Bayesian Inference in densely connected networks applied to CDMA

Juan P. Neirotti & David Saad

The Neural Computing Research Group, Aston University, Birmingham, B4 7ET, UK

Graphical models provide a powerful framework for modelling statistical dependencies between variables [1]. Message passing techniques are typically used for inference in graphical models that can be represented by a sparse graph. Iterative message passing is guaranteed to converge to the globally correct estimate when the system is tree-like; there are no such guarantees for systems with loops.

Two inherent limitations seem to prevent the use of message passing techniques in densely connected systems: 1) Their high connectivity implies an exponentially growing computational cost. 2) The existence of an exponential number of loops that render the method inconsistent. However, a new approach was suggested [2] for extending Belief Propagation (BP) techniques to densely connected systems. In this approach, messages are grouped together, giving rise to macroscopic random variables drawn from a different Gaussian distribution of varying mean and variance for each of the nodes.

In a separate development [3], BP was extended to Survey Propagation (SP). This new algorithm has succeeded in solving hard computational problems [3], far beyond other existing approaches.

Inspired by the extension of BP to SP we have extended the approach of [2], designed for inference in densely connected systems, in a similar manner by including an average over multiple pure states. However, for highlighting the advantages with respect to the original method [2], we apply it to the problem of signal detection in CDMA.

Code Division Multiple Access [4] is based on spreading the signal by using K individual random binary spreading codes of spreading factor N. We consider the large-system limit $N \to \infty$, $K \to \infty$ with $\beta = K/N \sim \mathcal{O}(1)$. The received aggregated, modulated and corrupted signal is of the form:

$$y_\mu = \frac{1}{\sqrt{N}} \sum_{k=1}^{K} s_{\mu k} b_k + \sigma_0 n_\mu$$

where b_k is the bit transmitted by user k, $s_{\mu k}$ is the binary spreading chip value, n_μ is the Gaussian noise variable drawn from $\mathcal{N}(0,1)$, and y_μ the received message. The goal is to get an accurate estimate of the vector $\mathbf{b}$ for all users given the received message vector $\mathbf{y}$ by approximating the posterior $P(\mathbf{b}|\mathbf{y})$.

A solution can be obtained by averaging over the various solutions, inferred from the same data, in a similar manner to the SP approach. Meanwhile, the messages in the current case are more complex.

Using Bayes rule one obtains the BP equations:

$$\begin{aligned} P^{t+1}(y_\mu|\mathbf{b}_k,\{y_{\nu\neq\mu}\}) &\propto \sum_{\mathbf{b}_{l\neq k}} P\left(y_\mu|\mathbb{B}\right) \prod_{l\neq k} P^t\left(\mathbf{b}_l|\left\{y_{\nu\neq\mu}\right\}\right) \\ P^t\left(\mathbf{b}_l|\left\{y_{\nu\neq\mu}\right\}\right) &\propto \prod_{\nu\neq\mu} P^t\left(y_\nu|\mathbf{b}_l\,,\left\{y_{\sigma\neq\nu}\right\}\right) . \end{aligned} \tag{1}$$

An explicit expression for the likelihood is required for deriving the posterior

$$P\left(\mathbb{B}\,|\,\mathbf{y}\right) = \frac{\prod_{\mu=1}^{N} P\left(y_\mu\,|\,\mathbb{B}\right)}{\mathsf{Tr}_{\{\mathbb{B}\}} \prod_{\mu=1}^{N} P\left(y_\mu\,|\,\mathbb{B}\right)} . \tag{2}$$

The latter is derived from the noise model (assuming zero mean and variance σ^2)

$$P\left(y_\mu\,|\,\mathbb{B}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\left(\mathbf{y}_\mu - \boldsymbol{\Delta}_\mu\right)^{\mathsf{T}} \mathbb{I}\left(\mathbf{y}_\mu - \boldsymbol{\Delta}_\mu\right)}{2\sigma^2}\right\}, \tag{3}$$

where $\mathbf{y}_\mu = y_\mu \mathbf{u}$, $\mathbf{u}^{\mathsf{T}} \equiv \overbrace{(1,\, 1,\, \cdots,\, 1)}^{n}$ and $\boldsymbol{\Delta}_\mu \equiv \frac{1}{\sqrt{N}} \sum_{k=1}^{K} s_{\mu k} \mathbf{b}_k$. Understanding the correlation between the replicated solutions is at the heart of the new approach. An explicit expression for the statistical dependence between solutions is required for obtaining a closed set of update equations. We assume a dependence of the form

$$P^t\left(\mathbf{b}_k \mid \{y_{\nu\neq\mu}\}\right) \propto \exp\left\{h^t_{\mu k}\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k + \frac{1}{2n} g^t_{\mu k}\left(\sum_{\mathrm{a}=1}^{n} b^{\mathrm{a}}_k\right)^2\right\}\ , \tag{4}$$

Using Eqs.(1), one can then calculate the expected value of b^{a}_k

$$\widehat{m}^{t+1}_{\mu k} = \left(\sigma^2+\beta\left(1-Q^t\right)+\beta R^t\right)^{-1}\left(\frac{y_\mu \mathbf{s}_\mu}{\sqrt{N}} - \beta\left(\mathbb{P}_\mu - \mathbb{I}/K\right)\mathbf{m}^t_\mu\right)_k \tag{5}$$

$$m^t_{\mu k} = \tanh\left(\sum_{\nu\neq\mu} \mathrm{artanh}\left(\widehat{m}^t_{\nu k}\right)\right) \qquad m^t_k \simeq \tanh\left(\sum_{\mu=1}^{N}\widehat{m}^t_{\mu k}\right) \tag{6}$$

where $\mathbb{P}_\mu \equiv (1/K)s_{\mu k}s_{\mu l}$, $\mathbb{I} \equiv \delta_{kl}$, $m^t_{\mu k}$ are the messages at time t from b nodes to y nodes and $\widehat{m}^t_{\mu k}$ are the messages at time t from y nodes to b nodes, respectively. The variables Q^t and R^t are related to the diagonal and off-diagonal elements of the covariance matrix of the macroscopic messages $\mathbf{\Delta}_\mu$.

The main difference between equation (5) and the equivalent equation in [2] is the emergence of an extra term in the prefactor, βR^t, reflecting correlations between different solutions groups (replica). This extra degree of freedom can be used to minimise the bit error probability at each time step. The bit error probability is the mean value of the discrepancy between the true bits sent (b_k) and the estimates ($\mathrm{sgn}\left(m^t_k\right)$). When the bit error probability is optimised with respect to R^t one obtains straightforwardly that $R^t = (\sigma_0^2 - \sigma^2)/\beta$. If the noise estimate is identical to the true noise, the term vanishes and one retrieves the expression of [2]; otherwise, an estimate of the difference between the two noise values is required for computing the prefactor of Eq. (5). Using the received signal to calculate the variance of the noise we can rewrite the update equation for $\widehat{m}_{\mu k}$, equation (5), as

$$\widehat{m}^{t+1}_{\mu k} = \left\{\frac{1}{N}\sum_{\mu=1}^{N} y_\mu^2 - \beta Q^t\right\}^{-1}\left(\frac{y_\mu \mathbf{s}_\mu}{\sqrt{N}} - \beta\,\left(\mathbb{P}_\mu - K^{-1}\mathbb{I}\right)\,\mathbf{m}^t_\mu\right)_k \tag{7}$$

where no estimate on σ_0 *is required.*

This transforms the inference algorithm into a highly practical technique as it obviates the need for a prior belief of the noise level. The inference algorithm merely requires an iterative update of equations (7,6) and converges to a reliable estimate of the signal. The computational complexity of the algorithm is of $\mathcal{O}(NK^2)$ (reducing back to $\mathcal{O}(K^2)$ once the noise has been estimated).

To test the performance of our algorithm we studied the CDMA signal detection problem under typical conditions. Error probability of the inferred signals has been calculated for a system load $\beta = 0.25$, where the true noise level is $\sigma_0^2 = 0.25$ and the estimated value is $\sigma^2 = 0.01$, as shown in Figure 1(a). In this scenario we expect the original algorithm [2] to fail due to the discrepancy between the two noise levels. The solid line represents the expected theoretical results, knowing the exact values of σ_0^2 and σ^2, while circles represent simulation results obtained via the suggested *practical* algorithm, where no such knowledge is assumed. The results presented are based on 10^5 trials per point and a system size $N = 2000$ and are superior to those obtained using the original algorithm [2].

Another performance measure one should consider is $D^t \equiv \frac{1}{K}\left|\mathbf{m}^t - \mathbf{m}^{t-1}\right|^2$. It provides an indication to the stability of the solutions obtained. In the inset of Figure 1(a) we see that results obtained using our algorithm show convergence to a reliable single solution in stark contrast to the results obtained by the original algorithm [2]. The physical interpretation of the difference between the two results is related to a replica symmetry breaking phenomena.

To analyse the critical properties of the system we studied the asymptotic regime of the error per bit probability P_b, for different values of σ_0^2 (Figure 1(b)). For low values of the noise variance and large values of β, P_b is a decreasing function of β^{-1}. For a given value of $\beta_C\left(\sigma_0^2\right)$ there is a

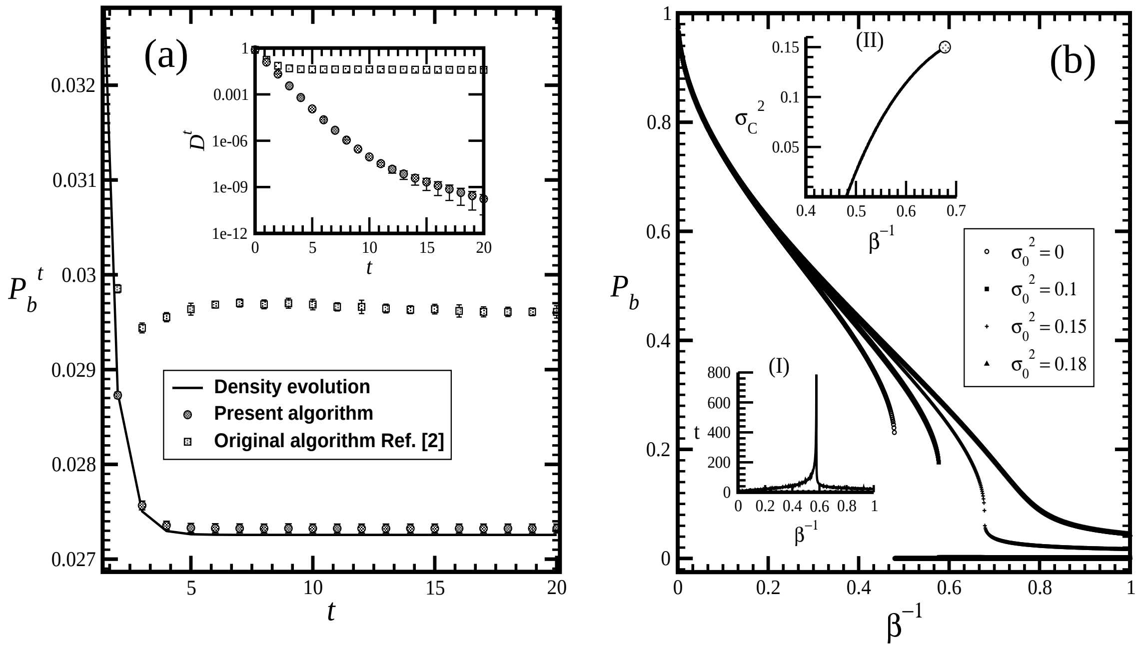

Figure 1: (a) Error probability of the inferred solution as a function of time. The system load $\beta = 0.25$, true and estimated noise levels $\sigma_0^2 = 0.25$ and $\sigma^2 = 0.01$, respectively. Squares represent results obtained by the original algorithm [2], solid line the dynamics obtained from our equations; circles represent results obtained from the suggested *practical* algorithm. Variances are smaller than the symbol size. In the insetset we present the measure of convergence D of the obtained solutions, as a function of time. (b) Error probability as a function of β^{-1} for several values of σ_0^2. Below $\sigma_0^2 \simeq 0.15$ all curves are discontinuous for a critical value β_C^{-1}. The number of iterrations needed to reach the steady state at the criticality diverges, as it is shown in the inset (I) for $\sigma_0^2 = 0.1$. For values of $\sigma_0^2 > 0.15$ all curves are analytical. The critical points $\left(\beta_c^{-1}, \sigma_0^2\right)$ are presented in the inset (II).

discontinuity in P_b. At this point, the number of iterations needed to reach the steady state diverges (Figure 1(b), inset (I)). There is a last value of $\sigma_0^2 \simeq 0.15$ for which there is a non-analytical point in P_b. At this value of the noise parameter, P_b becomes a continuous curve with a singularity in its first derivative. The critical points $\left(\beta_C^{-1}, \sigma_0^2\right)$ are presented in the inset (II) of the Figure 1(b).

Support from EVERGROW, IP No. 1935 in FET, EU FP6 is gratefully acknowledged.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann Publishers, San Francisco, CA (1988)
[2] Y. Kabashima, J. Phys. A **36** 11111 (2003)
[3] M. Mézard, G. Parisi and R. Zecchina, Science **297** 812 (2002)
[4] S. Verdú, *Multiuser Detection*, Cambridge University Press UK (1998)