

Encoding the Temporal Statistics of Markovian Sequences of Stimuli in Recurrent Neuronal Networks

Alessandro Usseglio Viretta¹, Stefano Fusi², and Shih-Chii Liu¹

¹ Institute of Neuroinformatics, University and ETH Zurich
Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[ale, shih]@ini.phys.ethz.ch

² Institute of Physiology, University of Bern
Bühlplatz 5, CH-3012 Bern, Switzerland
fusi@cns.unibe.ch

Abstract. Encoding, storing, and recalling a temporal sequence of stimuli in a neuronal network can be achieved by creating associations between pairs of stimuli that are contiguous in time. This idea is illustrated by studying the behavior of a neural network model with binary neurons and binary stochastic synapses. The network extracts in an unsupervised manner the temporal statistics of the sequence of input stimuli. When a stimulus triggers the recalling process, the statistics of the output patterns reflects those of the input. If the sequence of stimuli is generated through a Markov process, then the network dynamics faithfully reproduces all the transition probabilities.

1 Introduction

Our current understanding of the neuronal mechanisms that permit biological systems to encode and recall temporal sequences of stimuli is still marginal. In the last decade many studies attempted to relate the problem of encoding temporal sequences to the generation of associations between visual stimuli [1, 2,3]. Interestingly, in one simple case [4], in which the stimuli were presented in a fixed temporal order, it was possible to study the neural correlate of this type of associative memory. Cortical recordings displayed significant correlations between the patterns of activity elicited by neighboring stimuli in the temporal sequence. Hence these internal representations of the visual stimuli encode the temporal context in which stimuli were repeatedly presented during training. These patterns of activities were stable throughout long time intervals and have been interpreted as global attractors of the network dynamics [3]. Despite the fact that the learning rule used by Griniasty [3,5] makes use of only the information about the contiguity of two successive stimuli, these attractors are correlated up to a distance of 5 in the temporal sequence, similar to that observed in the experiment of Miyashita [4]. Here we extend these mechanisms to a more general situation. First, we show that the network in the presence of noise

of sufficient amplitude can spontaneously jump to a pattern of activity representing a different stimulus. The pattern of connectivity between the internal representations of the stimuli encodes the transition probability, and the presentation of a single stimulus can trigger the recalling of a sequence of patterns of activity corresponding to temporally correlated stimuli. Noise is exploited as in [2], and time is essentially encoded in the escape rates from the attractors. Second, the pattern of connectivity encoding the transition probabilities can be learned when the network is repeatedly exposed to the temporal statistics of the stimuli. The transition probabilities are automatically extracted during this “training phase” and encoded in the synaptic matrix. The learning rule was inspired by the one introduced in [3] and it makes use of the information carried by the current stimulus and by the pattern of activity elicited by the previous stimulus. A possible mechanism for making this rule local in time has been suggested in [6,5] and relies on the stable activity that is sustained by the network in the interval between two successive stimuli.

2 The Model

We implemented a recurrent neural network with N excitatory neurons, labeled by index i , $i = 1 \dots N$. The state of neuron i is described by the variable S_i : $S_i = 1$ ($S_i = 0$) corresponds to a firing (quiescent) neuron. The network is fully connected with binary excitatory synapses J_{ij} from neuron j to neuron i [6]. The neuron’s state is updated using the *Glauber dynamics* (see, for example, [7]), in which $S_i=1$ with probability $g_\beta(h_i) \equiv \frac{1}{1+\exp(-2\beta h_i)}$, where $1/\beta$ is the *pseudo temperature* of the network and $h_i = \sum_j J_{ij} S_j - I$ the synaptic input, or *field*, to the neuron. The global inhibition I dynamically adjusts the activity of the network. I depends on the fraction $F = \frac{1}{N} \sum_k S_k$ of neurons that are active as expressed by the following equation:

$$I(t + 1) = \begin{cases} s_0(F(t) - s_1), & \text{If } F > f_m \\ I_m & \text{If } F \leq f_m \end{cases} \quad (1)$$

where f_m is the threshold for the global activity, and I_m the minimum inhibition. The parameters s_0 and s_1 are chosen so that $s_0(f_0 - s_1) = I_0$, where I_0 is usually chosen between the maximum input to quiescent neurons and the minimum input to active neurons when the network state corresponds to a learned pattern of activity, and f_0 is the average activity of the learned patterns.

2.1 Learning Rule

Following [8], we implemented learning as a stochastic process. During the presentation of a pattern η_i^μ , the μ -th pattern of a temporal sequence, the neuron states are set to $S_i = \eta_i^\mu$ and the binary synapses are updated according to the following rules:

- If both the pre and the post-synaptic neurons are active, then a transition to the potentiated state occurs with a probability q_+ .

- If only one of the two neurons connected by the synapse is active, then a transition to the depressed state of the synaptic weight occurs with a probability q_- .
- If both the neurons are inactive, the synapse is left unchanged.
- If the pre-synaptic activity imposed by the previous stimulus in the sequence is high and the post-synaptic activity induced by the current stimulus is also high, then a transition to the potentiated state of the synaptic weight $J_{ij} : 0 \rightarrow 1$ occurs with a probability $q_\times = \lambda_f q_+$, where $\lambda_f < 1$ (see [6, 5]). This part of the rule allows one to connect events that are separated in time and to encode the information about the temporal context in which the stimuli are presented.

In the case of random patterns, if the average activity is f , the probability for two randomly chosen neurons to be both active is f^2 , whereas the probability for the two neurons to have different activities is $2f(1 - f)$. In order that the probability for long term depression is approximately equal to the probability of long term potentiation, we choose $q_- = \frac{fq_+}{2(1-f)}$.

The learning of a transition results in making neurons that belong to the same pattern have a finite probability of providing a nonzero synaptic input to neurons belonging to a different pattern. In the case where several patterns and transitions have been learned, the transition probability is shown, for a suitable choice of the network parameters, to be a monotonically increasing function of the relative frequency of presentation of the patterns during the learning phase.

2.2 Learning Markov Processes

We assume that the sequence of patterns to be learned is generated according to a Markov process, that is to a random process in which the probability for which a pattern is chosen depends only on the immediate preceding chosen pattern. Given a Markov matrix M , the element $M_{\mu\nu}$ is the probability of transition from state s_μ to state s_ν . To teach the network to encode a Markov process, a pattern η_i^μ is randomly chosen from a pool of p patterns. The next pattern η_i^ν is chosen according to the transition probability of the Markov process from state s_μ to state s_ν . The two patterns are presented to the network, which thus learns them and the transition from the first to the second. This process is repeated until the synaptic matrix has reached its asymptotic configuration. Alternatively, the probability for each synapse to be potentiated can be analytically calculated in the limit for vanishing q_+ and a pattern sequence of infinite length [9]. Calculating the weight matrix in this way is computationally less expensive than on-line learning. Given the auxiliary variables P_{ij} and Q_{ij} , which are respectively proportional to the number of events leading to synapse potentiation and depression,

$$\begin{cases} P_{ij} = \sum_{\mu=1}^p (\eta_i^\mu \eta_j^\mu q_+ + \sum_{\nu=1}^p (\eta_i^\nu \eta_j^\mu M_{\mu\nu} q_\times)) \\ Q_{ij} = \sum_{\mu=1}^p ((1 - \eta_i^\mu) \eta_j^\mu q_- + q_- \eta_i^\mu (1 - \eta_j^\mu)) \end{cases} \quad (2)$$

the probability for the synapse J_{ij} to be 1 is given by $p_{ij} = \frac{P_{ij}}{P_{ij} + Q_{ij}}$.

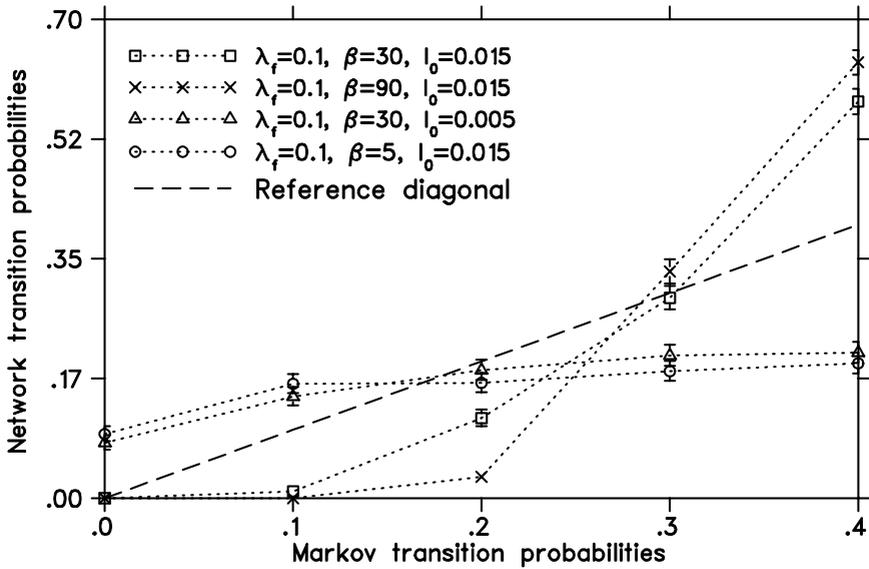


Fig. 1. Transition probabilities as a function of the Markov process transition probabilities for different combinations of β and I_0 . Upper and lower error bars have been calculated using the formula in Meyer [10], with $k = 1$ and $n = 700$ (490 neurons, 7 patterns). An increase in β decreases the transition probabilities corresponding to low Markov transition probabilities and increases those corresponding to high Markov transition probabilities. A low I_0 or β makes the dynamics less dependent on the connectivity pattern. The corresponding plot is shallower, indicating a poor capability in reproducing the statistics of the input sequences.

3 Results

The performance of the network was evaluated on a sequence generated by a Markov chain. The network was able to reproduce the statistics of the input sequences without an external input. Given an initial pattern of activity, the network made spontaneous transitions to other patterns with a probability close to the corresponding Markov chain probability. The performance of the network strongly depended on the temperature $1/\beta$ and inhibition I_0 . The on-line learning procedure and the analytical derivation of the synaptic matrix led to qualitatively equivalent results (data not shown).

To evaluate the performance we computed the transition probability matrix $T_{\mu\nu}$. This matrix describes the probability that the network makes at some time t a transition from an initial state highly resembling pattern μ (the overlap $m^\mu = \frac{1}{N} \sum_i S_i \eta_i^\mu$ [7] between the network state and pattern η^μ is the maximum), into a state that resembles maximally pattern η^ν .

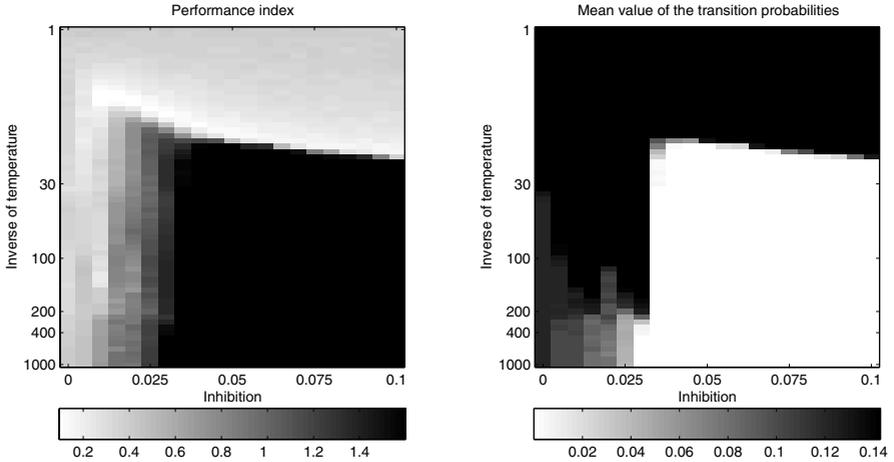


Fig. 2. Left: performance index Π as a function of I_0 and β . Lower values of Π correspond to better performance ($\Pi \geq 0$). Right: mean value of the transition probabilities ($\lambda_f = 0.1$). Both diagrams show the average over 10 networks with different synaptic configurations: the other parameters are identical. The transition probability matrix was measured after 100 transitions per pattern from the starting state and allowing a maximum of 100 updates of all neurons to make a transition. The performance of the network is maximal in a region around $I_0 = 0.01$ and $\beta = 14$. For I_0 up to 0.04 the optimal β is roughly proportional to I_0 and for higher inhibition the best performance lies in the area just above the sharp boundary in the plot. In the right-hand plot, the sharp boundary goes together with a sudden decrease in the mean value of the transition probability for an increasing β . Very low temperatures trap the network in the initial basin of attraction.

3.1 Performance Evaluation

The Markov matrix M used to generate the input patterns has been compared with the transition probability matrix T . The closer the elements of the transition probability matrix $T_{\mu\nu}$ are to the corresponding values of the Markov matrix $M_{\mu\nu}$, the better the network is at reproducing the statistics of the Markov process. We have analyzed the performance of the network as a function of I_0 and β using a Markov matrix whose rows are randomly reshuffled versions of $\{0, 0, 0, 0.1, 0.2, 0.3, 0.4\}$. Figure 1 shows the performance of the network for four combinations of λ_f , β and I_0 . Depending on the combination of the parameters, the network reproduces more or less faithfully the statistics of the temporal sequence used for training. Π has been chosen as the performance index of a network that has learned P patterns.

$$\Pi = \frac{1}{P^2} \sum_{\substack{\mu, \nu = 1 \\ M_{\mu\nu} \neq 0}}^P \frac{|M_{\mu\nu} - T_{\mu\nu}|}{\frac{M_{\mu\nu} + T_{\mu\nu}}{2}} \quad (3)$$

Figure 2 shows the performance index I and the mean value of the transition probabilities as a function of I_0 and β for $\lambda_f = 0.1$.

4 Conclusions

We described a recurrent neuronal network consisting of binary neurons and binary synapses that is able to learn and reproduce the statistics of the input sequences used in the learning phase. The performance has been evaluated by comparing the transition probability matrix with the Markov matrix used in generating the input sequences. The performance of the network strongly depends on the temperature and global inhibition. The binary coding of the variables, the stochastic nature of the network and its scalability makes this architecture attractive for implementation in hardware.

References

1. D. Kleinfeld and H. Sompolinsky. Associative neural network model for the generation of temporal Patterns. *Biophysical Journal*, 54:1039–1051, 1988.
2. J. Buhmann and K. Schulten. Noise-driven temporal association in neural networks. *Europhysics Letters*, 4(10):1205–1209, 1987.
3. M. Griniasty, M. V. Tsodyks, and D. J. Amit. Conversion of temporal correlations between stimuli to spatial correlations between attractors. *Neural Computation*, 5:1–17, 1993.
4. Y. Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335(6193):817–820, October 1988.
5. V. Yakovlev, S. Pusi, E. Berman, and E. Zohary. Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. *Nature neuroscience*, 1(4):310–317, August 1998.
6. N. Brunel. Hebbian learning of context in recurrent neural network. *Neural Computation*, 8:1677–1710, 1996.
7. D. J. Amit. *Modeling Brain Function*. Cambridge University Press, New York, 1989.
8. D. J. Amit and S. Fusi. Learning in neural networks with material synapses. *Computation*, 6:957–982, 1994.
9. N. Brunel, F. Carusi, and S. Fusi. Slow stochastic hebbian learning of classes of Stimuli in a recurrent neural network. *Network*, 9:123–152, 1998.
10. P. L. Meyer. Introductory probability and statistical applications. Addison-Wesley, Reading, MA, 1965.