A Spin-Glass model of a Markov Random Field

B. Caputo, J. Hornegger, D. Paulus and H. Niemann
Computer Science Department, Chair for Pattern Recognition,
University of Erlangen,
Martensstrasse 3, D-91058, Erlangen, Germany

In the last few years there has been a growing interest within the machine learning comunity in Spin-Glass Theory (SGT) [8] and its possible applications in learning and recognition tasks [9], [2]. SGT was first used in physics to describe magnetic materials in which the interactions between the magnetic moments (spins) are random and conflicting [8]. The attempt to understand the cooperative behaviour of such systems has led to the development of new concepts and techniques which have been finding applications and extensions in many areas such as attractor neural networks [1], combinatorial optimization problems, prebiotic evolution [8], and recently Gaussian Processes [9] and shape recognition [2]. This contribution describes a new model that makes it possible to use SGT results in a Maximum A Posteriori-Markov Random Field (MAP-MRF, [7]) framework. Many vision problems can be posed as labeling problems; labeling is also a natural representation for the study of MRFs [7]. Two major tasks in MRF modeling are how to define the neighborhood system for irregular sites, and how to choose the energy function for a proper encoding of constraints. How to define the neighbor relations between sites is related to their regularity; in the irregular case (i.e. object recognition problems, [7]), the neighborhood system must be defined by means of an "ad-hoc" distance that will be feature-dependent. If the application problem is 3-D object recognition, we have the additional problem of choosing invariant features, or we should incorporate the pose parameters in the energy formulation and in the neighbor relations definition, with a dramatical increase in complexity. The energy function is a quantitative cost measure of the quality of a solution, which defines the best solution as its minimum. In the case of irregular sites, the energy function's formulation can become something of an art, as it is generally done manually.

SGT provides a way to deal with these problems in an elegant manner: full connectivity makes the neighborhood definition irrelevant, and the energy function is defined independently from the considered application; this makes it possible to find the analytical properties of the minima and may make it unnecessary to construct fast algorithms for seaching the absolute minima. To our knowledge, there are no previous works attempting to integrate SGT results in a MRF-MAP framework. Two basic properties of SG are *disorder* and *frustration*; these features are readily visualized in the energy function $E = (-1/N)\sum_{(i,j)} J_{ij}\ s_i\ s_j$, where $\boldsymbol{s} = (s_1, \ldots, s_N)$ is a generic configuration, the s_i are random variables taking values in $\{\pm 1\}$ and $\boldsymbol{J} = [J_{ij}\ ], i, j = 1, \ldots, N$ is the connection matrix. It has been proved that (see [1], chap. 4-6) choosing $J_{ij} = (1/N)\sum_{\mu=1}^{p} \xi_i^{(\mu)}\ \xi_j^{(\mu)}$ (while $\boldsymbol{\xi}^{(\mu)} \perp \boldsymbol{\xi}^{(\nu)}\ \forall \mu \neq \nu$ and $p \ll N, N \to \infty$ hold), the $\{\boldsymbol{\xi}^{(\mu)}\ |\mu = 1 \ldots p\}$ (a chosen set of configurations) are the absolute minima of E. These results can be extended from the discrete to the continuous case (i.e. $\boldsymbol{s} \in [-1, +1]^N$, see [5]). With this choice for the connection matrix is straightforward to recognize that E is a function of the scalar product between a generic configuration $\boldsymbol{s}$ and a particular configuration $\boldsymbol{\xi}^{(\mu)}$ which we want to be an absolute minima of E.

We propose to use SGT results as follows. Consider a generic pattern recognition problem: let $R = \{1, \ldots, m\}$ be a set of m sites corresponding to the features, and $L_x = [x_l, x_h] \subset \Re$ a continuous label set; then $\boldsymbol{f} = \{f_1, \ldots, f_m\}$ will be a labeling configuration in the configuration space G; given the observed data D, we define the optimal labeling $\hat{\boldsymbol{f}}$ to be the one which satisfy a MAP criterion. Now suppose we map the data from G to a space $H \equiv [-1, +1]^N$, with $N \to \infty$, using a mapping $\Phi : G \to H$. As E is a function of the scalar product, this allows us to look for kernel functions K such that $K(\boldsymbol{f}_1, \boldsymbol{f}_2) = \Phi(\boldsymbol{f}_1) \cdot \Phi(\boldsymbol{f}_2)$, and thus to use the kernel K without explicitly knowing Φ. The Mercer's condition [3] tells us for which kernels there exist a pair $\{H, \Phi\}$; the Gaussian kernel $K(\boldsymbol{f}_1, \boldsymbol{f}_2) = \exp\{-||\boldsymbol{f}_1 - \boldsymbol{f}_2||^2/2\sigma^2\}$ is a Mercer's kernel [3], that is to say it is the scalar product between two generic vector in a space $H = [-1, +1]^N, N \to \infty$ [3], which is the space where the SG energy E lives. Thus, the kernel trick allows us to realize a Spin-Glass model of a Markov Random Field.

We tested this model on two texture classification problems: in the first experiment we classified five types of textures (see Figure 1); for each class we had a sample set of 64 images, each of dimension 64×64.

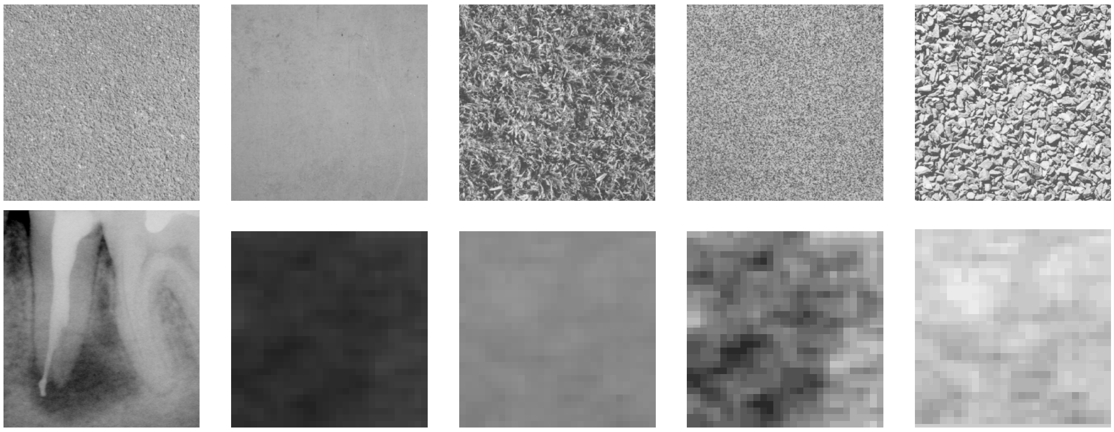

Figure 1: Textures databases: on the top are five textures taken from the Meastex Database (available at *www.cssip.elec.uq.edu.au/ guy/meastex/meastex.html*); on the bottom an example of radiographic image of periapical lesion [4], and four regions of interest representing healtly and lesioned bone.

The second experiment was done on a database of 228 Regions Of Interest (ROIs) which were extracted from radiographic images of periapical lesions (see Figure 1, [4]). These ROIs represented regions of the image where the disease could be detected. The dimensions of these ROIs varied from 10×10 to 20×20; each region was selected on the basis of visual criteria and clinical results. Features were extracted in both cases by means of the co-occurrence matrices method [6], thus we obtained a vector of five features for each sample. The experiments were performed with a leave-one-out technique and prototypes were evaluated in three different ways: mean value, median value and Gaussian mixtures. Results reported in Table 1 show that Gaussian mixtures give the best performance in both cases.

mean value	90.6%	85.1%
median value	90.3%	86.0%
gaussian mixture	91.8%	86.8%

Table 1: Experimental results

References

[1] D. J. Amit, *"Modeling Brain Function"*, Cambridge University Press, 1989.

[2] Y. Amit and M. Mascaro, "Attractor networks for shape recognition" *Tech. Rep. no. 495*, Department of statistic, University of Chicago, available at http://galton.uchicago.edu/ amit.

[3] C. J. Burges, " A Tutorial on Support Vector Machines for Pattern Recognition", *Data Miniming and Knowledge Discovery*, Vol. 2, N. 2, 1998.

[4] B. Caputo, G. E. Gigante, "Analysis of Periapical Lesion Using Statistical Textural Features", *Proc. of MIE2000*, pp 1231-1234, August 2000.

[5] T. Fukai, S. Shiino, "Large suppression of spurious states in neural networks of nonlinear analog neurons", *Phys. Rev. A*, Vol. 42, N. 12, Dec. 1990.

[6] R. M. Haralick, K. Shammugam, I. Dinstein. "Textural Features for Image Classification", *IEEE Trans. on Sys., Man and Cyb., Vol 3, No 6, 1973*.

[7] S. Z. Li, *"Markov Random Field Modeling in Computer Vision"*, Computer Science Workbench, Springer, 1995.

[8] M. Mezard, G. Parisi, M. Virasoro, *" Spin Glass Theory and Beyond"*, World Scientific Singapore, 1987.

[9] M. Opper and O. Whinter, "Gaussian Processe for Classification: Mean Field Algorithms", *to appear in Neural computation*, vol 12, issue 11 (2000).