ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 289-294

On the short-term-memory of WTA nets

Brijnesh J. Jain, Fritz Wysotzki
Technical University Berlin, Computer Science Department,

Franklinstr. 28/29, 10587 Berlin, Germany

Abstract. An exact solution of a system of coupled differential equations describing the dynamics of a special class of winner-take-all networks is given. From the solution two properties of the short-term-memory traces are derived: (1) information preservation and (2) a discrimination measure. These properties justify a biologically inspired fault tolerant extension of the network using differentiating neurons.

1 Introduction

One way to deal with the maximum selection from a set of inputs within a connectionist framework are *winner-take-all* (WTA) networks ([4]). The operation of these networks is a mode of contrast enhancement and pattern normalization where only the unit with the highest activation fires and all other units in the network are inhibited after some setting time. References to common competitive architectures to select the maximum or minimum from a set of data can be found in [10]. Exemplary for a practical application field of WTA nets we mention classification tasks and knowledge discovery of structured objects ([12] and references therein). Results on the computational power of competitive nets can be found in [11].

There is a growing body of mathematical results on competitive neural systems describing the dynamics of competition. Well known series of articles concerning the stability analysis includes the work of Wilson and Cowan, Grossberg et al., Amari et al. ([9] and references therein), and the Lyapunov method in the Cohen-Grossberg ([1]) and Hopfield-Tank ([7]) models. Further interesting results on the dynamics of competitive models are given in [2], [3], [6], and [8].

All articles mentioned have in common, that they examine models for which the corresponding system of differential equations can't be solved exactly for the general case. Thus, results stating that a dynamical system must converge to an equilibrium do not specify which equilibrium in fact is approached. Though it can be shown that competitive networks often arrive at a certain choice, but the nature of the choice is highly dependent on the network parameters, the initial activation, and the external input. In general, knowledge of a solution enables us to uncover and prove basic properties of the networks behavior in a more simplified fashion. Solving dynamical systems assumes simplifications in the model of consideration. The understanding of a simplified model then can suggest principles that can also be used in more complex models.

This paper investigates the dynamics of a special class of mutually inhibitory WTA networks with linear transfer function and external input. The dynamics of the net can be described by a system of coupled differential equations which can be solved exactly (Sect. 2). From the solution we derive and prove two properties of the network's behavior: (1) a form of information preservation in the short-term memory (STM) trace; (2) a form of an inherent discrimination measure (Sect. 3). Both properties give rise to a robust, fault tolerant extension of our model using differentiating neurons (Sect. 4). The extended model detects local extrema in the STM traces and stabilizes the system.

We denote vectors by bold letters (e.g. $\mathbf{x}$). For any vector $\mathbf{x} = (x_1, \ldots, x_n)$ we set $\bar{x} := \frac{1}{n}\sum_i x_i$ and $\bar{\mathbf{x}} := (\bar{x}, \ldots, \bar{x})$.

2 The WTA model

The dynamical system we will work with is a linear WTA model with laterally inhibitory connections of the form

$$\dot{x}_i(t) = -dx_i(t) + \sum_{j=1,\, j\neq i}^{n} w_{ij} f(x_j(t)) + I_i, \qquad x_{0i} := x_i(0) \tag{1}$$

where $x_i(t)$ is the activation or the STM trace (in the sense of Grossberg, see e.g. [5]) of unit i, $w_{ij} = -w < 0$ represents the inhibitory strength of the synapse connecting unit i and unit j, $d > 0$ is the selfinhibition and I_i an external input. Here we assume $w > d \geq 0$. f is a transfer function of the form

$$f(x) = \begin{cases} \phi & : \text{ for } x \geq \phi \\ x & : \text{ for } x \in \,]\psi, \phi[\\ \psi & : \text{ for } x \leq \psi \end{cases}$$

whereby $\psi < \phi$ are arbitrary constants. This network is a special case of an *additive short-term-memory* model (see [5]). The aim of this network is to discriminate input patterns, i.e. the components of an input vector.

In the following we focus on the dynamics of the net where the STM traces $x_i(t)$ are in the range of the open interval $]\psi, \phi[$. To investigate the general tendency of the network's behavior we will assume ϕ and ψ to be sufficiently large or shifted towards $+\infty$ and $-\infty$. So we are concerned with a linear transfer function $f(x) = x$. Then a solution of eqn. (1) can be given in closed form: The vectors $\mathbf{v_1} = (-1, 1, 0, \ldots, 0)^T$, ..., $\mathbf{v_{n-1}} = (-1, 0, \ldots, 0, 1)^T$, $\mathbf{v_n} = (1, \ldots, 1)^T$ form an eigenbasis of W with distinct eigenvalues $\lambda_1 = w - d$ and $\lambda_2 = -(n-1)w - d$ where $\mathbf{v_1}, \ldots, \mathbf{v_{n-1}} \in Eig(W, \lambda_1)$ and $\mathbf{v_n} \in Eig(W, \lambda_2)$. To verify this statement check $W\mathbf{v_i} = \lambda \mathbf{v_i}$ $(1 \leq i \leq n)$ for a suitable $\lambda \in \{\lambda_1, \lambda_2\}$ and show that the $\mathbf{v_i}$ are linear independent. With knowing the eigenvectors and their corresponding eigenvalues we are able to derive a solution.

Proposition 1 *The solution of the differential equation* $\dot{\mathbf{x}} = W\mathbf{x} + \mathbf{I}$ *in* $\mathbb{R}^n$ *with initial condition* $\mathbf{x}(0) = \mathbf{x_0}$ *is given by the formula*

$$\mathbf{x}(t) = \left(\mathbf{x_0} - \bar{\mathbf{x}}_\mathbf{0} + \frac{\mathbf{I} - \bar{\mathbf{I}}}{\lambda_1}\right) e^{\lambda_1 t} + \left(\bar{\mathbf{x}}_\mathbf{0} + \frac{\bar{\mathbf{I}}}{\lambda_2}\right) e^{\lambda_2 t} - \frac{\mathbf{I} - \bar{\mathbf{I}}}{\lambda_1} - \frac{\bar{\mathbf{I}}}{\lambda_2}.$$

Proof: The proof is a straightforward matter of differentiating the solution and plugging the derivative into the differential equation to check whether it is correct. □

A global stabilty analysis shows that the origin $\mathbf{0}$ is a saddle point, i.e. the system is unstable. Trajectories with $\mathbf{x_0} = \bar{\mathbf{x}}_\mathbf{0}$ go to $\mathbf{0}$ on the line in direction of positive and negative multiples of the eigenvector $\mathbf{v_n}$, and those with $\bar{x}_0 = 0$ go to ∞ on the hyperplane defined by $Eig(W, \lambda_1)$. All other trajectories are superpositions of these motions. Furthermore it can be shown that the net performs *contrast enhancement*[1] and *pattern normalization*[2] whereby the ratios of contrasts remain constant the whole time.

3 Analysis of the STM traces

In the following some properties of the STM traces $x_i(t)$ are derived. A key result is the ability of the system to preserve information given by the input patterns in the short-term-memory of units with above average initial activation. During discrimination the stored information is used to provide a measure of how plausible the choice is. For convenience, we call units with initial activation $x_{0i} > \bar{x}_0$ *dominating units.*

For a dominating unit i the STM trace $x_i(t)$ decreases until it reachs a minimum at t_*^i. Subsequently the activation $x_i(t)$ is monotonously increasing for $t > t_*$. On the other hand, if $x_{0i} < \bar{x}_0$, then $x_i(t)$ is monotonously decreasing for $t > 0$. In the homogeneous case this is evident for $x_i < \bar{x}_0 \neq 0$, since $\lambda_2 < 0$ and by the following inequality

$$x_i(t) = (x_i - \bar{x}_0)e^{\lambda_1 t} + \bar{x}_0 e^{\lambda_2 t} < \bar{x}_0 e^{\lambda_2 t}.$$

Of interest are the STM-traces $x_i(t)$ of dominating units i, since they reveal a form of information preservation and an inherent discriminating measure when they arrive at their minimum. To show these properties, we first have to determine the minimum of $x_i(t)$.

Proposition 2 *Let* $u_i := (I_i - \bar{I})/\lambda_1 + \bar{I}/\lambda_2$, $y_i := \lambda_1(x_{0i} - \bar{x}_0) + I_i - \bar{I}$, $z_i := -\lambda_2 \bar{x}_0 - \bar{I}$. *If the following conditions are satisfied*

$$(1)\ x_{0i} > \bar{x}_0 \qquad (2)\ I_i \geq \bar{I} \qquad (3)\ z_i \neq 0 \qquad (4)\ y_i/z_i > 0$$

then $x_i(t)$ *has a global minimum at*

$$t_*^i = \frac{1}{\lambda_2 - \lambda_1} \ln\left(\frac{y_i}{z_i}\right) \qquad \textit{with} \qquad x_i(t_*^i) = \frac{\lambda_2 - \lambda_1}{\lambda_1 \lambda_2} y_i^{\frac{\lambda_2}{\lambda_2 - \lambda_1}} z_i^{\frac{\lambda_1}{\lambda_1 - \lambda_2}} - u_i.$$

[1] contrast enhancement: The absolute difference $|x_i(t) - x_j(t)|$ increases with time t.

[2] pattern normalization: the total activation $\sum_i x_i$ approaches 0 with increasing time t.

Proof: Using u_i, y_i, and z_i in Prop. 1 gives us

$$x_i(t) = \frac{1}{\lambda_1} y_i e^{\lambda_1 t} - \frac{1}{\lambda_2} z_i e^{\lambda_2 t} - u_i.$$

Differentiating $x_i(t)$ and equating with 0 leads to $\dot{x}_i(t) = y_i e^{\lambda_1 t} - z_i e^{\lambda_2 t} = 0$. Solving this equation to t yields t_*^i. Since $x_i(t)$ is continuously differentiable and $\ddot{x}_i(t) > 0$, t_*^i is indeed a global minimum. To conclude the proof, plug t_*^i into $x_i(t)$. □

Conditions (1) and (2) restrict the statement to all STM traces $x_i(t)$ of dominating units i, whereas conditions (3) and (4) exclude the stable state and normalized patterns, i.e. patterns with $\bar{x}_0 = 0$. The STM traces of stable states and normalized patterns are either constant or monotoneously increasing and decreasing, respectively.

Now let us turn to the key result of this paper, the ability of the network to preserve the supplied information and to use this infomation to discriminate the input patterns by a reasonable measure. At time t_*^i the STM of unit i retrieves approximately the difference $x_{0i} - \bar{x}_0$ of the initial activation of unit i and the average initial activation $\bar{x}_0$. This approximation improves for an increasing number n of units (see Corallary 1). *The dominating units do not only signal the choice made, but also provide a measure of how plausible and how focussed the decision is.* According to Prop. 2 the measure is a monotoneously increasing function d of the difference $x_{0i} - \bar{x}_0$.

Corollary 1 *Consider the conditions of Prop. 2. Then*

$$\lim_{n\to\infty} x_i(t_*^i) = x_{0i} - \bar{x}.$$

Proof: Follows directly from Prop. 2 by taking $\lim_{n\to\infty}$. □

Corollary 1 suggests two extensions of our model (see Sect. 4). To justify the extensions we have to prove two order preserving properties of the STM traces. The first one says that the network preserves the order of the activations.

Lemma 1 *Let $x_{0i} < x_{0j}$ and $I_i \leq I_j$. Then $x_i(t) < x_j(t)$ for $t \geq 0$.*

Proof: Using the solutions $x_i(t)$ and $x_j(t)$ given in Prop. 1, we obtain

$$x_i(t) - x_j(t) \quad = \quad (x_{0i} - x_{0j})e^{\lambda_1 t} + \frac{I_i - I_j}{\lambda_1}(e^{\lambda_1 t} - 1).$$

By assumption $x_{0i} - x_{0j} < 0$ and $I_i - I_j \leq 0$. Since $\lambda_1 > 0$, we have $\frac{I_i - I_j}{\lambda_1} \leq 0$ and $e^{\lambda_1 t} - 1 \geq 0$ for $t \geq 0$. Putting all inequalities together proves the assertion. □

A similar statement also holds for the chronological order of the STM traces arriving at their minima. The STM trace $x_i(t)$ of unit i passes the minimum before the STM trace $x_j(t)$ of unit j if $x_{0i} > x_{0j}$. More precisely:

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 289-294

Lemma 2 *Consider the assumptions in Prop. 2. Let $x_{0i} > x_{0j} > \bar{x}_0$ and $I_i > I_j > \bar{I}$. Then $t_*^i < t_*^j$.*

Proof: First note that $z_i = z_j$ for all $1 \leq i, j \leq n$. By assumption

$$y_i - y_j = \lambda_1(x_{0i} - x_{0j}) + I_i - I_j > 0$$

holds. Hence, $y_i/y_j > 1$. With $\lambda_1 = w - d$ and $\lambda_2 = -(n-1)w - d$ we get

$$t_*^i - t_*^j = \frac{1}{\lambda_2 - \lambda_1}\left\{\ln\left(\frac{y_i}{z_i}\right) - \ln\left(\frac{y_j}{z_j}\right)\right\} = -\frac{1}{nw}\ln\left(\frac{y_i}{y_j}\right) < 0$$

□

4 Conclusion

The result of Corollary 1 justifies two biologically motivated extensions of our model with similar behavior in their STM traces: (1) an enlarged network using (2) differentiating neurons. Extension (1) leads to a robust, fault tolerant network whereas extension (2) keeps the STM traces bounded and guarantees a stable behavior of the system.

(1) Robust, fault tolerant networks. The network can be enlarged without changing the behavior of the STM traces by the following procedure. Increase the number n of units by adding artifical units without changing the average activation $\bar{x}_0$. This can be achieved by connecting a fixed number k of copies of each unit to the net where the activations of the copies are slighty perturbed by a random noise. This enlarges the system to kn mutually inhibited units consisting of n groups each with k units of nearly equal initial activation. If the noise is Gaussian distributed with expectation 0 and small variance σ^2 compared to the initial activation of the original units, then the expectation of $\bar{x}_0$ of the enlarged system corresponds to the average of the original system. Now let x_{0i} be the initial activation of a representative of group i. Then at t_*^i the activations of group i retrieve a better averaged approximation of $x_{0i} - \bar{x}$ than unit i of the original system (Corollary 1). Thus, a wider distribution of the input patterns does not only provide a system which is robust against failure of single neurons but also sharpens the focus on the given problem.

(2) Differentiating neurons. We call neurons which are able to recognize local extrema of their STM traces differentiating neurons. With differentiating neurons we can construct a network keeping the activations bounded and leading to a stable state within finite time. In mathematical terms, we may describe a differentiating neuron i by the following pair of iterative equations[3]:

$$x_i(t+1) = (1-d)x_i(t) + \textstyle\sum_{j \neq i} w_{ij} y_j(t) + I_i$$
$$y_i(t+1) = x_i(t+1) \cdot f[x_i(t) - x_i(t+1)]$$

where $x_i(t+1)$ is the activation of neuron i at time $t+1$, $f(x)$ the threshold function with $f(x) = 0$, if $x \leq 0$ and $f(x) = 1$ otherwise. Finally $y_i(t+1)$ is the output signal of the neuron.

[3] For convenience we consider iterative equations instead of differential equations.

ESANN'2001 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 25-27 April 2001, D-Facto public., ISBN 2-930307-01-3, pp. 289-294

Now assume $d = 0$ and w sufficiently small, such that the iterative system is a good approximation of the underlying system of differential equations. From Lemma 1 and 2 we know, that the STM trace of unit i with maximal initial activation arrives first at its minimum. This leads to an ouptut signal $y_i(t_*^i) = 0$. In the next iteration step the net will arrive at a stable state. The winning neuron i is the one first entering a stable state. By construction, neuron i has an approximately identical STM trace in the interval $(0, t_*^i]$ as given in Prop. 2.

References

[1] M.A. Cohen and S. Grossberg. Absolute stability of global pattern formation and parallel memory storage by competitive neural networks. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(5):815–826, 1983.

[2] B. Ermentrout. Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks*, 5:415–431, 1992.

[3] Y. Fang, M.A. Cohen, and T.G. Kincaid. Dynamics of a winner-take-all neural network. *Neural Networks*, 9(7):1141–1154, 1996.

[4] J.A. Feldmann and D.H. Ballard. Connectionist models and their properties. *Cognitive Science*, 6:205–254, 1982.

[5] S. Grossberg. Nonlinear neural networks: Principles, mechanisms, and architectures. *Neural Networks*, 1:17–61, 1988.

[6] M.W. Hirsch. Convergent activation dynamics in continuous time networks. *Neural Networks*, 2:331–349, 1989.

[7] J.J. Hopfield. Neurons with graded respose have collective computation properties like those of two-state neurons. *Proceedings National Academy of Sciences*, 81:3088–3092, 1984.

[8] M. Lemon and B.V.K. Vijaya Kumar. Emulating the dynamics for a class of laterally inhibited neural networks. *Neural Networks*, 2:193–214, 1989.

[9] D.S. Levine. *Introduction to neural and cognitive modelling.* Lawrence Erlbaum Associates, Inc., Hillsday, New Jersey, 1991.

[10] R.P. Lippman. An introduction to computing with neural nets. *IEEE ASSP Magazine*, pages 4–22, April 1987.

[11] W. Maass. On the computational power of winner-take-all. *Neural Computation*, 12(11):2519–2536, 2000.

[12] K. Schädler and F. Wysotzki. Comparing structures using a Hopfield-style neural network. *Applied Intelligence*, 11:15–30, 1999.