ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

An Optimization Neural Network Model with Time-Dependent and Lossy Dynamics

Z. Heszberger J. Bíró E. Halász T. Henk
High Speed Networks Laboratory
Department of Telecommunications and Telematics
Budapest University of Technology and Economics
H-1117, Pázmány P. s. 1/D, Budapest, Hungary
Biro@ttt-atm.ttt.bme.hu

Abstract. The paper deals with continuously operating optimization neural networks with lossy dynamics. As the main feature of the neural model time-varying nature of neuron activation functions is introduced. The model presented is general in the sense that it covers the cases of neural networks for combinatorial optimization (Hopfield-like networks) and neural models for optimization problems with continuous decision variables. Besides the brief stability analysis of the proposed neural network we also show how to derive from it lossy versions of improved Hopfield neural models .

1. Introduction

In the last decade considerable attention has been paid for optimization neural networks. Such systems are considered as potentially efficient hardware solutions for large-scale or hard optimization problems [1], [2]. Although many problematic, and therefore challenging question arises in connection with the hardware realization, an optimization neural network could work very fast as a parallel computational structure in a truly distributed implementation.

One of the first pioneers in this field were Hopfield and Tank who presented a 'neural like' network for solving combinatorial problems [1]. This network, since then always referred to as Hopfield neural network (HNN) is a continuously operating model being very close to analog circuit implementation. Since 1985 a wide variety of Hopfield-like neural networks have been designed for improving the performance of the original model, i.e. for avoiding local optima or spurious states with high probability. (The term "local optima" stands for locally optimal stable equilibrium points). Besides the continuously operating networks which fit to analog circuit realizations [3], [4], [5] discrete versions being more suitable for computer implementations [2], have been also developed. In the paper, we concentrate on the continuous models of which operation can be described by differential equations with special regard to lossy dynamics.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

The adjective *lossy* means in this context that the time derivatives of the state variables in the network are proportional *not* with the gradient of the objective function, but the gradient plus the corresponding state variable itself. In a circuit realization, it means that the integrators are *leaky*, that is their input resistances are not infinitely large, which is doubtless a better model of real circuits. This lossy property implies the drawback that the energy function and the objective function don't match each other. The difference can be arbitrarily small with arbitrarily steep activation function at decision neurons.

In the original Hopfield network the analog dynamics is lossy and some other papers also discussed the drawback of lossy dynamics. In this paper, we reintroduce lossy dynamics for a broader class of optimization neural nets and show how to utilize the lossy property for improving network performance provided time-variation of activation function is allowed.

2. The optimization neural network model

For easy reference let's call the neural model to be introduced $TONN$ standing for time-varying optimization neural network. One of the main characteristics of $TONN$ is that it is a continuously operating system seeking for a local minimizer of an unconstrained objective function in a gradient manner. It means that the operation can be described by the following set of differential equations.

$$\frac{dz_k}{dt} = -G_k z_k - \frac{\partial E(x)}{\partial x_k} \ , \quad k = 1, \ldots, n \tag{1}$$

where G_k are positive constants, the n dimensional vector x comprises the decision variables and z_k's are inner state variables. The unconstrained objective function to be optimized is denoted by $E(x)$. The decision variables x_k are obtained by

$$x_k = \Theta_k(z_k/T_k(t)) \tag{2}$$

where functions $\Theta_k(.)$ are sigmoid-like functions with finite positive and negative saturations, that is $\Theta_k(\infty) = X_{kmax} > 0$ and $\Theta_k(-\infty) = X_{kmin} < 0$. A further assumption is that $\Theta(0) = 0$. For example, a widely used activation function in neural network world is $\Theta(.) = \tanh$. The steepness of $\Theta_k(.)$ can be controlled by finite $T_k(t)$ which is allowed to vary in time in $TONN$ in such ways that $T_k(t)$ can be strictly monotone increasing or strictly monotone decreasing or constant in time and $\lim_{t\to\infty} T_k(t) = \hat{T}_k \geq 0$).

If $E(x)$ is an objective function with discrete decision variables (x_k should be $0, 1$ or $-1, 1$) then $TONN$ can be used for solving combinatorial optimization problems. Otherwise, if $E(x)$ is derived from an optimization problem with continuous decision variables (x_k's can take any values in a certain range) then $TONN$ can be attached to the group of nonlinear (or linear) programming neural networks.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

Before investigating the stability properties of $TONN$, let's briefly consider the lossless dynamics neural network based on gradient search. In this case, all G_k's are zero and the qualified Lyapunov function of this system is $E(x)$ provided it is bounded from below. A simple but very important observation is that the steepness of Θ_k does not affect the objective function of which the local minimizer is retrieved by the network. Further, the time-varying nature of $T_k(t)$ even does not influence the Lyapunov function $E(x)$. In spite of this fact in many cases the performance can rely on gradually increasing steepness of the activation function, for example, in case of $HANN$ [3].

In connection with lossy dynamics we encounter the problem that the original objective function to be optimized and the Lyapunov function which is really minimized by the network do not match each other. For a moment let us consider a lossy system which can be described by a similar equation to (1) but with the difference that $T_k(t)$'s are constants in time like in the original neural model of Hopfield and Tank [1]. Then the Lyapunov function of such systems is as follows.

$$L(x) = E(x) + \sum_{k=1}^{n} T_k G_k \int_0^{x_k} \Theta^{-1}(\xi) d\xi \qquad (3)$$

A minor but important observation, which will be also referred later (in Section 3.1.2.) in connection with the relations between $TONN$ and other neural systems, is that parameters T_k and G_k play the *same* role in the Lyapunov function.

As regards $TONN$ the questions arise that whether the network remains stable, and if yes, what is the function which is minimized by $TONN$. The following theorem sheds light on the results in connection with these problems.

Theorem 1 *If $E(x)$ is bounded from below and the function*

$$H_k(x_k) = \int_0^{x_k} \Theta^{-1}(\xi) d\xi$$

is bounded on the set $\{X_{kmin} \leq x_k \leq X_{kmax}\}$, $\forall k$ then $TONN$ is asymptotically stable in Lyapunov sense and converges to a local minimizer of the function

$$E(x) + \sum_{k=1}^{n} \hat{T}_k G_k \int_0^{x_k} \Theta^{-1}(\xi) d\xi$$

where $\hat{T}_k$, $k = 1, \ldots, n$ are the limit values of $T_k(t)$'s.

Proof: see [6]

Remarks:

The boundedness of $H_k(x_k)$ is reasonable for the following reasons. For example, if $\Theta_k(z_k, T_k(t)) = \tanh(z_k/T_k(t))$ then $X_{kmax} = -X_{kmin} = 1$ and

$$H_k(1) = \int_0^1 \tanh^{-1} \xi d\xi$$

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

can be described as an improprius integral $\lim_{z\to\infty}(z\tanh z - \int_0^z \tanh\zeta d\zeta)$ which is equal to $\lim_{z\to\infty}(z\tanh z - \ln(\cosh z))$ Since $\tanh z$ tends to 1 and $\cosh z$ converges to $e^z/2$ as $z \to \infty$ the limit above is $\ln 2$. Consequently, $\sup_{x_k\in\mathcal{X}_k} H_k(x_k) = \ln 2$ and therefore $\hat{H}_k$ can be $\ln 2 + \varepsilon$ where ε is any small positive number.

It can be seen that in the Lyapunov function of $TONN$ $T_k(t)$ and G_k are similarly the weights of the additional terms. It implies that their role may be exchanged, that is G_k can be time varying and T_k can be constant (if it was not constant) without changing the objective function.

If the time-varying $T_k(t)$ tends to zero then the network finally converges to a minimizer of $E(x)$. This issue is acceptable in case of Hopfield-like networks (nonlinearities becamehard limiters) but shouldn't be concerned with optimization neural nets producing continuous decision variables.

3. Relations to other optimization neural networks

In this section we discuss what is the relation between $TONN$ and Hopfield-like optimization neural network models.

Combinatorial optimization neural networks like the Hopfield model is essentially based on the gradient descent seeking for an optimum of the objective function. In case of the Hopfield neural network the energy function that should be minimized can be given by a general quadratic from

$$E = \frac{1}{2}x^tWx + b^tx \tag{4}$$

where x comprises the decision variables x_k, $x_k = 1$ or -1. (No matter to transform it such that $x_k = 1$ or 0). W is an $n \times n$ symmetric matrix and b is an n dimensional input vector. The operation of a *lossless* dynamics network can be described by

$$\frac{dz_k}{dt} = -\frac{\partial E(x)}{\partial x_k}\ ,\ x_k = \tanh(z_k/T)\ ,\ k = 1..n \tag{5}$$

where T is a positive constant. In fact, this network performs a continuous relaxation of the discrete optimization problem, therefore, x_k should be digitized after the convergence. A *lossy* version of the network above (in fact the original Hopfield model was presented as a lossy system) can be obtained from $TONN$ if $T_k(t)$ are positive constants in time and $\Theta_k = \tanh$, $\forall k$. The main drawbacks of these models that the equilibrium state represents only a local minimizer of $E(x)$ or some of x_k's do not satisfactorily converge towards 1 or -1.

3.1. Hardware annealing neural network

In hardware annealing neural network ($HANN$) the scalar T is designed to be time-varying in such a way that the steepness of the sigmoid activation is

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

gradually increasing in time. It resulted in a similar effect to that of simulated annealing (SA), thus, providing better chance to avoid local optima [3]. In this case, the governing equations are similar to (5) except that T is decreasing in time.

In [3] the operation of the network is modeled by lossless dynamics like (5). The better performance relies on the time-varying nature of the activation function. Moreover, the $HANN$ minimizes an energy function in the form of $E(x)$, therefore, the better performance can not analytically be caught through the Lyapunov function.

A lossy version of $HANN$ can be derived from $TONN$ in a way that $\Theta_k(.) = \tanh(.)$ and all $T_k(t)$'s are identical and strictly monotone decreasing functions of time tending to 0. $HANN$ with lossy dynamics is certainly a better approach of real circuit behaviours. The network remains stable according to Theorem 1 and the Lyapunov function $L(x)$ makes clear that why the neural network has chance to avoid local optima. To support this latter statement let's consider the Lyapunov function in (3). The additional term besides $E(x)$ is convex because H_k is strictly monotone increasing function. Generally, if an appropriate convex function is added to a function to be optimized, then some of the local optima of the objective function can be eliminated at the expense of changing the minimizers including the global one. However, in this modified $HANN$ the additional convex term is gradually disappearing as T is approaching to 0. If this process is slow enough the network output may track the time-varying global optimum finally converging probably to the best minimizer of the original objective function. At the same time the steepness of Θ is increasing, in this way, the decision variables are really forced tending to -1 or 1. A similar phenomenon can be observed in simulated annealing regarding the objective function and the probability density function of states.

3.2. Matrix graduated neural network

In [4] a neural network is proposed with time-varying main diagonal entries of W. w_{ii}'s start from positive values and are decreasing in time in a *discrete* manner. The network is based on the matrix graduated nonconvexity ($MGNC$) algorithm, therefore, hereafter we refer to this neural system as matrix graduated neural network ($MGNN$). In this model, the activation function is constant in time and piece-wise linear. It is shown that the network can produce better optimum than that of the original Hopfield model, for instance, in solving the traveling salesman problem. The dynamics of $MGNN$ is lossless and the activation function is piece-wise linear, that is $x_k = \Theta_{MGNN}(z_k) = z_k$ if $|z_k| \leq 1$, otherwise $x_k = \text{sign}(z_k)$.

Now, we show how to derive a lossy version of $MGNN$ from $TONN$. Obviously, we should choose Θ_k as Θ_{MGNN} defined above. The role of $T_k(t)$ and G_k should also be exchanged so that $T_k = 1, \forall k$ (due to the definition of Θ_{MGNN}) and $G_k(t)$ are time-varying with the properties of $\dot{G}_k(t) < 0$, $G_k(t) \to 0$ and $\dot{G}_k(t) \to 0$ as $t \to \infty$. In this case $H(x_k) = x_k^2/2$, therefore, the qualified Lyapunov function of the system is $L(t) = E(x) + \frac{1}{2}\sum_k T_k G_k(t) x_k^2 =$

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 287-292

$\frac{1}{2}\sum_{i,j;i\neq j} w_{ij}x_i x_j + \frac{1}{2}\sum_k \Big(w_{kk} + T_k G_k(t)\Big) x_k^2$. It evidently implies that the main diagonal elements of $\hat{W}$ $\hat{w}_{kk} = w_{kk} + 1/2T_k G_k(t)$ are decreasing in time while the shape of activation function doesn't change, that is we have an $MGNN$-like network with lossy dynamics and continuously decreasing main diagonal entries. This results has two-fold significance because besides taking into account nonideal integrators through lossy dynamics the continuous evolution of $\hat{w}_{kk}(t)$ in time may provide fully analog implementation.

Conclusion

A time-varying optimization neural network model with lossy dynamics referred to as $TONN$ was introduced. The non-trivial stability properties was presented. It was also shown how to derive from $TONN$ lossy versions of known Hopfield-like neural networks with improved performance.

References

[1] Hopfield, J.J., Tank, D.W.: 'Neural' Computation on Decisions Optimization Problems. *Biological Cybern.* **52** (1985) 141–152

[2] Peterson, C., Söderberg, B.: A New Method for Mapping Optimization Problems onto Neural Networks. *Int. Journal of Neural Systems*, **1**(1989), 3–22.

[3] Lee, B.W., Sheu, B.J.: Hardware Annealing in Electronic Neural Networks. *IEEE Trans. on Circuits and Systems*, **38**, (1991) 134–137.

[4] Abe, S., Gee, A.H.: Global Convergence of the Hopfield Neural Network with Nonzero Diagonal Elements. *IEEE Trans. on Circuits and Systems.* **42** (1995) 39–45.

[5] Takefuji, Y., Lee, K.C.: An Artificial Hysteresis Binary Neuron: a Model Suppressing the Oscillatory Behaviours of Neural Dynamics. *Biological Cybernetics.* **64** (1991) 353-356.

[6] J Bíró, Optimization Neural Networks, Ph.D. dissertation, Technical University of Budapest, 1998.

[7] Bíró, J., Henk, T., Boda, M. "Analog Neural Optimization for ATM resource Management", *IEEE J. on Sel. Ar. of Comm.*, February, 1997.

[8] Bíró, J., Koronkai, Z., L. Ast, T. Trón, M. Boda: Analyses of Extended and Generalized Optimization Neural Networks. *Journal on Artificial Neural Networks*, (ABLEX, USA), **4**(1995), 401–410.