Boltzmann Learning of Parameters in Cellular Neural Networks [1]

Lars Kai Hansen
CONNECT, Electronics Institute, build. 349
Technical University of Denmark, DK-2800 Lyngby, Denmark
Tlf.: +45 45 93 12 22 ext. 3889, Fax: +45 42 88 01 17
Email: lars@eiffel.ei.dth.dk

We use Bayesian methods to design cellular neural networks for signal processing tasks and the Boltzmann Machine learning rule for parameter estimation. The learning rule can be used for models with "hidden" units, or for completely unsupervised learning. The latter is exemplified by unsupervised adaptation of an image segmentation cellular network, in particular we apply the learning rule to adaptive segmentation of satellite imagery.

1 Introduction

The Bayesian or *Maximum Posterior* approach is a very successful device for signal processing [15, 7, 11], a particular attraction is that it leads to algorithms that map well onto networks of locally connected, simple processing elements ie. cellular neural networks [4]. With the advent of low-cost massively parallel hardware, this virtue may end up being decisive for (near) future real life applications. However, while the Bayesian approach allows formulation of collective models that solve complex signal processing tasks, general and flexible tools for parameter estimation are lacking. Direct maximum-likelihood estimation is hampered by the difficulty of obtaining analytical expressions for derivatives of normalization constants [1]. Besag has introduced two methods based on approximate maximum-likelihood estimation, the *coding* method [1], and the *pseudo-likelihood* method [2]. Both methods pose difficulties if the image model includes non-observable attributes ie. *hidden* units. In this contribution we discuss the use of the Boltzmann Machine learning rule [10] for parameter estimation. The learning rule may be applied to general situations with hidden units without complication. We address in this presentation *unsupervised* learning.

Since the phase-space distribution of a dynamical system in contact with a heat-bath is a Gibbs distribution; sampling from such distributions have been central to simulations of statistical physics systems. The standard simulation tool is due to Metropolis et al. [13]. Geman and Geman [7] introduced Metropolis sampling from Gibbs distributions as a simulation tool for visual reconstruction and showed that a *Simulated Annealing* strategy [12] could improve the speed of the sampling process. The sampling process implements a stochastic neural network with symmetric connections. Hinton and Sejnowski [10] studied supervised learning in such networks and introduced the term *Boltzmann Machines* to emphasize the relation to statistical physics. Invoking the *Mean Field* approximation, Peterson and Anderson derived a deterministic selfconsistent set of equations for the time-averages of the dynamical variables. Using these

[1] *Proceedings of the Second Int. Workshop on Cellular Neural Networks and their Applications CNNA'92, Munich, October 1992*, pp 62-67. Ed.: J.A. Nossek. IEEE Service Center Piscataway, NJ, USA (1992).

averages in the learning rule, they obtained substancial improvements in speed and performance [14].

In the next section the Bayesian approach to signal processing is outlined, and we discuss the design of cellular networks using the Mean Field approach. In section three we show how the Boltzmann Machine learning rule may be applied for parameter adaptation. Section four contains experiments on image segmentation and concluding remarks.

2 Bayesian Signal Processing

The Bayes approach to signal processing has a long tradition, see e.g. [11, 7], or [5] for a recent introduction. The basic idea is to consider both the source (un-degraded) signal and the degradation as stochastic processes. The Bayes formula can then be used to construct the distribution of the reconstructed signal (x), conditioned on the observed degraded signal (y):

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{1}$$

According to the standard interpretation the conditional distribution is the product of the distribution of the degradation process: $P(y|x) \equiv P(x \rightarrow y)$, and the *prior* distribution of the reconstructed signal $P(x)$. $P(x|y)$ of equation (1) is referred to as the *posterior* distribution. A useful estimate of the recontructed signal is given by the location of the mode of the posterior distribution, the socalled *Maximum A Posteori* (MAP) estimate. In the following we derive the posterior distributions for an image segmentation model.

2.1 Image Segmentation

Segmentation is an important step in many computer vision systems, however even in its simplest form: binarization of a grey-scale image there exist no established standard solution!. Here we use the Bayes scheme to derive a simple cost-function that can be minimized by a cellular neural network. The resulting cost-function is identical to the one used by Carnevali *et al.* [3]. The target signal is a smooth binarization of a grey-scale image d_j, in terms of two-valued pixels $S_j \in \{-1, +1\}$. The prior distribution is designed to emphasize smoothness:

$$P[S] = Z_1^{-1} \exp\left(- \sum_{j=1}^{N} \sum_{j'=1}^{N} M(j,j')(S_j - S_{j'})^2 \right) \tag{2}$$

$M(j,j')$ defines the connectivity, hence the unit cell of the cellular network. Here we just use the nearest neighbors. More complex connectivity structures have been designed for modelling textural features [6].

We assume the signal degradation to consist in addition of white Gaussian noise. This degradation process leads to the following conditional distribution:

$$P[d|S] = Z_2^{-1} \exp\left(-\frac{1}{2\sigma^2} \sum_{j=1}^{N} (S_j - d_j)^2 \right) \tag{3}$$

We want to approach real data for which the noise variance is unknown, hence this parameter has to estimated as part of the learning process. As above we use Bayes to combine and obtain the posterior distribution. Clearly it is of the Gibbs form[2], with a cost-function given by the negative logarithm of the posterior distribution: $-\log P[S|d]$. We note that the state dependent part of the cost-function is linear in the parameters $M(j, j')$ and $w_d \equiv 1/\sigma^2$.

2.2 Network Design

The Mean Field annealing method for estimation of averages over Gibbs distributions is well documented in the litterature see e.g. Hertz *et al.* [9]. The cellular neural network is designed to minimize the Mean Field *free energy* F, and this can be done either in analog mode:

$$\tau \frac{\partial \langle S_j \rangle}{\partial t} = -\frac{\partial F}{\partial \langle S_j \rangle} = -\langle S_j \rangle + \tanh\left(\beta^t \left(\sum_{j'=1}^{N} M(j, j') \langle S_{j'} \rangle + w_d d_j \right) \right) \tag{4}$$

or in discrete time mode:

$$\langle S_j^{t+1} \rangle = (1 - \frac{\Delta}{\tau}) \langle S_j^t \rangle + \frac{\Delta}{\tau} \tanh\left(\beta^t \left(\sum_{j'=1}^{N} M(j, j') \langle S_{j'}^t \rangle + w_d dj \right) \right) \tag{5}$$

where the time-scale Δ/τ can be used to regularize the stability of the iteration process in digital implementation[14]. β^t is quantify the annealing schedule, in this work we use the simple schedule: $\beta^t = \beta^1 + (t/t_{max})(\beta^2 - \beta^1)$.

In summary, the unit cell of the cellular network contains one unit approximating the local thermodynamic average: $\langle S_j^t \rangle$, and one input unit d_j. We assume that $M(j, j')$ connects nearest neighbors symmetrically with weight w_n, and the weights to the input units are all w_d.

3 Boltzmann Machine Learning

In order to apply the cellular network above we have to estimate the parameters (denoted $w = (w_n, w_d)$). Since our network is based on the Gibbs distribution we invoke the Boltzmann Machine learning rule [10, 14, 9]. The objective of the Boltzmann Machine learning rule is to minimize the Kullback information distance between a target distribution $P_{w^*}[S|d]$, (of which the training set is a finite sample), and the distribution $P_w[S|d]$ sampled by the current (stochastic) network with parameters w.

[2] A distribution of the form $P(x) = Z^{-1} exp(-E(x)/T)$, where $E(x)$ is a cost-function, bounded from below, and T is a parameter

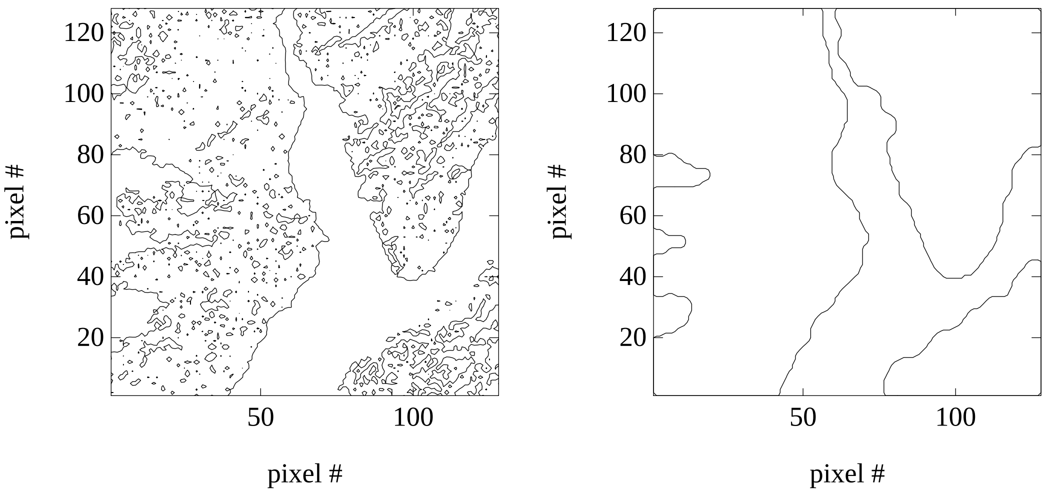

Figure 1: Subsampled, 128x128 pixel, preprocessed, Landsat image providing the raw (thresholded) evidence for water in the Igaliko region of Greenland (left), and the output evidence obtained by the unsupervised cellular neural network (right)

The learning rule is formulated for a general system specified in terms of inputs, hiddens, and outputs: (x, h, y). Since the states of the hidden units are unknown for the learning examples, we compare the marginal distributions, i.e. the distributions integrated over the hidden variables:

$$D[P_{w^*}, P_w] = \int Dx \int Dy \int DhP_{w^*}[x,h,y] \log \frac{\int DhP_{w^*}[x,h,y]}{\int DhP_w[x,h,y]} \tag{6}$$

$\int Dx$ is short for $\int \prod_j dx_j$. The learning algorithm is derived by gradient descent minimization of the information distance. The recursive learning algorithm reads:

$$w_\nu^{n+1} - w_\nu^n = \eta \left(\langle \Omega_\nu (x,h,y)) \rangle_{clamped} - \langle \Omega_\nu (x,h,y) \rangle_{free} \right) \tag{7}$$

for a any cost-function linear in the parameters w_ν: $E(x,h,y) = \sum_\nu \Omega_\nu (x,h,y) \, w_\nu$, where $\Omega_\nu (x,h,y)$ is an expression in the stochastic variables. $\langle ... \rangle_{clamped}$ indicates that the average is performed with respect to the current Gibbs distribution, with *fixed* values for all the stochastic variables that are specified as *input or output* variables in an example of the database, i.e. (x,y). Similarly $\langle ... \rangle_{free}$ is the average with only *input* variables fixed. In brief we can characterize the learning process as follows: the parameters are adjusted to minimize the difference between the correlations in situations with and without the teacher specifying the correct output [10, 9].

We can employ the above formalism for unsupervised learning if we let our output units S play the role of hidden units and partition the input image d into a Boltzmann "input" part and a Boltzmann "output" part. The Boltzmann learning process then adapts the model until we reach a parameter set for which the "output" part of the image is estimated correctly from "input" part. Our procedure can be viewed as an example of statistical *cross-validation*.

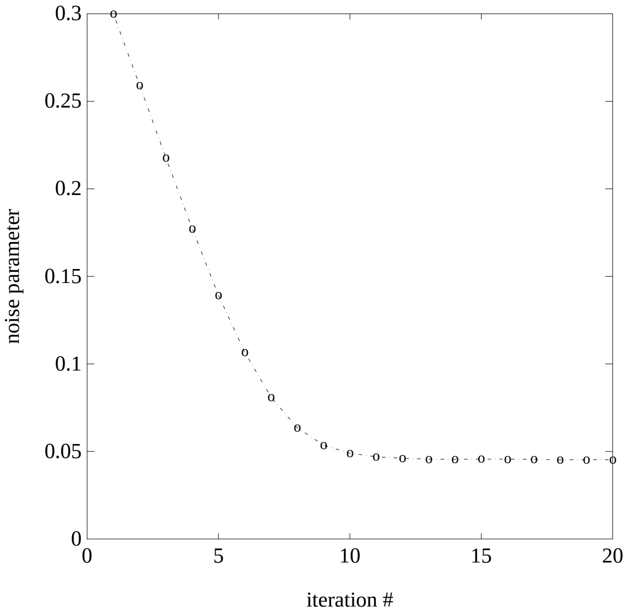

Figure 2: Unsupervised Boltzmann learning of the noise parameter of the cellular segmentation network.

4 Experimental and concluding remarks

Our case-study concerns the segmention of a subsampled Landsat satellite image. The input signal is a preprocessed $128 * 128$ pixel image representing the evidence for water in the *Igaliko* region in Greenland. The preprocessing scheme establish the evidence using four frequency bands. Part of the image has been classified manually into five classes and the evidence is the result of a simple linear model relating the intensities in the four bands to the classification. Thresholding the evidence at zero results in Figure 1a). We adapt the noise parameter of the cellular network in unsupervised mode using a gradient descent parameter of 0.15. In Figure 1b) we show the segmentation of the adapted network. The convergence of the noise-parameter w_d towards the selfconsistent optimal value is presented in Figure 2.

In conclusion we have shown that the Boltzmann Machine learning rule can be used for identification of parameters in cellular neural networks designed using Bayesian reasoning. By invoking a crossvalidation-like procedure we were able to adapt the parameters of the cellular network without supervision, hence generalizing our earlier results on parameter estimation in cellular networks with hidden units [8].

Acknowledgement

This research is supported by the Danish Research Councils for the Natural and Technical Sciences through the Danish Computational Neural Network Center. I thank Bjarne K. Ersbøll for providing the Igaliko imagery.

References

[1] J. Besag: *Spatial Interaction and the Statistical Analysis of Lattice Systems.* J.Roy.Statist.Soc, **B36**, 192-236, (1974).

[2] J. Besag: *On the Statistical Analysis of Dirty Pictures.* J.Roy.Statist.Soc, **B48**, 259-302, (1986).

[3] P. Carnevali, L. Coletti and S. Paternello: *Image processing by simulated annealing.* IBM Journal of Research and Development **29**, 569-579, (1985).

[4] L.O. Chua and L. Yang: *Cellular Neural Networks: Theory.* IEEE Transactions on Circuits and Systems **35**, 1257-1272, (1988).

[5] J.B. Cole: *The Statistical Mechanics of Image Recovery and Pattern Recognition.* Am.J.Phys. **59**, 839-42, (1991).

[6] G.R. Cross and A.K. Jain: *Markov Random Field Texture Models.* IEEE Transactions on Pattern Analysis and Machine Intelligence **5**, 25-39, (1983).

[7] D. Geman and S. Geman: *Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images.* IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-6**, 721-741, (1984).

[8] L.K. Hansen: *Boltzmann Learning of Parameters in Bayes Visual Reconstruction.* Proceedings of the First Danish Conference on Pattern Recognition and Image Analysis. Ed.: S.I.Olsen. Department of Computer Science, University of Copenhagen, 92/8, (1992).

[9] J. Hertz, A. Krogh and R.G. Palmer: *Introduction to the Theory of Neural Computation.* Addison Wesley, New York (1991).

[10] G.E. Hinton and T.J. Sejnowski : *Learning and Relearning in Boltzmann Machines.* In Parallel Distributed Processing, vol. 1, chap 7. Eds. Rumelhart D.E. et al. Cambridge: MIT Press.

[11] B.R. Hunt: *Bayesian Methods in Nonlinear Digital Image Restoration.* IEEE Transactions on Computers **C-76**, 219-229, (1977).

[12] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi: *Optimization by Simulated Annealing.* Science **220**, 671-680, (1983).

[13] N. Metropolis et al.: *Equation of State Calculations by Fast Computing Machines.* J. Chem. Phys. **21**, 1087-1099 (1953).

[14] C. Peterson and J.R. Anderson: *A Mean Field Learning Algorithm for Neural Networks.* Complex Systems **1** 995-1019, (1987).

[15] M.W. Roth: *Survey of Neural Network Technology for Automatic Target Recognition.* IEEE Transactions on Neural Networks **1**, 28-43, (1990).