How to maximize Neural Complexity

Lorenzo Zambotti (Paris 6)
(joint with J. Buzzi)

Kochi, 5th december 2011

- G. Edelman, O. Sporns and G. Tononi [PNAS 1994] have proposed a definition of complexity for neural networks. This concept can be interpreted as a functional on probability laws on a finite space
- A complex (random) system should display a combination of high *differentiation* (local independence) and high *integration* (global correlation).
- The aim of our work is to explore this concept mathematically and in particular to explain properties of random systems with high neural complexity.
- We study the order of magnitude of maximal neural complexity for fixed system size and the properties of maximizers as the system size grows to infinity.

Entropy

Let X be a E-valued r.v. with E finite. The entropy of X is

$$\mathrm{H}(X) := -\sum_{x\in E} P_X(x)\,\log(P_X(x)), \qquad P_X(x) := \mathbb{P}(X = x),$$

where we adopt the convention

$$0 \cdot \infty = 0.$$

The entropy is a measure of the randomness of X. We recall that

$$0 \le \mathrm{H}(X) \le \log|E|,$$

with $\mathrm{H}(X) = 0$ iff X is constant and $\mathrm{H}(X) = \log|E|$ iff X is uniform.

Mutual Information

Given a couple (X, Y) we have

$$\max\{H(X), H(Y)\} \leq H(X, Y) \leq H(X) + H(Y)$$

and

1. $H(X, Y) = H(X)$ iff Y is a function of X
2. $H(X, Y) = H(X) + H(Y)$ iff (X, Y) is independent.

Then we define the Mutual Information of (X, Y)

$$MI(X, Y) := H(X) + H(Y) - H(X, Y) \geq 0.$$

MI is a measure of the dependence between (X, Y), more precisely of the randomness shared by the couple.

Neural complexity

Edelman-Sporns-Tononi [PNAS 1994] consider a finite system of $N = |I|$ r.v. $X = (X_i)_{i \in I}$ with $X_i \in \{0,1\}$ and define the neural complexity as

$$\sum_{k=0}^{N} \frac{1}{\binom{N}{k}} \sum_{S \subset I, \, |S|=k} \mathrm{MI}(X_S, X_{S^c}),$$

where

$$X_S := (X_i, i \in S), \qquad X_{S^c} := (X_i, i \in S^c).$$

By convention, $\mathrm{MI}\,(X_\emptyset, X_I) = \mathrm{MI}\,(X_I, X_\emptyset) = 0$.

The neural complexity of X is zero whenever

1. X is an independent family (*chaos*)
2. X is a deterministic family (*order*).

Neural complexity

We adopt rather the following definition

$$\mathcal{I}(X) := \frac{1}{N+1} \sum_{k=0}^{N} \frac{1}{\binom{N}{k}} \sum_{S \subset I,\, |S|=k} \mathrm{MI}(X_S, X_{S^c}).$$

Then

1. $\mathcal{I}(X)$ is invariant under permutations of $(X_i)_{i \in I}$
2. $\mathcal{I}$ is weakly additive, i.e. $\mathcal{I}(X, Y) = \mathcal{I}(X) + \mathcal{I}(Y)$ whenever X and Y are independent

Maximal neural complexity

It is easy to find systems with minimal (null) neural complexity. But what about systems with *maximal* neural complexity? This is harder. Let

$$\mathcal{I}_N := \sup\{\mathcal{I}(X) : \ X = (X_i)_{i\in I}, \, |I| = N\}.$$

By super-additivity we find $\displaystyle\lim_{N\to\infty} \frac{\mathcal{I}_N}{N} = \sup_{N\geq 1} \frac{\mathcal{I}_N}{N}$.

What is this limit?

We define

1. $X = (X_1, \ldots, X_N)$ is a maximizer if $\mathcal{I}(X) = \mathcal{I}_N$
2. $(X^N)_N$ is an approximate maximizer if

$$\lim_{N\to\infty} \frac{\mathcal{I}(X^N)}{N} = \lim_{N\to\infty} \frac{\mathcal{I}_N}{N}.$$

What do maximizers and approximate maximizers look like?

Maximizers

We can characterize maximizers only for $N = 2, 3$, since in this case it is possible to maximize each mutual information separately. For large N we know that

1. Exchangeable systems have small neural complexity. More precisely

$$\sup_{(X_1,\ldots,X_N) \text{ exchangeable}} \mathcal{I}(X) = o(N^{2/3+\epsilon}), \qquad N \to +\infty,$$

   for any $\epsilon > 0$. In particular maximizers are neither unique nor exchangeable.

2. if X is a maximizer, then its support does not exceed a fixed proportion of the configuration space.

The first property is an example of a spontaneous symmetry breaking.

Main result

1. For any sequence $X^N = (X_1^N, \dots, X_N^N)$

$$\limsup_{N\to\infty} \frac{\mathcal{I}(X^N)}{N \log 2} \le \frac{1}{4}.$$

2. For any sequence $X^N = (X_1^N, \dots, X_N^N)$ such that

$$\lim_{N\to\infty} \frac{\mathrm{H}(X^N)}{N \log 2} = x \in [0,1],$$

   we have

$$\limsup_{N\to\infty} \frac{\mathcal{I}(X^N)}{N \log 2} \le x(1-x).$$

3. For all $x \in [0,1]$ there is at least a sequence X^N such that

$$\lim_{N\to\infty} \frac{\mathrm{H}(X^N)}{N \log 2} = x, \qquad \lim_{N\to\infty} \frac{\mathcal{I}(X^N)}{N \log 2} = x(1-x).$$

Random Sparse Configurations

Let $N \geq 2$ and $1 \leq M \leq N$ be an integer. We denote

$$\Lambda_n := \{0,1\}^n, \qquad \forall\ n \geq 1.$$

We consider a family $(W_i)_{i \in \Lambda_M}$ of i.i.d. variables, each uniformly distributed on Λ_N. We define a *random* probability measure on Λ_N

$$\mu^{N,M}(x) := 2^{-M} \sum_{i \in \Lambda_M} \mathbb{1}_{(x=W_i)}, \qquad x \in \Lambda_N.$$

Theorem

Let $x \in]0,1[$. We have a.s. and in L^1

$$\lim_{N \to +\infty} \frac{\mathrm{H}(\mu^{N,\lfloor xN \rfloor})}{N \log 2} = x \tag{1}$$

$$\lim_{N \to +\infty} \frac{\mathcal{I}(\mu^{N,\lfloor xN \rfloor})}{N \log 2} = x(1-x). \tag{2}$$

Proof

By the symmetries

$$\mathbb{E}\left(\frac{\mathcal{I}(\mu^{N,\lfloor xN \rfloor})}{N\log 2}\right) = \frac{2}{N+1}\sum_{k=0}^{N} h_k - h_N.$$

One can show a sharp transition for h_k

1. For $k \leq M$ we have $k - 2^{\frac{k-M}{2}} \leq h_k \leq k$
2. For $k > M$ we have $M - 2^{M-k} \leq h_k \leq M$

Approximate maximizers

This sequence satisfies the following property: as $N \to +\infty$,

1. if $y \in]0, x]$ then for *almost* all subsets S with $|S| = \lfloor yN \rfloor$, X_S is *almost* uniform, i.e. almost independent;
2. if $y \in [x, 1[$ then for *almost* all subsets S with $|S| = \lfloor yN \rfloor$, X is *almost* a function of X_S.

It turns out that the same property is shared by any sequence of approximate maximizers.
This property describes the interplay between *differentiation* and *integration* that biologists expect to find in complex systems.

New questions and problems

1. How to maximize under further constraints?
2. Is there any evolutionary (learning) mechanism with interesting interplay with neural complexity?
3. How to combine N.C. with an underlying geometry?
4. How to estimate the N.C. of a real system?
5. Is it possible to detect critical phenomena (epidemics)
6. How to compute N.C. of classical systems, like the Ising model?
7. Is there an interpretation in terms of Information theory?