization. *Neural Computation*, 1:151–160, 1989.

[2] A. Delorme, J. Gautrais, R. VanRullen, and S. J. Thorpe. Spikenet: A simulator for modeling large networks of integrate and fire neurons. *Neurocomputing*, 26:989–996, 1999.

[3] A. Delorme and S. Thorpe. Face processing using one spike per neuron: resistance to image degradation. *Neural Networks*, 14:795–804, 2001.

[4] L. Devroye, L. Gyorfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer Verlag, 1996.

[5] R. Durbin, C. Miall, and G. Mitchinson, editors. *The computing neuron*. Addison-Wesley, 1989.

[6] P. Koiran and E.D. Sontag. Neural networks with quadratic vc dimension. *Advances in Neural Information Processing System*, 8:197–203, 1996.

[7] D. Marr. *Vision*. W.H. Freeman and Co., 1982.

[8] L.G. Novak and J. Bullier. *The Timing of Information Transfer in the Visual System*, volume 12 of *Cerebral Cortex*, chapter 5, pages 205–241. Plenum Press, New York, 1997.

[9] S. Thorpe, A. Delorme, and R. VanRullen. Spike based strategies for rapid processing. *Neural Networks*, 14:715–726, 2001.

[10] S. Thorpe, D. Fize, and C. Marlot. Speed of processing in the human visual system. *Nature*, 381:520–522, 1996.

[11] S.J. Thorpe and M. Fabre-Thorpe. Seeking categories in the brain. *Science*, 291:260–263, 2001.

[12] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[13] T. Viéville and S. Crahay. A deterministic biologically plausible classifier. Technical Report 4489, INRIA, June 2002.