MICROCANONICAL ANNEALING ON NEURAL NETWORKS.

G. Cattaneo and N. Cesa-Bianchi. Dipartimento di Scienze dell'Informazione, Università di Milano, via Moretto da Brescia 9, 20133 Milano, Italy.

The implementation of the classical simulated annealing algorithm on neural networks has given rise to Boltzmann machines, a model that finds its major applications in a variety of learning [1] and combinatorial optimization [2] tasks. We present a new connectionist model based on the algorithm known as microcanonical annealing [3] which is efficient and easy to implement.

The basis of microcanonical annealing is the Creutz criterion [4]. This algorithm introduces an extra degree of freedom, a "demon", carrying a variable amount of energy E_D. At every time $t = 0, 1, 2, \ldots$, a transition to a new state s' is proposed. If the change $\Delta E = E(s') - E(s)$ in the energy of the system is negative, then the transition is accepted and E_D is increased by $|\Delta E|$. Otherwise, the transition is accepted only if $E_D \geq \Delta E$; in this case E_D is decreased by $|\Delta E|$. This procedure amounts to a random walk through a surface of costant energy $E_T = E(s) + E_D$. To approximate the uniform distribution on this surface, the Creutz criterion satisfies a form of detailed balance. Microcanonical annealing is accomplished by a procedure that periodically removes energy from the demon. The resulting process can be modeled by a Markov chain whose transition matrix depends upon E_T. Because of the deterministic character of the Creutz algorithm, the annealing process causes an increasing number of transitions to occur with zero probability. This prevents the probabilistic convergence of microcanonical annealing on the global minima of E. In the neural implementation, at every time t the demon chooses randomly an unit and attemps to flip its state following the Creutz criterion. The change in energy ΔE, for unit u_i, is given by the usual formula:

$$\Delta E = (2s_i - 1)\left(\sum_{j=1}^{n} w_{ij}s_j - \tau_i\right)$$

where w_{ij} are the (symmetrical) weights, τ_i is the threshold and $s_i \in \{0, 1\}$ for $i = 1, 2, \ldots n$.

As a test bed for the algorithm, we have given a 0-1 formulation of the (NP-complete) graph bipartition problem. Every graph of m vertices is mapped on a net of $2m$ units, where each vertex v_i is represented by an ordered couple (u_i, u_i') of units. A configuration of the net is a feasible solution to the problem only when the following three conditions are simultaneously satisfied:

$$(i)\ \sum_{i=1}^{m} s_i s_i' = 0, \quad (ii)\ \sum_{i=1}^{m} s_i = \frac{m}{2}, \quad (iii)\ \sum_{i=1}^{m} s_i' = \frac{m}{2}.$$

Some weights of the nets are fixed by the problem instance, the remaining ones are chosen in order to assure that: (P1) every solution of the problem corresponds to a local minimum of the energy function of the net; (P2) the lower the cost of a solution, the lower the energy of the corresponding configuration. For this combinatorial problem, simulations show that microcanonical models and Boltzmann machines obtain similar results on networks of 128 units. However, w.r.t. the Metropolis criterion, which is the basis of classical simulated annealing, Creutz algorithm is computationally less expensive and therefore faster and suitable for an implementation on fine-grained architectures like the Connection machine.

References.

[1] G.E. Hinton, T.J. Sejnowski and D.H. Ackley, Boltzmann Machines: Constraint Satisfaction Machines that Learn, *Department of Computer Science technical report* CMU-CS-84-119, CMU, Pittsburgh, PA 15213.

[2] J.H.M. Korst and E.H.L. Aarts, Combinatorial Optimization on a Boltzmann Machine, *Proc. European seminar on neural computing*, February 1988, London.

[3] S.T. Barnard, Stereo Matching by Hierarchical Microcanonical Annealing, *Proc.* IJCAI-87, Milano, August 1987, 832–835.

[4] M. Creutz, Microcanonical Monte Carlo Simulation, *Phys. Rev. Lett.*, 50(1983)1411–1414.