Covariance Phasor Neural Network as a Mean field model

Haruhisa Takahashi

Department of Infromation and Communication Engineering
The University of Electro-Communications, Chofu, Tokyo, 182 Japan
takahasi@ice.uec.ac.jp

Abstract

We present a phase covariance model that can well represent stimulus intensity as well as feature binding (i.e., covariance). The model is represented by complex neural equations, which is a mean field model of stochastic neural model such as Boltzman machine and sigmoid belief networks.

1 Introduction

We consider here complex avtivation neural networks called covariance phasor networks which allow us to implement covariance Hebbian learning as well as to represent correlation of firing. The covariance phasor network is a mean field model of stochastic neural netowrks such as Boltzmann machine and sigmoid belief network. Unlike the ordinal mean field methods, the covariance phasor network can represent covariance between units. This enables us to apply it as a convinient tool for graphical models. In this report the model is established and numerically examinined as a deterministic model of stochasticneural networks.

2 Covariance model

2.1 Phasor representation of correlation

We introduce a phasor representation for culculating covariance, which is convenient for neural network analysis. The set of zero-mean (and finitely deviated) random processes on e.g. R^1 are considered as a vector space with inner product:

$$< X, Y >= \int_A xydP(x,y).$$

The angle θ between X and Y represents correlation. More precisely, the correlation coefficients are given by the cosine as

$$\cos\theta = \rho_{XY} = \frac{< X, Y >}{\sqrt{< X, X >< Y, Y >}}.$$

We now consider the linear combination of the random processes

$$Y = \sum_{i=1}^{n} a_i X_i. \qquad (1)$$

From

$$\begin{aligned} Cov(Y,Y) &= \sum_{i=1}^{n} a_i < Y, X_i > \qquad (2) \\ &= \sum_{i=1}^{n} a_i Cov(Y, X_i), \end{aligned}$$

the following holds

$$\sigma_Y \rho_{YY} = \sum_{i=1}^{n} a_i \sigma_i \rho_{YX_i},$$

where $\sigma_Y^2 = Cov(Y,Y)$, $\sigma_i^2 = Cov(X_i, X_i)$, and ρ_{YX_i} repersents the coefficient of deviation.

Let us represent each random process by phase θ and standard deviation σ as a phasor $\sigma \exp(\mathbf{i}\theta)$. Note that the value of θ itself makes no sense solely, but when phasors are multiplied as the inner product the real part represents covariance between corresponding processes. Correponding to eq.(1),(2),

$$\begin{aligned} z &= \sigma_Y \exp(\mathbf{i}\theta) \\ &= \sum_{i=1}^{n} a_i \sigma_i \exp(\mathbf{i}\theta_i), \qquad (3) \end{aligned}$$

from which

$$\sigma_Y = \sum_{i=1}^{n} a_i \sigma_i \exp(\mathbf{i}(\theta_i - \theta)) = \sum_{i=1}^{n} a_i \sigma_i \cos(\theta_i - \theta). \quad (4)$$

The phase θ can be determined from the imaginary part as

$$\sum_{i=1}^{n} a_i \sigma_i \sin(\theta_i - \theta) = 0.$$

The following property holds.

Property 1 *Let X and Y be related with $Y = aX + b$, and let their phasor representations be $\sigma_X \exp(\mathbf{i}\theta_X)$ and $\sigma_Y \exp(\mathbf{i}\theta_Y)$, respectively. Then $\theta_Y = \theta_X$ and $\sigma_Y = a\sigma_X$.*

We construct the phasor network in the next section.

2.2 Phase equation

We assume that each neuron behaves as an on-off random process $\{s_i\}$. The probability of $s_i = 1$ is

$$p_i = f(\sum_j w_{ij} s_j + b_i), \quad (5)$$

where f is the logistic function. Phasor network is a neural network which accounts for pulse timing as well as correlation. The fire timing of each neuron i is represented by a phase $-\pi \leq \phi_i < \pi$, and the mean firing rate (probability) is represented by $0 \leq r_i \leq 1$. According to the discussion in the previous subsection, cosine of the phase difference between two neurons can be interpreted as representing correlation. To apply the phasor analysis for neural networks we need to estimate the phasor of output process $\{s_i\}$. The difficulty to perform this is twofold; it is due to nonlinearity of logistic function and random firing with the rate in eq.(5) .

To provide the phasor analysis for neural networks, we characterize each neuron by a complex variable $\sigma_i \exp(\mathbf{i}\phi_i)$, where $\mathbf{i} = \sqrt{-1}$. According to eq.(3) we let each process be governed by the next phase equations:

$$\frac{dz_j}{dt} = -z_j + \sum_{i=1}^{n} w_{ji} \sigma_i \exp(\mathbf{i}\phi_i), \quad (6)$$

where $z_j = \tilde{\sigma}_j \exp(\mathbf{i}\tilde{\phi}_j)$ represents a phasor for

$$u_j = \sum_i w_{ji} s_i + b_j,$$

and σ_i is simply given by the standard deviation of on-off random process $\{s_i\}$ with mean r_i as

$$\sigma_i = \sqrt{r_i(1 - r_i)}. \quad (7)$$

We now wish to obtain the phase output equation for $\{s_i\}$ taransforming $\tilde{\phi}_j$ to ϕ_j. Let a phasor representation for $\{p_j\}$ be $\hat{\sigma}_j \exp(\mathbf{i}\hat{\phi}_j)$.

Property 2 *Let X be a random process, and let $\hat{X} = X - \mathbf{E}\{X\}$. Then for $i \neq j$*

$$< \hat{s}_i, \hat{s}_j > \; = \; < \hat{s}_i, \hat{p}_j > \; = \; < \hat{p}_i, \hat{p}_j >$$

and for $i = j$

$$< \hat{s}_j, \hat{p}_j > = \hat{\sigma}_j^2 \leq \sigma_j^2.$$

Thus for $i \neq j$

$$\sigma_j \cos(\phi_i - \phi_j) = \hat{\sigma}_j \cos(\phi_i - \hat{\phi}_j), \quad (8)$$

$$\sigma_i \cos(\phi_i - \hat{\phi}_j) = \hat{\sigma}_i \cos(\hat{\phi}_i - \hat{\phi}_j). \quad (9)$$

Using $\tilde{\phi}_j = \hat{\phi}_j$ (see Property 1), $z_j = \tilde{\sigma}_j \exp(\mathbf{i}\tilde{\phi}_j)$ may be replaced by

$$z_j = \tilde{\sigma}_j \exp(\mathbf{i}\hat{\phi}_j). \quad (10)$$

The corresponding phase equation is

$$\frac{dz_j}{dt} = -z_j + \sum_{i=1}^{n} w_{ji} \hat{\sigma}_i \exp(\mathbf{i}\hat{\phi}_i). \quad (11)$$

Property 3 *Let $r_j = \mathbf{E}\{s_j\}$. Assume that $\tilde{\sigma}_j$ is not so large that the sigmoid function can be linearly approximated around $f^{-1}(r_i)$. Then*

$$\hat{\sigma}_j \approx \alpha \sigma_j^2 \sqrt{\tilde{\sigma}_j},$$

where $\sigma_j = \sqrt{r_j(1 - r_j)}$, and $\alpha > 0$ is taken empirically as $\alpha = 1.0$. [1]

[1] Since the approximatin is first order, it cannnot be assured for large $\tilde{\sigma}_j$. Taking $\hat{\sigma}_j < \sigma_j$ into account, we can use a huristic modification of $\hat{\sigma}_j = \alpha \sigma_j^2 \tanh(\sqrt{\tilde{\sigma}_j})$.

3 Rate equation

The phase equation discussed in the previous section represents the correlation among firing activation on the units. However to complete eqs.(6) and (10) we need to evaluate r_i. We will develop a mean field approach to approximately culculating the firing rate in probabilistic neural networks.

3.1 Probabilistic neural networks

The energy function of Boltzmann machine is defined by

$$H_\alpha = -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij}s_i s_j - \sum_{i=1}^{N} b_i s_i. \quad (12)$$

3.2 Mean field approximation

Replacing $< \hat{s}_i, \hat{s}_j > = < s_i s_j > - r_i r_j = < \hat{p}_i, \hat{p}_j >$ in the avarage of Eq.(12) by $\hat{\sigma}_i\hat{\sigma}_j\cos(\hat{\phi}_i - \hat{\phi}_j)$, and $< s_i >$ by r_i, we obtain a phasor covariance energy representation.

$$\begin{aligned} \mathbf{E}\{H_\alpha\} &= -\frac{1}{2}\sum_{ij=1}^{n} w_{ji}\{r_i r_j + \hat{\sigma}_i\hat{\sigma}_j\cos(\hat{\phi}_i - \hat{\phi}_j)\} \\ &\quad -\sum_{i=1}^{n} b_i r_i. \end{aligned}$$

We seek an optimal equation for $r_i = < s_i >$ by minimizing the KL-divergence between P_α and the marginal distribution $Q_\alpha = \Pi_i(s_i r_i + (1-s_i)(1-r_i))$,

$$\begin{aligned} KL(P_\alpha \| Q_\alpha) &= \sum_\alpha P_\alpha \ln\frac{P_\alpha}{Q_\alpha} \\ &= -\sum_{i=1}^{n} r_i \ln r_i + (1-r_i)\ln(1-r_i) \\ &\quad -\mathbf{E}\{H_\alpha\} - \ln Z, \quad (13) \end{aligned}$$

with respet to the marginal probabilities r_i. Setting the gradient of eq.(13) equal to zero, we obtain

$$r_j = f\left(-\frac{\partial \mathbf{E}\{H_\alpha\}}{\partial r_j}\right),$$

which is an equilibrium of the gradient dynamical system:

$$\frac{du_j}{dt} = -u_j + \sum_{i=1}^{n} w_{ji}\{r_i + \frac{\partial\hat{\sigma}_j}{\partial r_j}\hat{\sigma}_i\cos(\hat{\phi}_i - \hat{\phi}_j)\} + b_j, \quad (14)$$

From Property 3

$$\frac{\partial\hat{\sigma}_j}{\partial r_j} = \alpha\frac{(1-2r_j)}{2\sigma_j}\sqrt{\hat{\sigma}_j},$$

and we reach

$$\begin{aligned} \frac{du_j}{dt} &= -u_j + \sum_{i=1}^{n} w_{ji}\{r_i + \alpha\frac{(1-2r_j)}{\sigma_j} \cdot \\ &\quad \sqrt{\hat{\sigma}_j}\hat{\sigma}_i\cos(\hat{\phi}_i - \hat{\phi}_j)\} + b_j. \quad (15) \end{aligned}$$

Finally it can be also shown that $\mathbf{E}\{H_\alpha\}$ decreases according to eq.(6).

Property 4 *Let*

$$E = \Re[-\frac{1}{2}\sum_{ij=1}^{n} w_{ij}\hat{\sigma}_i\exp(\mathbf{i}\phi_i)\hat{\sigma}_j\exp(-\mathbf{i}\phi_j)].$$

Then E decreases according to eq.(11). Thus, $\mathbf{E}\{H_\alpha\}$ decreases according to eq.(11).

4 Mean field Boltzmann learning

Correponding to the Boltzmann Machine learning, we obtain

$$\begin{aligned} \Delta w_{ij} &= \eta\{r_i^+ r_j^+ + \hat{\sigma}_i^+\hat{\sigma}_j^+\cos(\hat{\phi}_i^+ - \hat{\phi}_j^+) \\ &- [r_i^- r_j^- + \hat{\sigma}_i^-\hat{\sigma}_j^- \cdot \\ &\quad \cos(\hat{\phi}_i^- - \hat{\phi}_j^-)]\}, \quad (16) \end{aligned}$$

where $\eta > 0$.

5 Numerical results

We compared a covariance phasor network with a mean field network. Since these are deterministic approximation of a stochastic network, a Boltzman machine with four units is calculated for the bench mark probability distribution.

The weights are chosen randomly from $[0, 1]$ and multiplied by a coefficient β. which moves from 0 to 4.0. Figure 1 compares the Kullback-Leibler distance between Boltzman machine and deterministic models: covariance model and a variational mean field model. Figure 2 shows the ratio of covariance coefficients in covariance model v.s. BM.

Finally we performed an experiment on learning ability of a Phasor neural network. The learning target is a simple two states Markov processes which can

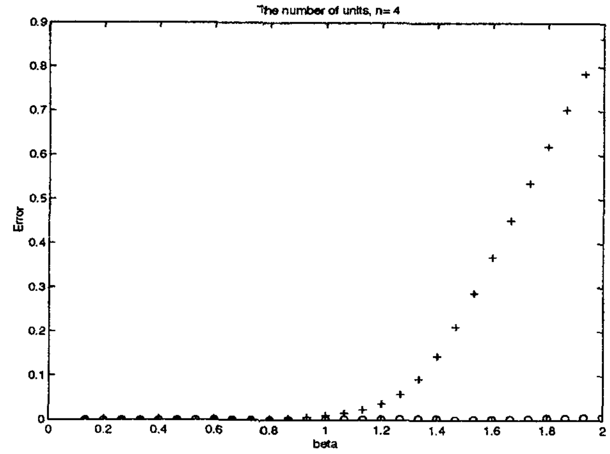

Figure 1: Kullback-Leibler distance of Marginal probability distribution: + : Mean field; o : Covariance.

be switched according to inputs. The network architecture is tow outputs and six hidden units. The network could learn the target as shown in Figure 3.

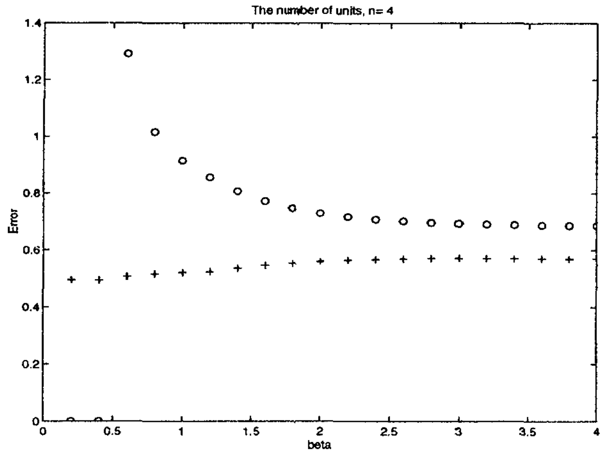

Figure 2: Ratio of correlation coefficients of Covariance netowirk: + : Units1-4; o : Unit1-2; when the ratio is one, the approximation is correct.

Acknowledgments

This work was supported in part by Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

References

[1] M.C. Mozer, et.al., Learning to segment images using dynamic feature binding, Neural Computation 4,650-665 (1992).

[2] D. Wang, D. Terman, Image segmentation based on oscillatory correlation, Neural Computation 9, 805-836 (1997).

[3] M. Takeda, T. Kishigami, Complex neural fields with a Hopfield-like energy function and an analogy to optical fields generated in phase-conjugate resonators, J. Optical Soc. of America A., pp 2182-2191 (1992).

[4] H.J.Kappen, W.J. Wiegerinck, Mean field theory for graphical models, in Advanced Mean Field Methods-Theory and Practice,ed. M. Opper, D. Saad, MIT Press (2001).

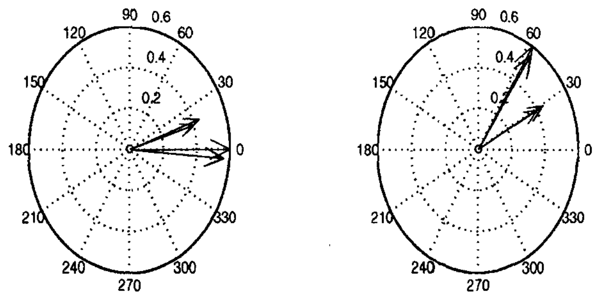

Figure 3: Target vectors are represented as two complex variables of poler representation correponding to two outputs. They are learned using learning rule of eq.(15).