ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 1-6

Concerning the Formation of Chaotic Behaviour in Recurrent Neural Networks

Thorsten Kolb*, Karsten Berns**

* Technische Universität Cottbus, Lehrstuhl f. Numerische und Angewandte Mathematik
** Forschungszentrum Informatik Karlsruhe, Fachgruppe Interaktive Planungstechnik

Abstract

The formation of chaotic behaviour in neural networks has been analyzed and some principles have been established, which illustrate the effects accounting for that behaviour. The use of networks which act in a chaotic manner will also be discussed. Finally the applicability of methods used in chaos theory to the field of connectionism will be examined.

1 Introduction

To generate sequences one often uses recurrent neural networks to get a distributed representation and generalisation behaviour. This is for example important in control problems. In some cases a learning process can result in chaotic behaviour. There is then no fault tolerance and no generalisation of the net. It is therefore desirable to widen the understanding of principles causing chaos.

Chaotic behaviour has often been presented in Neural Networks. Articles have been written on the topic of chaos in learning processes [7], [8] or chaos resulting from time-delayed connections [3]. By using nets common in practice the number of units used to produce chaos decreased from infinity [6] to 128 [2] and further to eight [5]. In 1991 Xin Wang presented an example of chaotic behaviour in a very simple neural network consisting only of two units wich are fully connected to each other [9]. He also gave a mathematical proof of the behaviour found being chaotical. The formation of this behaviour has not been looked into by him.

In order to answer this question we first give a definition of chaos. Next some methods of chaos theory are examined. Then we introduce the neural model which we try to describe. In section 3.2 we present a chaotic acting network examining the formation of chaotic phenomena. Finally we try to evaluate the described principles.

2 Fundamentals of chaos theory

Different definitions of chaos is the major reason why there is confusion in the field of research on chaotic phenomena. One major reason for confusions in the field of research on chaotic phenomena is caused by the different definitions of chaos. This

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 1-6

is the reason why the results of many papers can not be compared. Other difficulties arise in complex n-dimensional systems, where phenomena which cannot be directly described are interpreted as chaotical. Finally the use of techniques, developed for the analysis one dimensional systems cannot be transferred in any case to n-dimensional systems. For this reasons, some foundations of chaos theory are described below.

2.1 A definition of chaos

Although there are many other definitions of chaos a simple one concernes with sensitive dependence of the initial conditions. Based on this, Devaney [1] defined chaos as follows:

Definition (Chaos) Let $f : M \rightarrow M$ be a map, where M is a metric space. The map f is said to be chaotic if

1. f has sensitive dependence on initial conditions, that is there exist $\delta > 0$ such that, for any $x \in M$ and any neighborhood U of x, there exists $y \in U$ and $n > 0$ such that $|f^n(x) - f^n(y)| > \delta$.

2. f is topological transitive, that is for any pair U, V of open sets, there exists $k \geq 0$ such that $f^k(U) \cap V \neq \emptyset$.

3. Periodic points of f are dense in M.

One can prove that part (1) of this definition is superfluous, because (2) and (3) imply (1). In the given form it has the advantage of being readily verified in many cases. In the following subsection some methods on detecting this type of chaos are presented.

2.2 Techniques on handling chaos

One of the most used techniques to demonstrate chaotic behaviour, the principle of bifurcation (i.e. a qualitative change of the limit behaviour of a dynamical system) does not depend on any dimensions. This is not the case when using bifurcation diagrams to visualize such changes. Hence, bifurcation diagrams are in general not really meaningful to use when describing the dynamics of a neural network.

Having a description of an attractor, one can make use of the belonging Hausdorff dimension to detect chaotic behaviour. The attractor is called fractal (and the underlying system is called chaotic), if the Hausdorff dimension is not an integer.

The principle of topological equivalence can be used to transform a given function into another one, being easier to analyze or where the behaviour is known. Thereby two functions f and g are called topological equivalent, if there is an injective function h so that $g = h^{-1} \cdot f \cdot h$.

Characteristic exponents (or Liapunov exponents) give a measure of the drifting of two neighbouring points apart from each other. An n-dimensional system is characterized by n Liapunov exponents. The existence of a positive Liapunov exponent implies chaotical behaviour.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 1-6

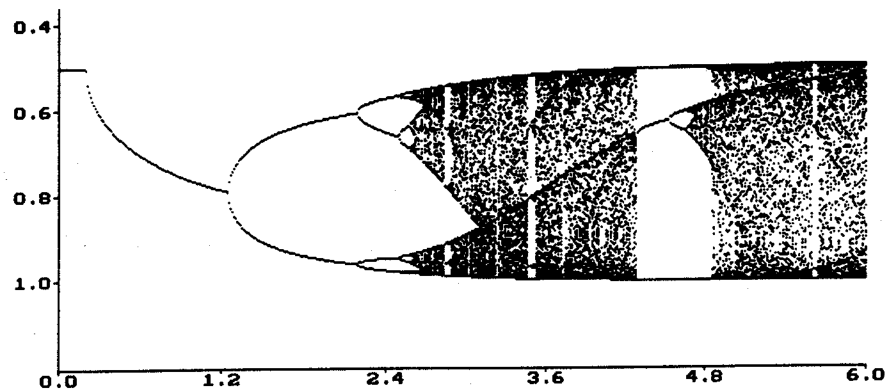

Figure 1: Bifurcation diagram of the Wang-attractor.

3 Chaos in neural networks

3.1 A simple neural network model

Models in technical use are the background for the model we analyse. It is simplified as much as possible to make the conclusions clearer. The weighted inputs of a unit are added and one applies a sigmoid activation function to the result in order to limit the output. The units are fully connected to each other and update their output by using a synchronous iteration mode. The outputs of the units can hence be combined in a state vector. The function $\boldsymbol{f}$ realised by a net can be formalised as the transition from one state vector to another, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}_{akt}(\mathcal{W} \cdot \boldsymbol{x})$. $\boldsymbol{x}$ is here the actual state vector,

$$\boldsymbol{f}_{akt}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \tanh(x_1) \\ \vdots \\ \tanh(x_n) \end{pmatrix}$$

is the n-dimensional expansion of the activation function $f_{akt}(x) = \tanh(x)$ and $\mathcal{W}$ is the weight matrix, which is constructed by rows of weights belonging to one unit.

3.2 The Wang attractor

The chaotic behaviour found by Xin Wang occurs in a net with only two units and the activation function $f_{sig}(x) = (1 + e^x)^{-1}$. The weight matrix is

$$\mathcal{W} = a \cdot \begin{pmatrix} -5 & 5 \\ -25 & 25 \end{pmatrix}.$$

The parameter a is used to vary the dynamical properties of the net. The bifurcation diagram in figure 1 demonstrates the period doubling to chaos for increasing values of a.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 1-6

Wang described his attractor as aspecial case of a net with the matrix

$$\mathcal{W} = a \cdot \begin{pmatrix} x & kx \\ y & ky \end{pmatrix}.$$

He theoretically proved the existence of chaos by transforming this family of functions into a topologically equivalent family of one-dimensional functions with chaotic members. By using this method however, there is no way to find out why these special members act in a chaotic manner.

By using f_{akt} qualitatively the same properties can be observed. To compensate the bigger gradient of f_{akt} one can use the weight matrix

$$\mathcal{W} = a \cdot \begin{pmatrix} -1 & 1 \\ -5 & 5 \end{pmatrix}.$$

3.3 Analysis of the behaviour

The main intention here is to give an insight into the principles leading to chaotic behaviour. We therefore take a closer look at the state transition function $\boldsymbol{f}$ composed of a linear mapping and a locally defined activation function being limited and increasing monotonously. The two of them can not cause chaotic behaviour separately. This behaviour must hence result from the interaction of the two components. The linear mapping can be analysed by determining the eigenvalues and eigenvectors. The activation function retains the ordering, only disorting the output of the linear mapping.

The weight matrix

$$\mathcal{W} = a \cdot \begin{pmatrix} -1 & 1 \\ -5 & 5 \end{pmatrix}$$

has eigenvalues $\lambda_1 = 0$ and $\lambda_2 = -4a$ with corresponding eigenvectors $e_1 = (1,1)^T$ and $e_2 = (1,5)^T$.

Any starting point is therefore mapped in one calculation step directly onto the line $\boldsymbol{G} = \{\boldsymbol{x} \in \mathbb{R}^2 \mid 5x_1 = x_2\}$, while its distance at $\boldsymbol{G}$ from the origin is scaled by a factor $-4a$ as shown in figure 2. The following application of the sigmoid activation function $\boldsymbol{f}_{akt}$ curves $\boldsymbol{G}$ in the direction of the main diagonal, which is reached at the points $(-1,-1)$ and $(1,1)$ for the distant point of $\boldsymbol{G}$.

After having demonstrated a case of formation of chaotic behaviour, one can easily formulate a suitably sufficient criterion on the bases of the knowledge acquired above. Firstly, at least one of the eigenvalues must be equal to zero to get a convergence independent of the parameter a in the direction of the origin.

Further, the curving caused by $\boldsymbol{f}_{akt}$ acts only in direction of the diagonals, so the space covered by eigenvectors with vanishing eigenvalues must include a diagonal, i.e. a vector with components contained in the set $\{-1,0,1\}$. Finally, a linear combination of the eigenvectors with non-vanishing eigenvalues must be in the same quadrant as the diagonal. This must be so, to guarantee the curving of this linear combination towards the right diagonal.

If we define the term $\frac{x}{|x|}$ to be zero if $x = 0$, we can then summarize the above considerations by the theorem below.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 1-6

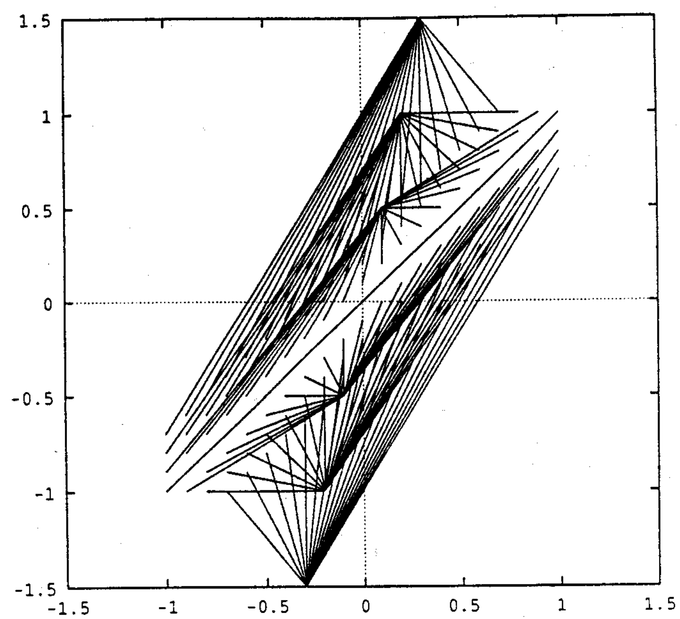

Figure 2: Demonstration of the effect of multiplying different starting points with $\mathcal{W}$.

Theorem 1 (chaos criterion) *A net with activation function* f_{akt} *and weight matrix* $\mathcal{W}$ *so that the eigenvectors* $e_1, \ldots, e_m, e_{m+1}, \ldots, e_n$ *have corresponding eigenvalues* $\lambda_1, \ldots, \lambda_m$ *with absolute value greater than zero and* $\lambda_{m+1}, \ldots, \lambda_n$ *equal to zero, can act chaotically if a linear combination of* $e_1, \ldots, e_m$ *reaches a quadrant whose diagonal is described by a linear combination of* $e_{m+1}, \ldots, e_n$*. Hence, if one can find suitable values* $x_1, \ldots x_n$ *to define*

$$\begin{aligned} y &= x_1 e_1 + \cdots + x_m e_m \\ z &= x_{m+1} e_{m+1} + \cdots + x_n e_n \end{aligned}$$

one can prove, that

$$z = \begin{pmatrix} \frac{y_1}{|y_1|} \\ \vdots \\ \frac{y_n}{|y_n|} \end{pmatrix} \quad \Longrightarrow \quad \textit{net acts chaotically} \qquad (1)$$

4 Conclusions

In this article we have demonstrated, under which circumstances chaotic behaviour can occur in neural networks. A sufficient criterion was formulated. Further investigations [4] have shown the general advantages of analysing weight matrices to get new insight into the behaviour of neural nets.

There are several articles postulating the practical use of neural nets acting in a chaotic manner. These articles are trying to put this effect into practice. But by considering the sensitive dependence of initial conditions characterising chaotic behaviour, we can see that in principle it may be possible to describe a given chaotic acting system by such a net. On the other hand it is obviously impossible to approximate the behaviour of a chaotic system by learning processes. The reason being that a tiny deviation from the optimal parameter setting can cause the net to act in an extremely different chaotic way.

References

[1] Devaney, R.: Dynamics of Simple Maps. In: Devaney, R., Keen, L. (eds.): Chaos and Fractals – The Mathematics Behind the Computer Graphics. Proceedings of Symposia in Applied Mathematics 39, American Mathematical Society (1989), S.1-24.

[2] Doyon, B., Cessac, B. et al.: Destabilization and Route to Chaos in Neural Networks with Random Connectivity. Advances in Neural Information Processing Systems 5 (1993), S.549-555.

[3] Ikeguchi, T., Adachi, M., Aihara, K.: Chaotic Neural Networks and Associative Memory. In Prieto, A. (ed.): Artificial Neural Networks – IWANN'91. Lecture Notes in Computer Science 540, Springer-Verlag, Berlin (1991), S.17-24.

[4] Kolb, T.: Mathematische Analyse des dynamischen Verhaltens rekurrenter Neuronaler Netze. Thesis, Forschungszentrum Informatik Karlsruhe (1993).

[5] Renals, S.: Chaos in Neural Networks. In Almeida, L.B., Wellekens, C.J. (eds.): Neural Networks – EURASIP Workshop 1990. Lecture Notes in Computer Science 412, Springer-Verlag, Berlin (1990), S.90-99.

[6] Sompolinsky, H., Crisanti, A., Sommers, H.J.: Chaos in Random Neural Networks. Physical Review Letters 61.3 (1991), S.259-262.

[7] Van der Maas et al.: A Note on Chaotic Behaviour in Simple Neural Networks. (1990).

[8] Verschure: Chaos-based Learning. Complex Systems (1991).

[9] Wang, X.: Period-Doubling to Chaos in a Simple Neural Network: An Analytical Proof. Complex Systems 5 (1991), S.425-441.