Boltzmann Machines for signals decomposition. Application to Parkinson's Disease control

Rémi SOURIAU[1,*], Vincent VIGNERON[1], Jean LERBET[2], Hsin CHEN[3]

[1]IBISC, Univ Evry, Université Paris-Saclay, 91025, Evry, France.

[2]LaMME, CNRS, Univ Evry, Université Paris-Saclay, 91037, Evry, France.

[3]NEL, Dept. of Electrical Engineering, National Tsing Hua University, Taiwan.

remi.souriau@univ-evry.fr

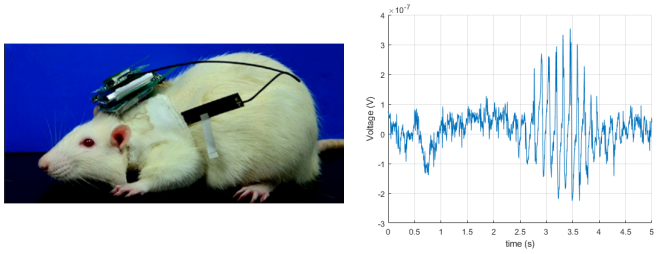
Résumé – Cet article présente une nouvelle méthode de décomposition des signaux à l'aide d'un modèle d'apprentissage non supervisée: la Machine de Boltzmann Restreinte continue (cRBM) basée sur la structure des réseaux de diffusion. Une application pour la détection des pics d'amplitude de tension enregistrée dans une région profonde du cerveau est également présentée.

Abstract – This article presents a new method of decomposition of signals with an unsupervised training model: the continuous Restricted Boltzmann Machine (cRBM) based on the structure of the Diffusion Network. An application for the detection of *High-Voltage Spindle* (HVS) in signals recorded in the brain is also presented.

Introduction

The Parkinson's Disease (PD) is a progressive *neurodegenerative* disease. The depletion of the dopamine in the basal ganglia network leads to several symptoms like rigidity, posture instability, slow motion or pain for example. Applying an electronic periodic signal to the subthalamic nucleus in the deep brain is an efficient treatment for advanced PD patient. However open-loop Deep Brain Stimulation (DBS) leads to side effects like psychiatric ones. Another consequence of applying permanently the DBS is the need to replace batteries which requires surgery. Recent studies show that the arrival of symptoms can be predicted by detecting the presence of High-Voltage Spindles (HVS) (see Fig.1b) in Local Field Potentials (LFPs) [3]. The HVS are synchronous spike-and-wave patterns oscillating in the 5-13 Hz frequency band. Suppressing HVS signals is found useful for delaying the progress of PD and deleting symptoms.

The detection of the HVS is a challenging problem for two reasons. First, symptoms of the PD appear some milliseconds after the first HVS. Second, PD is a progressive disease : signals depend on the patient, on the location of recording probes or on the advanced state of the disease. A fast and robust model capable to learn automatically from the data on real time is then required. A previous work on Gaussian Process leads to satisfying results [6]. However, the absence of *ground truth* for the training step makes results highly variable among individuals due to major differences in signals recording between different people. In this article, we focus on an unsupervised graphical model, the Boltzmann Machine.



(a) PD rat with a neural prostheses (b) HVS : Temporal representation

FIGURE 1 – Signals recorded in LFPs. Fig.1a is a PD rat with an implant for recording signals and apply DBS. Fig.1b plots a brain signal. The HVS is located between 2.5 and 5 seconds. it is characterized by a fundamental frequency between 5 and 13 Hz.

The seminal works of Hopfield [4] lead to the emergence of a large family of models, in particular to the Boltzmann Machine [9] and to the Diffusion Network (DN), a continuous stochastic neural network, studied by Movellan [5]. Chen and Murray [1, 2] first proposed a continuous Restricted Boltzmann Machine (cRBM) architecture based on the DN and its VLSI implementation. The cRBM is used in this paper for the first time for signal decomposition. The paper is organized as follows. Section 1 presents the principles of the DN and the cRBM. The convergence properties of the cRBM for the signal decomposi-

*This work is supported by the doctoral school of Université Paris-Saclay and the Ministry of Science and Technology of Taïwan (MOST)

tion is detailed in section 2. Finally, we discuss about the application of the detection of HVS using the cRBM.

1 Principles of the Diffusion Network

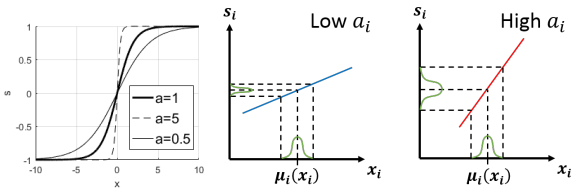
1.1 Diffusion Network

The DN is a continuous stochastic model described by a Stochastic Differential Equation (SDE) [5]. The aim of the DN is to model the dynamic dependencies between signals. We note $\boldsymbol{X(t)} = (x_1(t), \ldots, x_n(t))^T$ the signal at instant t, where ^{T} is the transpose operator. For all neurons j, the SDE is given by :

$$\mathrm{d}x_j(t) = \mu_j(X(t))\mathrm{d}t + \sigma \mathrm{d}B_j(t). \quad (1)$$

The first term in the right member of Eq.1 is the *drift* term and the second term is the *diffusion* term. $\mathrm{d}B_j(t)$ is a Brownian motion and σ is the noise standard deviation. Training the DN consists to find the parameters of the drift $\mu_j(.)$. The formulation of the drift term of neuron j is :

$$\mu_j(\boldsymbol{X(t)}) = \kappa_j \left(-\rho_j x_j(t) + \xi_j + \sum_{i=1}^{n} W_{ij}\phi_i(x_i(t)) \right) \quad (2)$$



(a) Neuron response. (b) Regularization of noise.

FIGURE 2 – Influences of the parameters a_i. Fig.2a displays the activation between -1 and 1 for three different values of a_i. The bigger/smaller a_i is, the more the neuron will have a binary behavior. In Fig.2b, the two schemes describe how the noise influences the state of the neuron. At a time instant t, $x_i(t)$ is a Gaussian random variable centered in $\mu_i(X(t))$ with a variance σ (see Eq.1). On the left figure, a_i has a low value, the slope of the function is almost horizontal. The dispersion of the noise is "squeezed" and the state of the neuron is almost deterministic. If a_i is high like on the right scheme, the slope of the function is almost vertical and the neuron state is stretched.

In Eq.2 κ_j, ρ_j, ξ_j and $\phi_i(x_j)$ are resp. the inverse capacitance, the inverse resistor, the bias and the state of the neuron j. W_{ij} is the weight between input i and neuron j. ϕ_j is a sigmoïd function (see Fig.2a).

$$s_j = \phi_j(x_j) = \theta_L + (\theta_H - \theta_L)\frac{1}{1+\exp(-a_j x_j)} \quad (3)$$

θ_L and θ_H are respectively the lower and the upper bounds of the function and a_j is the slope parameter of the activation function. The influences of the parameter a_j are double as illustrated in Fig.2.

1.2 Continuous Restricted Boltzmann Machine

In a cRBM [1] using the structure of neurons of a DN, we suppose we have the following equality for all the neurons $\rho_i \kappa_i \Delta t = 1$. Links in a cRBM are symmetric and there are no connection between neurons in the same layer. The structure of the neuron j is given in Fig.3. Each input is weighted and summed with a bias. The added noise in the neuron structure explains the stochastic behavior of the model and helps the cRBM to not fall in a local minimum during the training step.

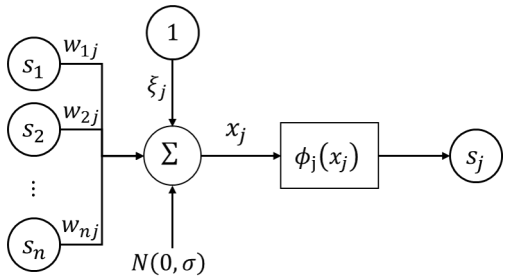


FIGURE 3 – Structure of the neuron j of a cRBM. The expression of the activation function $\phi_j(x_j)$ is given in Eq.3. $W = \{W_{ij}\}$ is the transfer matrix and $\boldsymbol{\xi} = \{\xi_i\}$ is the bias vector.

The energy function of the cRBM is the same as the energy function of the Hopfield Network [4], *i.e.*

$$E_{cRBM}(\boldsymbol{s} = \{\boldsymbol{v}, \boldsymbol{h}\}) = -\boldsymbol{h}^T W \boldsymbol{v} - \boldsymbol{v}^T \boldsymbol{\xi_v} - \boldsymbol{h}^T \boldsymbol{\xi_h} + \sum\nolimits_i \int_0^{s_i} \phi_i^{-1}(s')\mathrm{d}s' \quad (4)$$

where $\boldsymbol{h}, \boldsymbol{v}$ are resp. the hidden and the visible unit values. The continuous integration term ϕ_i^{-1} between 0 and s_i in Eq. 4 makes sense only if 0 is in the range of variation of the neurons. The lower bound 0 in the integral term ensures that $E_{cRBM}(\boldsymbol{0}) = 0$. Note that the bounds θ_L and θ_H can be different between layers. We choose specific bounds for visible units to force neurons to cover the range of variation. Learning the cRBM consists to estimate the set of parameters $\boldsymbol{\lambda} = \{W_{ij}, \xi_i, a_i\}$ which maximize the log-likelihood :

$$\mathcal{L}(\boldsymbol{\lambda}) = \sum_{\boldsymbol{v},\boldsymbol{h}} \log P_{cRBM}(\boldsymbol{v}, \boldsymbol{h}) = \sum_{\boldsymbol{v},\boldsymbol{h}} -E_{cRBM}(\boldsymbol{v}, \boldsymbol{h}) + c \quad (5)$$

c being a constant. Contrastive Divergence rule proposed by Hinton [9] is used to train the cRBM. The update law for the transfer matrix and the bias remains the same as the RBM. The update rules for the weights and the activation function parameters are :

$$\begin{aligned} \Delta W_{ij} &\sim < s_i s_j >_0 - < s_i s_j >_1 \\ \Delta \xi_i &\sim < s_i >_0 - < s_i >_1 \\ \Delta a_i &\sim \frac{1}{a_i} \int_{<s_i>_1}^{<s_i>_0} \phi_i^{-1}(s')\mathrm{d}s' \end{aligned} \quad (6)$$

with $< . >_0$ the expected value over the training set and $< . >_1$ the expected value after one step of Gibbs Sampling.

2 Unsupervised signal decomposition

Suppose a short time window $\boldsymbol{v} = (v(t_1), \ldots, v(t_m))^T$. In a cRBM composed of n hidden neurons bounded between $\theta_H = +\theta$ and $\theta_L = -\theta$, we note h_i, the $i-$th component of the hidden layer defined as $h_i = \phi_i(x_i)$ where $x_i \sim \mathcal{N}(z_i, \sigma))$ with :

$$z_i = \xi_i^h + \sum_{j=1}^{m} W_{ij} v(t_j) \tag{7}$$

Let W_i the $i-$th line vector of the transfer matrix W. W_i can be seen as a temporal vector $W_{ij} = W_i(t_j)$. For a signal without offset the hidden bias vector $\boldsymbol{\xi^h}$ tends to zeros during the training and z_i becomes :

$$z_i = \sum_{j=1}^{m} W_i(t_j) v(t_j) = \Gamma_{W_i v}(0) \tag{8}$$

The sum over i of the correlation function $\Gamma_{W_i \boldsymbol{v}}$ is also present in the energy function :

$$-h^T W v = -\sum_{i=1}^{n} h_i z_i = -\sum_{i=1}^{n} h_i \Gamma_{W_i \boldsymbol{v}}(0) \tag{9}$$

The hidden units h_i are bounded. Then the log-likelihood $\mathcal{L}$ (Eq.5) requires to maximize all the z_i. z_i is a scalar product, the Cauchy–Schwartz inequality tell us :

$$z_i = W_i \boldsymbol{v} \leq \|W_i\| \times \|\boldsymbol{v}\| \tag{10}$$

The equality between the visible units and the line vector of transfer matrix are co-linear, *i.e.* $W_i^T = \alpha \boldsymbol{v}$. Intuitively, we could think the components of W will continuously augment during the learning because the bigger α is, the more the cross-correlation increases and the energy decreases. But for $\alpha \to \infty$, hidden and visible unit values become binary which makes the model unable to reconstruct data.

The learning step of the cRBM in Eq.6 captures successfully the frequencies of the signal. Unfortunately, the phase ψ of the visible layer with vector W_i changes when we shift the visible layer in time : W_i and $\boldsymbol{v}$ can be correlated ($\psi = 0$), non-correlated ($|\psi| = \pi/2$) or anti-correlated ($|\psi| = \pi$). Eq.11 gives the interpretation of the variation of h_i in function of ψ :

$$\begin{cases} h_i \to +\theta & \text{if} \quad \psi \to 0 \\ h_i \to -\theta & \text{if} \quad |\psi| \to \pi \\ h_i \to 0 & \text{if} \quad |\psi| \to \pi/2 \end{cases} \tag{11}$$

For a given short time window $\boldsymbol{v}$, each hidden units h_i gives an expected image of the correlation between the visible layer and the $i-$th line of the transfer matrix $\Gamma_{W_i \boldsymbol{v}}(0)$. The product $h_i \Gamma_{W_i \boldsymbol{v}}(0)$ tends to stay positive and to decrease the energy according to Eq.9. In the non-correlated case ($\Gamma_{W_i \boldsymbol{v}}(0) = 0$), the energy component of the $i-$th hidden unit is set back to zero. Fig.4 gives the evolution in time of the energy term $\boldsymbol{h}^T W \boldsymbol{v}$ for a cRBM trained with a sinusoidal signal. The upper plot is the original signal, the three last ones are the $\boldsymbol{h}^T W \boldsymbol{v}$ terms with, resp. $1 \leq n \leq 3$ hidden units. We note for a cRBM with two or three hidden units, all hidden neurons capture the same frequency but with a delay of $\pi/2$ and $\pi/3$ for the cRBM with resp. two and then three hidden units. The introduced delay between hidden units allows the cRBM to keep the energy as stable as possible.

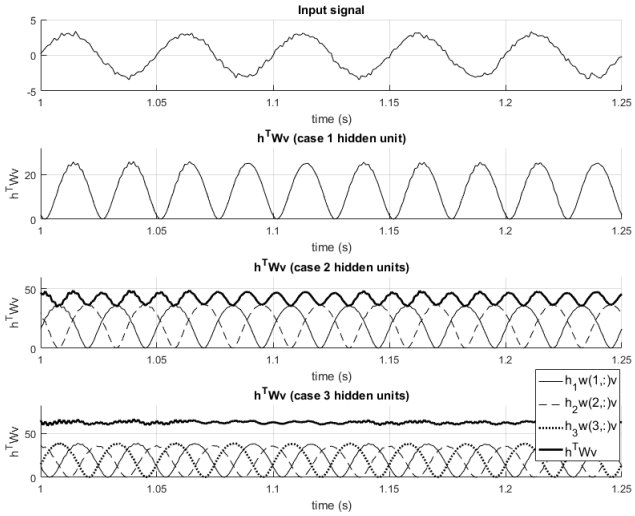


FIGURE 4 – Variation of the term $\boldsymbol{h}^T W \boldsymbol{v}$ in function of the time for a 50Hz signal. Results for three cRBMs with respectively 1,2 and 3 hidden unit(s).

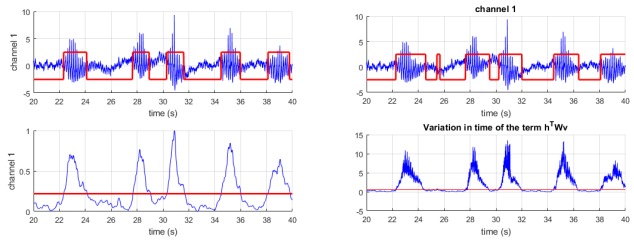
Once the cRBM is trained, weights of the transfer matrix capture the *principle components* of signals. Multiple hidden units allow to reconstruct learned frequencies with the corresponding phase and intensity.

3 Experiment

In this section, we use the cRBM model to detect the HVS in LFPs. The process of data extraction is described in Vigneron *et al.* [8]. The LFPs were recorded from 4 or 8 different brain regions. Several sessions of 60s with 1 kHz sampling rate were recorded on PD rats. The data were then standardized by subtracting the mean and dividing by the standard deviation of each channel.

We defined a *ground truth* only for the result evaluation. The presence of HVS is characterized by a burst of *spike-and-wave* patterns with a fundamental frequency between 5 and 13 Hz. The cosine wavelet transform (CWT) is computed for each channel. HVS is detected from the sum of the CWT coefficients between 5 and 13 Hz (see Fig.5a) and it is simultaneously present on at least ¾ of all channels.

The first stage consists to train the cRBM with 6 hidden units : at each iteration, a short time window is randomly selected to update the parameters. Choosing a 200 ms windows with 20 ms between each observation as visible units of the cRBM is sufficient to capture the fundamental frequency and first's harmonics of the HVS. The non-stationary offsets are removed by centering each visible neurons. Once the model is trained, we



(a) Ground truth. (b) cRBM classification.

FIGURE 5 – HVS detection (zoom 20s-40s). *Top*) First channel signal is represented in blue on the top figures. *Bottom*) The sum over 5-13 Hz band of the cosine wavelet transform coefficients is plotted in blue on the bottom left while the evolution term $\boldsymbol{h}^T W \boldsymbol{v}$ in time is given on the bottom right in blue. The red horizontal lines are thresholds for the detection of HVS defined by Otsu's method.

compute the $\boldsymbol{h}^T W \boldsymbol{v}$ term (Fig.5b) and apply Otsu's method to define a threshold. On the testing set, we use a *hysteresis filter* to cancel the noise during the transition between the two states of the signal ($\pm 50\%$ of the threshold). This method is applied to all PD rats and provides the results given in Tab.1.

Data ID.	C	N	Sensitivity	Specificity	Delay (ms)
1	8	12	0.99	0.81	-45
2	4	20	0.99	0.46	-96
3	4	9	0.85	0.79	25
4	4	6	0.97	0.80	-44
5	4	3	0.99	0.97	-9
6	4	3	0.92	0.97	101
7	4	17	0.91	0.66	-4
8	4	9	0.98	0.85	-28
9	4	3	0.98	0.87	-250
10	4	2	0.99	0.99	108
11	4	3	0.95	0.93	21
12	4	3	1.00	0.49	-447
13	4	1	1.00	0.97	-12

TABLE 1 – Results of multiple testing for different PD rats. C is the number of channels and N the number of HVS detected (by the ground truth). For each rat, we trained 10 cRBM and we compute for each test the *sensitivity* and the *specificity* and the mean delay of detection of the cRBM model compared with the ground truth method. The mean over each test has been reported in the table. A negative value on the delay means the cRBM model detect the HVS before the ground truth. The value of the delay is not always significant because for some PD rats, the noise makes the ground truth unsustainable, in particular for PD rats 12 and 13.

4 Discussion

The cRBM successfully learns a representation of deep cortical signals. For most rats the sensitivity of the detection is close to 1, *i.e.* we (almost) never miss the HVS pattern. For some patient, cRBM model successfully detects HVS *before the ground truth*; the decrease of the specificity means a faster detection of the HVS. The quality of data can be very different from a rat to an other due to the signal-to-noise ratio per channel or the presence of HVS. Our ground truth definition may be too simple to evaluate properly the models. A comparison with more robust methods like in [7] may be fruitful. Properties of the cRBM provide various advantages and possible approaches of improvement are currently available. The cRBM is an unsupervised generative model capable to learn optimal frequencies to be detected and can be used as a predictor. The cRBM can extract non-correlated components but the separability depends on the data. A study of the architecture of the model is required to help hidden units to extract different components. To remove the time lag introduced by the use of an observation window, working directly on the Diffusion Network is another possible path of improvement.

Références

[1] Chen *et al.* Continuous restricted boltzmann machine with an implementable training algorithm. *IEE Proceedings-Vision, Image and Signal Processing*, 150(3) :153–158, 2003.

[2] Chen *et al.* Continuous-valued probabilistic behavior in a vlsi generative model. *IEEE Transactions on Neural Networks*, 17(3) :755–770, 2006.

[3] Dejean *et al.* Dynamic changes in the cortex-basal ganglia network after dopamine depletion in the rat. *Journal of neurophysiology*, 100(1) :385–396, 2008.

[4] Hopfield *et al.* Computing with neural circuits : A model. *Science*, 233(4764) :625–633, 1986.

[5] Movellan *et al.* A monte carlo em approach for partially observable diffusion processes : Theory and applications to neural networks. *Neural computation*, 14(7) :1507–1544, 2002.

[6] Souriau *et al.* Probit latent variables estimation for a gaussian process classifier : Application to the detection of high-voltage spindles. In *International Conference on Latent Variable Analysis and Signal Separation*, pages 514–523. Springer, 2018.

[7] Souriau *et al.* High-voltage spindles detection from eeg signals using recursive synchrosqueezing transform. In *Proc. GRETSI'19 (submitted)*, August 2019.

[8] Vigneron *et al.* Automatic detection of high-voltage spindles for parkinson's disease. In *Proceedings of 4th Multivariable Processing for Biometric Systems*, Lisbon, Portugal, Jan. 2015.

[9] Hinton. A practical guide to training restricted boltzmann machines. In *Neural networks : Tricks of the trade*, pages 599–619. Springer, 2012.