Toward a robust 2D spatio-temporal self-organization

Thomas Girod, Laurent Bougrain and Frédéric Alexandre

LORIA-INRIA
Campus Scientifique - B.P. 239
F-54506 Vandœuvre-lès-Nancy Cedex, FRANCE

Abstract. Several models have been proposed for spatio-temporal self-organization, among which the TOM model by Wiemer [1] is particularly promising. In this paper, we propose to adapt and extend this model to 2D maps to make it more generic and biologically plausible and more adapted to realistic applications, illustrated here by an application to speech analysis.

1 Spatio-temporal self-organization

Fundamental property, in biological as well as artificial systems, is that of adaptive information representation. Self-Organizing Maps (SOM), proposed by Kohonen [2] in the framework of cortical modeling and extensively used for various tasks of information processing, underline how, from simple learning and connectivity rules within a map of neurons, a topological representation can emerge, where similar data activate close regions of the map. The resulting representation is interesting for several reasons: using short connections in the brain saves energy; the neighborhood property is robust to noise and makes easier communication between neurons representing close stimuli. Even if there is no strong theoretical reasons, two-dimensional maps are generally used either with inspiration to the cortex where this kind of structure corresponds to an important step in evolution [3] or in classical information processing where a 2D representation is natural for visualizing data.

In early works on such topological maps, similarity between data that generates neighborhood in representation derives from a spatial distance, the choice of which is often very critical to obtain good performances. Nevertheless, several reasons related to biological plausibility and to the nature of stimuli to be processed have underlined the need to implement spatio-temporal self-organization. Indeed, the 'time' dimension is often reported as participating to the cortical mapping [4] and most information processing tasks include a temporal dimension necessary to integrate in the resulting representation, like for example in speech processing. To that end, numerous spatio-temporal self-organizing models have been proposed (see Kremer [5], Vaucher [6], Zehraoui and Fessant [7] for some reviews). These models adress temporal integration and representation in very different ways and, today, no generic approach can be distinguished, as it is the case for purely spatial SOM.

Among these works, we find the Time Organized Map (TOM) model by Wiemer [1] very interesting for several reasons. It is deeply inspired from biological data, it represents time in propagating waves and it proposes interesting performances for selected tasks. Nevertheless, it has several weaknesses, among which its unique implementation on a one-dimensional neuronal layer and for too artifical tasks. The goal of this paper is to propose tools (how to measure spatio-temporal self-organisation?) and mechanisms (how to represent spatio-temporal similarity in 2D?) to extend and adapt the TOM model to processing real-world spatio-temporal data in neuronal 2D maps.

2 TOM algorithm

The TOM architecture is based on Kohonen maps : a perception vector on which stimuli are displayed is fully connected to a layer of neurons. TOM integrates time with inter-stimuli intervals (ISI), which express the time interval between any pair of stimuli. When a stimulus is displayed at time n, it comes with an ISI value expressing the time relation between previously presented and current stimuli. In most cases, data are representing sequences of stimuli and an ISI value is associated to each element of the sequence. Thus the proposed ISI function is additive ($ISI(a, c) = ISI(a, b) + ISI(b, c), a < b < c$) and commutative ($ISI(a, b) = ISI(b, a)$).

The TOM algorithm aims at translating temporal durations into distances on a self-organizing layer. At each step, a wave is propagated from the previous winner ((2) in figure1). The distance of propagation depends on the ISI and a fixed propagation speed v. This wave will cause a shift (3) in the determination of the current winner (1), which will eventually alter the self-organization of the layer (5).

Moreover, rather than having an ensemble of learning neurons around the winner, there is only one learning neuron by step (6). As proposed by Lo and Bavarian [8], this is made possible by a decreasing noise (4).

As fully described and illustrated in [1], this iterative process results in a topology on the layer of neurons realizing a balance between spatial and temporal proximity between stimuli.

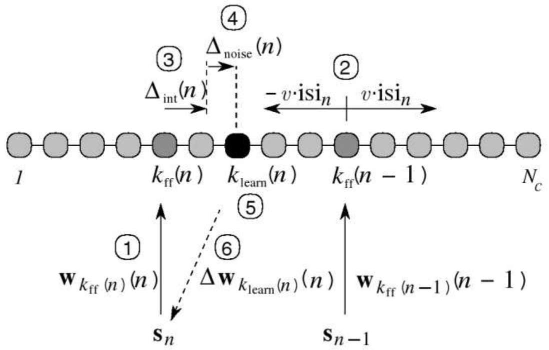

Fig. 1: One iteration of TOM algorithm on a 1D map (adapted from [1]).

3 Scaling TOM algorithm

The TOM algorithm has been developped on 1D layers of neurons. Our goal is to achieve a more biologically plausible algorithm, to make it more robust, and to obtain a visual representation of more complex and realistic sequences.

3.1 Quality measure of spatio-temporal self-organization

In order to analyze the behavior of TOM algorithm, we need to quantify the quality of the self-organization process. We do so by finding how prototypes from the training set are arranged on the layer using a spatial self-organization distance, and if this arrangement is well-founded according to a temporal self-organization distance.

3.1.1 Spatial self-organization

TOM algorithm is based on a *winner-take-all* policy for training. This tends to group similar neurons and avoid strong similarities between topological distant neurons. To better evaluate the preservation of the spatial information in a given layer, for each stimulus s, we compute its Euclidian distances with every prototype, and we define a subset containing the closest prototypes. Finally, we capture the topological coordinates of these prototypes into a footprint noted fp_s.

We can calculate a standard deviation from fp_s, giving us the scattering of the footprint. The lower is the scattering, the better is the corresponding spatial organization. Finally, we will reduce footprints to their barycenters.

3.1.2 Temporal self-organization

Temporal topology achieves its influence on the self-organization process through the lateral interactions. This tends to alter the footprints arrangement to satisfy the temporal constraints.

Let us take a pair of stimuli (s_1, s_2) from the training set. The temporal distance between them is $d_t = isi(s_1, s_2) * v$, v the propagation speed. fp_{s1} and fp_{s2} are s_1 and s_2 footprints (barycenters). Let the spatial distance $d_s = ||fp_{s1} fp_{s2}||$. If the topography of the layer satisfies the temporal constraint, then $d_t \approx d_s$.

From the temporal constraints, the self-organization is optimal when $d_t \approx d_s$ for any couple (s_1, s_2). Therefore $(d_s - d_t)^2$ is the error factor of temporal relation between s_1 and s_2. By averaging the error on several couples of stimuli we get a general error factor in temporal self-organization.

3.2 Extension to 2D maps

Wiemer description of TOM algorithm is concerned with 1D layers of neurons only. A logical evolution of TOM algorithm is to extend it to 2D maps. We will show that the process may not be as simple as it seems.

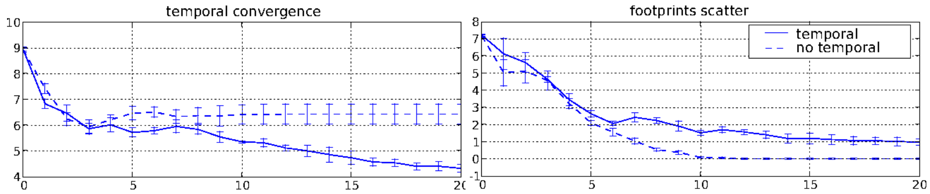

Fig. 2: An example of use of the spatial and temporal measures. The left part shows the mean scattering in time. The right part shows the temporal self-organization in time. Plain lines are a case with strong lateral interaction, dashed lines are a case without lateral interaction.

3.2.1 Lateral interaction

On a 2D map, one can visualize the wave propagation as a stone dropped in a pool. At step n the propagation of the wave from $k_{ff}(n-1)$ allows us to define $\tilde{k}(n)$, source of the lateral interaction at step n. $\tilde{k}(n)$ being the closest point of the wave to $k_{ff}(n)$, it is easy to get its coordinates geometrically.

Depending on isi and v, the wave may propagate partly or fully outside of the map. As $\tilde{k}(n)$ should refer to a neuron of the map, we may not be able to define $\tilde{k}(n)$ *stricto sensu*. To clear up this issue, we use $\tilde{k}(n)$ coordinates the way they are calculated, even if they do not fit into the map's limits.

3.2.2 Noise deviation

The second shift of activity is the noise vector ξ. On a 1D map, the noise is generated by a normal distribution $\mathcal{N}(0, \sigma^2_{noise}(n))$.

On a 2D map, the orientation distribution of the noise vector must be uniform. Moreover, the vector's norm distribution must follow the distribution $|\mathcal{N}(0, \sigma^2_{noise}(n))|$. Thus, we proceed as follows.

$\mathbf{\Theta}$ follows the zero-mean multivariate normal function with identity matrix of covariance $\mathcal{N}_2(\mathbf{0}, \mathbf{I})$

$$\xi = \frac{\mathbf{\Theta}}{||\mathbf{\Theta}||} * |\mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^{\mathbf{2}}_{\mathbf{noise}}(\mathbf{n}))| \tag{1}$$

3.3 Sequential input

TOM training algorithm is based on the same principles as SOM : the sequential choice of stimuli, randomly selected from a training set. Random choice has proven to be efficient to avoid maps folding. However in TOM, time is taken into account and the training set is supposed to be a sequence.

In examples chosen by Wiemer, only the ISI between stimuli is significant and no contextual effect is present from one stimulus to its close neighbours. It is thus possible to select, for learning, two stimuli randomly and simply evaluate their ISI. The learning of each stimulus is thus statistically influenced by every stimuli of the training set, and the prototypes arrangement on the map results from a global consensus.

We have chosen real-world applications, like speech processing, where contextual effects in time are very important and necessary to be learned. Displaying stimuli in order will reduce the number of possible lateral interactions. The learning of a stimulus will always be influenced by close neighbours in the sequence and no others. Therefore the general arrangement of prototypes on the map will be the combination of local arrangement between temporally close stimuli. Here, the combinatorial use of the corpus is no more possible, but real world applications like speech processing generate huge amounts of data. However, unlike the case of random display, the optimal arrangement will not necessarily be linear as the additivity of ISI is no more used, which could result in more difficult convergence.

3.4 Experiments

One major domain of application of spatio-temporal models is speech analysis. Spatio-temporal self-organizing networks can help to visualize and interpret main sequences of phonemes. They can also be used in an anticipatory way.

We have applied our model to the TIMIT Acoustic-Phonetic Speech Database [9]. We trained the network using the 38 American English speakers from New England. 12 MFCC (Mel Frequency Cepstral Coefficients) have been extracted from the 258 sentences available with constant sampling period of 10ms.

Figure 3 shows that phonemes are organized by zones. A phoneme can appear in several zones according to usual pre- and post-phonemes. This representation allows a more simple visualization of complex sequences.

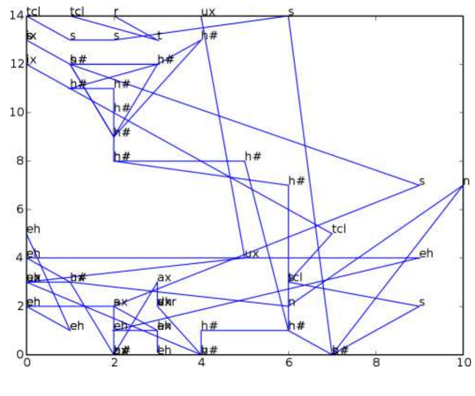

Fig. 3: Visualization on our 2-dimensional statio-temporal self-organized network of one sentence of the TIMIT speech database: "Nevertheless it's true". Beginning and ending silence regions are marked with h#.

4 Conclusion

From a very interesting spatio-temporal self-organizing model, restricted to 1D layers of neurons and to applications where stimuli have no contextual effects one on the others, we have proposed new tools and mechanisms toward a more robust spatio-temporal self-organization in 2D. Preliminary encouraging results have been reported on the difficult task of speech processing and additional work remains to be done to better interpret the results and improve the model.

More precisely, the study of the convergence along the temporal dimension indicates some problems that could be related to the initialization and to the non-transitive property of ISI. More systematic tests must be done on larger corpus. It seems now likely that, instead of adapting mechanisms built from global evaluation of the map as presented above, it would be more convenient to look for only local interaction rules. This is all the more interesting as this is consistent with a bio-inspired view of spatio-temporal self-organization.

References

[1] J.C Wiemer. The time-organized map algorithm: Extending the self-organizing map to spatiotemporal signals. *Neural Computation*, 15(5), 2003.

[2] T. Kohonen. *Self-organizing maps*. information science. Springer, Berlin, 3 edition, 2001.

[3] D.H. Ballard. Cortial connections and parallel processing: structure and function. *Behavioral and brain sciences*, 1986.

[4] F. Spengler, T. Hilger, X. Wang, and M. Merzenich. Learning induced formation of cortical populations involved in tactile object recognition. *Soc Neurosc Abstracts*, 22(1), 1996.

[5] S.C. Kremer, A. Guilherme, A. Barreo, and D. Chen. A taxonomy for spatio-temporal connectionist networks revised: the unsupervised case. *Neural Computation*, 15(6), 2003.

[6] G. Vaucher. *A la recherche d'une algèbre neuronale spatio-temporelle*. PhD thesis, université de Rennes 1, 1996.

[7] Y. Bennani, F. Zehraoui, and F. Fessant. *Apprentissage connexionniste*, chapter 7. 2006.

[8] Z. Lo and B. Bavarian. On the rate of convergence in topology preserving neural networks. *Biol Cybern*, 1991.

[9] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett. Darpa timit acoustic-phonetic continous speech corpus cd-rom. *NASA STI/Recon Technical Report N*, 93:27403–+, February 1993.