Journal of the International Neuropsychological Society (2000), **6**, 620–626.

Errors in a nonlinear graphic-semantic mapping task resulting from lesions in Boltzmann machine: Is it relevant to dyslexia?

AMIR B. GEVA, LIOR SHTRAM, AND SHAI POLICKER
Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, Beer Sheva, Israel

(RECEIVED August 26, 1999; REVISED October 1, 1999; ACCEPTED October 8, 1999)

Abstract

One of the most fascinating aspects of brain research is the subject of language. As in many other cases, the malfunctions that occur in different persons for various reasons give us insight on the mechanisms that support our ability to talk, read and listen. Following the work of Plaut and associates, we deal with the dyslexia disorder, which is the overall name for a large number of reading disorders. A Boltzmann machine neural network scheme was trained to implement the nonlinear mapping task of graphic representation into semantic representation, which may model the brain sections responsible for the translation of a written word into meanings and syllables. After training, various types of lesions were applied and the performance of the network was tested in order to measure the effect of each lesion on the error rate and type distribution that were detected. The system's errors were classified into several categories and the distribution of errors between the categories was studied. Using the simulations, it is demonstrated that a finite scheduling process in the Boltzmann machine causes the distribution of the network's errors to be unique and different from its expected error distribution. The phenomenon is given a mathematical explanation rooted in the statistical mechanics basics of the Boltzmann machine. Test results suggest the localization of certain reading functions within the network. Comparison is made to relevant types of dyslexia and shows resemblance in major symptoms as well as in certain known side effects. (*JINS*, 2000, *6*, 620–626.)

Keywords: Neural networks, Boltzmann machines, Nonlinear mapping, Simulated annealing, Limited scheduling, Lesions, Pruning, Errors analysis, Dyslexia

INTRODUCTION

The Reading System and Dyslexia

The widely accepted model of the human reading system was presented by Adams (1990). The model describes four specialized processors: graphic, phonetic, semantic, and context (Figure 1). The reading process utilizes two different pathways of translation: (1) graphic–phonetic–semantic (using the skills of the spoken language), and (2) graphic–semantic (imaging the word as a picture).

Dyslexia is a developmental disorder that characterizes the unexpected failure of a child to acquire the skills of reading. It is a name for a wide variety of reading disorders including (1) visual word–form dyslexia, (2) central dyslexia, (3) surface dyslexia, and (4) deep dyslexia. The most common disorders are surface and deep dyslexia. In the view of the Adams model of the reading system, deep dyslexia can be explained as a fault in the connection between the graphic and phonetic processors, while surface dyslexia can be explained as a fault in the connection between the graphic and semantic processors. These explanations give reasons for most of the disorder's symptoms, though there are some symptoms left unsolved.

Boltzmann Machines and Neural Networks

Artificial neural networks (ANN) are mathematical models motivated by the structure of neural cells. Wide spans of industrial applications utilize the ANN model, including pattern recognition, control systems, and adaptive filters. An interesting aspect of the ANN is the modeling of cognitive functions.

Reprint requests to: Dr. Amir Geva, Electrical and Computer Engineering Department, Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 84105, Israel. E-mail: geva@ee.bgu.ac.il

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

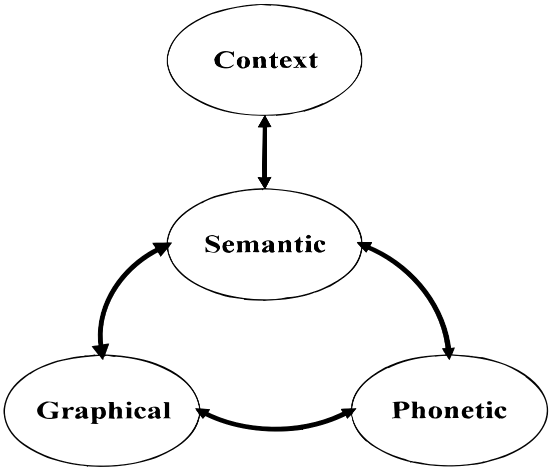

Fig. 1. The common model of the human reading system.

Boltzmann machines (Hinton & Sejnowski, 1983) are neural networks based on statistical thermodynamics. While the scheme is slower and more complex than others, more popular algorithms (like *back-propagation*), it has several features that make it more appealing for physiology-like nonlinear mapping tasks. The convergence (as well as the learning) algorithm of the system is a simulated annealing process that consists of continuous convergence cycles performed using a slowly decreasing random parameter called *temperature*. This random parameter enables the system to escape local minima and converge into the global minimum. The process is scheduled theoretically to reach the global minimum state at zero temperature. In actual implementations, however, practical reasons, like quantization and finite numbers handled by a computer, force the decrease in temperature to be neither continuous nor to reach zero value. Furthermore, the small residual final temperature can be regarded as a model for noise and for operation in extreme conditions.

Boltzmann machines have been applied to a number of problems including constraint satisfaction problems in vision (Hinton & Sejnowski, 1983), the encoder problem (Ackley et al., 1985; Parks, 1987), learning symmetries in two dimensions (Hinton & Sejnowski, 1986), statistical pattern recognition (Kohonen et al., 1988), and speech recognition (Lippmann, 1989). Although the algorithm is extremely slow, Boltzmann machines were found to be very effective. In a detailed comparison on a statistical decision tree (Kohonen et al., 1988), the Boltzmann machine achieved considerably better accuracy than a back-propagation network, and came close to the theoretical Bayes limit. Some specialized electronic and optoelectronic hardware has been developed for the Boltzmann machine.

In this work, we examine a Boltzmann machine implementation of a nonlinear graphic–semantic mapping task. The effect of the final temperature limitation on the system was specifically studied. In the first section of the introduction we define relevant parameters of the Boltzmann machine neural network. In its second section the nonlinear graphic–semantic mapping problem is described and the definitions of mistake categories are given. Different types of lesions were applied to the model and the simulation results are presented in the result section. Simulation results showed that the temperature limitation lesion causes a unique distribution of errors. The phenomenon receives a mathematical treatment in the appendix. We conclude with a discussion of possible applications and interpretations, which suggests that the results are valuable for performance enhancement and error detection and correction of the Boltzmann machine. The possible relevance of the work to understanding dyslexia is also discussed.

METHODS

The Boltzmann Machine

Boltzmann machines are made of *units* (neurons), which can be divided into three groups: *input*, *output* and *hidden* (Figure 2). The hidden units have no direct connection to the outside world, while the input units are set to the input vector on each activation of the network, and the output units will hold the output vector after a convergence process. The network is recurrent and the connections between units are symmetric $w_{i,j} = w_{j,i}$.

The units are stochastic, taking the value $S_i = +1$ with probability $g(h_i)$ or the value $S_i = -1$ with probability $1 - g(h_i)$, where h is defined:

$$h_i = \sum_j w_{i,j} S_j \tag{1}$$

and $g(h)$ is defined:

$$g(h) = \frac{1}{1 + e^{\frac{-2h}{T}}} \tag{2}$$

as the connections are symmetric, an energy function can be defined over the system. This energy function has minima whenever there is a stable state:

$$E_\alpha = -\frac{1}{2} \sum_{i,j} w_{i,j} S_i^\alpha S_j^\alpha . \tag{3}$$

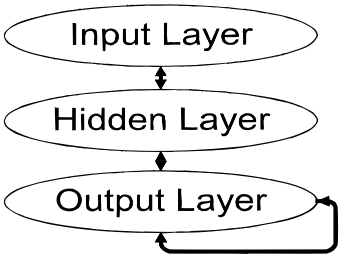

Fig. 2. The three layers of the Boltzmann machine.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

The probability of a change in the system's state is a function of the difference in the energies of the two states and is given by the Boltzmann distribution function:

$$P(S^{\alpha} \Rightarrow S^{\beta}) = P(E_{\alpha} - E_{\beta}) = \frac{1}{1 + e^{\frac{\Delta E}{T}}}. \tag{4}$$

Convergence of the system is achieved in a simulated annealing process, in which cycles of convergence are performed while the temperature parameter T is exponentially slowly decreased until reaching zero. When T reaches zero the system is completely deterministic; however, the exponentially slow decrease in the value of T (Geman & Geman, 1984) ensures that the system will converge into a global minimum of its energy function.

In order to speed up convergence on a computer simulation, it is common to use a mean field method, which is sometimes called deterministic Boltzmann machines. Using the mean field approximation method, we make the following assumption:

$$\langle S_i S_j \rangle \approx m_i m_j \tag{5}$$

where m is the mean of the activation of each unit:

$$m_i = \tanh\left(\frac{\sum_j w_{i,j} m_j}{T}\right). \tag{6}$$

Throughout the simulations that follow, the above approximation will be used.

The Model

A wide variety of tasks are analogous to the mapping task, especially when dealing with arbitrary mappings that can be nonlinear. Among those are coding and decoding, image recognition, semantic analysis and so on. We choose to view the problem with a semantic interpretation (Adams, 1990; Hinton et al., 1993; Plaut & Shallice, 1993).

Let M be a mapping from a set of vectors $V = \{v_i\}_{i=1}^{N}$ over the vector space G $(V \subset G)$ to another set of vectors $U = \{u_i\}_{i=1}^{N}$ over another vector space S $(U \subset S)$. Each element of the vector spaces is binary[1].

In our graphic–semantic mapping, the source vector space G is a graphic representation of a word. Its elements represent visual features of the written words, like diagonal lines, round objects, closed shapes and so on (Table 1). The target vector space S is a semantic representation of a word. Its elements represent semantic features of the word, like *good*, *big*, *green*, *heavy*, *mystical*, etc. (Table 2). The sets U and V are projections of the same set of words, in the graphic and semantic spaces respectively (Figure 3).

Table 1. Graphic representation of letters

	Graphic representation						
Letter	Side open	Top open	Bottom open	Round	Has angles	Horizontal lines	Vertical lines
A	0	0	1	0	1	1	0
B	0	0	0	1	1	1	1
C	1	0	0	1	0	0	0
D	0	0	0	1	1	1	1
E	1	0	0	0	1	1	1
F	1	0	1	0	1	1	1

The described mapping is nonlinear in the sense that the distance between two source vectors in the input vector space, tells us little or nothing about their distance in the output vector space, even though we demand that our mapping system apply some linearization over the stable points—which means that small changes in the input vector will result in changes as small in the output vector.

A Boltzmann machine was taught to realize the mapping M, in that, when given an input vector $v \in G$, the vector $u \in S$ produced by it is equal to M(v). Using the semantic analogy of the model, the Boltzmann machine was taught the task of reading–translation of the written word (the input vector $v \in V$) into concept (the output vector $u \in U$). The training process was performed until perfect performance of the system was achieved on the given vocabulary while in normal conditions. On the other hand, when applying lesions such as pruning or amplification, the system becomes error-prone.

The various mistakes produced will be classified into two main categories: *input-oriented* mistakes and *output-oriented* mistakes. An output mistake, analogous to a semantic mistake, occurs when the output vector u is not actually equal to $M(v)$, but is similar to it. An input or graphic mistake occurs when the output vector u is not close to the expected value $M(v)$, but is the mapping point $M(v')$ for another input vector v' similar to v as explained in Figure 4. Similarity is defined in the sense of Euclidean distance over the relevant vector space.

Table 2. Semantic representation of words

	Semantic representation							
Word	Inanimate	Fauna	Flora	Food	Fruit	Human	Family	Verb
Plum			+	+	+			
Tree			+					
Drum	+							
Dumb						+		
Ring	+							
Sing								+
Reef		+						
Gold	+							

[1] Meaning $G = Z^{D_G}$; $S = Z^{D_S}$; $Z = \{0,1\}$, thus the cardinal of each set is limited by: $|V| \leq 2^{D_G}$; $|U| \leq 2^{D_S}$ and is equal $|V| = |U| = N$.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

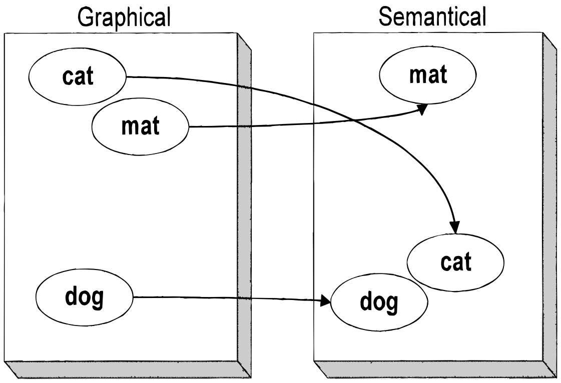

Fig. 3. An example of a nonlinear graphic–semantic mapping task.

The mistakes categories are more formally defined. The input vector $v \in V$, is given to the system and an output vector $u \in S$ is returned. If the u is sufficiently close to $M(v) \in U$:

$$d(u,M(v)) < \epsilon \tag{7}$$

and is closer to $M(v)$ from all other possible output vectors:

$$d(u,M(v)) \leq d(u,M(v_i)), \forall i \tag{8}$$

then we can say that the output is correct. Otherwise, a mistake category should be found. If the output vector u is sufficiently close to another possible output vector $M(v')$, and the two output vectors are close to each other:

$$d(u,M(v')) < \epsilon$$

$$d(M(v),M(v')) < d_o \tag{9}$$

then an *output* mistake has been made. On the other hand, if the output vector u is sufficiently close to another possible output $M(v')$ and its input vector v' is close to the original input vector v:

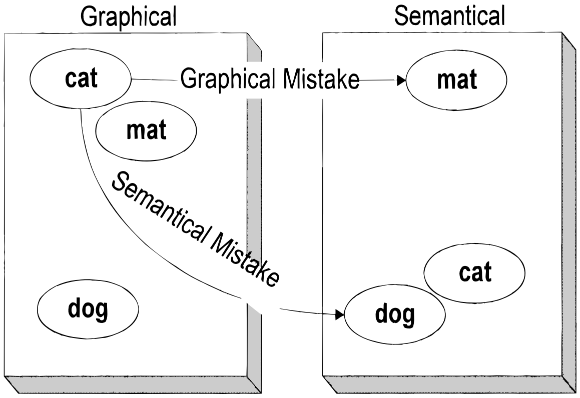

Fig. 4. Graphic *versus* semantic mistakes.

$$d(u,M(v')) < \epsilon$$

$$d(v,v') < d_i \tag{10}$$

then an *input* mistake has been made. An *input* mistake and an *output* mistake can also occur simultaneously, in which case we call it a *mixed* mistake. In a case where the mistake can not be categorized into any of the above, we assign it to the *other* mistake category.

The thresholds ϵ, d_o, and d_i are arbitrary; they adjust the trade-off between the false-detection of the system and its false alarms. Any setting of these parameters will set a working point to the system, to which all results should be related.

RESULTS

Different types of lesions, such as pruning of the connections between the layers, adding noise and bounding the lowest temperature, were applied to the network and the resulting errors were analyzed. The effect of scheduling limitation by stopping convergence before reaching zero temperature was demonstrated using several different simulations, in order to generalize the results. Here we bring one such simulation example followed by a mathematical treatment of the problem.

Simulation

A Boltzmann machine was taught the task of translating a word from its graphic representation to its semantic representation. The words were taken from a limited vocabulary of 43 words (Table 2). Each word consisted of three to four letters. Each element in the graphical representation of a word described a visual feature like *side open*, *top open*, and *horizontal lines* (Table 1). The elements of the semantic representation included semantic features like *inanimate*, *fauna*, etc. (Table 2).

The network, constructed of 128 neurons, is divided into three layers: an input layer with 48 neurons, a hidden layer with 40 neurons and an output layer with 40 neurons (Figure 2). Several connection schemes were tested and results proved to be independent of any specific architecture. The network was trained with a vocabulary of 43 English words of three to four letters each, until a near-zero error rate in the semantic translation was achieved and no mistakes were made over the given set of words. The vocabulary is far from being uniform and as such is very biased as regards distance between words in each projection space (semantic, graphic). The inherent grouping of the vocabulary was calculated and is presented in Figure 5. As we can see there is a much greater probability for graphic (input) errors than for semantic (output) errors[2].

[2] The inherent grouping is very dependent on the definitions of our group criteria; nevertheless, upon defining these criteria, we set a working point that all our results will be viewed with respect to.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

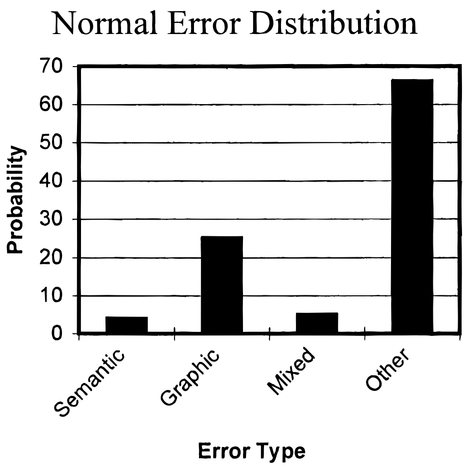

Fig. 5. Distribution of various error types.

Next, a series of lesions were applied on the network to model different kinds of defects. As we describe in details the effect of temperature lesions, it is valuable to consider, as a reference, the effect of other types of lesions on the same system. The lesion types that were applied to the network include: (1) *pruning*: cutting out some connections or cells; (2) *noise*: Changing the values of activation or connection weights; and (3) *temperature limitation*: Stopping the system's convergence before reaching the final stage. The various tests differ in the outcome distribution of errors, and can be explained in terms of the system's dynamics. The distributions of errors resulting from the pruning of connections between the different layers are shown in Figure 6. The frequency of *other* errors is lower than expected, but the graphic–semantic errors are close to the inherent probability.

The most extreme variation in the error distribution (from the inherent one shown in Figure 5) was found in the temperature limitation case and is shown in Figure 7. The next section and the Appendix focus on the temperature limitation, which gives insight to the system's dynamics and its robustness features. In this case we receive a totally different distribution, graphic errors are sparse or nonexistent, while semantic errors are common as is displayed in Figure 7. This occurs while the temperature is in a range close to the normal final temperature; larger temperature lesions result in the total malfunctioning of the system and produces almost no correct answers—in this case, the distribution of the error types is closer to the inherent one.

Simulation of a General Nonlinear Graphic–Semantic Mapper

In this realization, the network is constructed of three layers and a total of 18 neurons. An input layer which con-

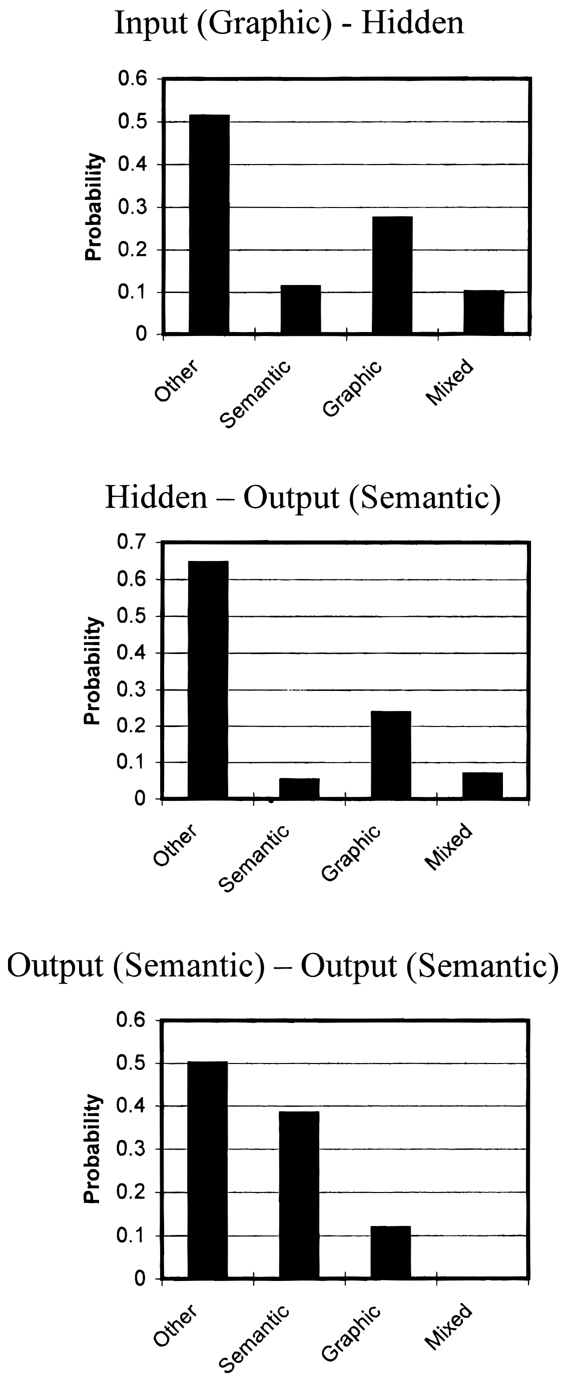

Fig. 6. Effect of connection pruning between the different layers.

sisted of five neurons, a hidden layer of eight neurons and an output layer of five neurons. Connections between the layers were full between the input and hidden layers and between the hidden and output layer, while there are no direct connections between the input and output layer. The network was taught a mapping of 10 input–output vector pairs and the task was learned successfully after approximately 100 iterations.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

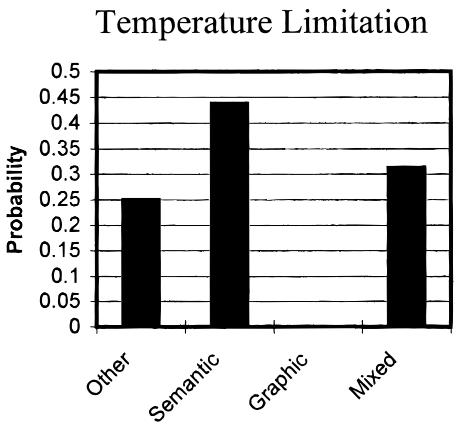

Fig. 7. Error distribution under temperature lesion.

Again, various lesions were tested. In this case we choose a mapping that is less biased in respect to the distance between the various vectors in each vector space, which is important to the generalization of the results. The inherent error distribution in this system is close to the uniform distribution; as the mapping is quite arbitrary, chances for an input or an output mistake are similar. The system indeed showed distributions close to the uniform one for all types of lesions, though the slight variations between the lesions applied can show localization of tasks. Yet, the temperature limitation lesion again showed extreme results, in which output mistakes are made much more often than input ones. The mathematical analysis of this interesting case is given in the Appendix.

DISCUSSION

In Boltzmann machines, as well as in other neural networks, robustness and resistance to damage is a noted fact. It is said that applying a small change to the parameters of the network usually produces only a small degradation in its performance—a feature hardly implemented in any arbitrary system. The robustness of the Boltzmann machine is particularly interesting when using temperature as the damaging parameter, while this might indicate time saved in a simulation, quantization error in a digital system, noise on a VLSI circuit, or the effect of stress and drugs in a biological system.

Extensive testing was carried out on the network in order to examine its functioning under different types of lesions. The major points that stood out in the tests were (1) the error probability is proportional to the extent of the lesion at all sites; (2) lesion of the graphic intermediate connection increases all types of errors dramatically as it blocks the input from the system; (3) by analyzing the corresponding error types it was possible to interpret the major processing task of each connection path, as follows:

A. Lesions in the graphic–hidden path increase the graphic errors that occur right at the beginning of the network convergence process. It seems that this path is related to the separation between visually similar words, that is, it maps visually similar words into different inner representation.

B. The hidden–semantic path corresponds to the process of convergence into the pretrained semantics. Lesion in this path increases the semantic errors that occur at the end of the network convergence process.

C. The semantic internal connections are responsible for the process of convergence into the exact semantic representation. Lesions in this path lead to a moderate increase in the semantic error rate, but the recognition is late and not clear. It suggests that this path is responsible for the final conversion into the right attractor. Lesion of these internal high level connections may explain some symptoms of slowness and bad prediction of the word's meaning in dyslexia.

D. Limiting the minimal temperature of the network causes mainly semantic errors and few or no graphic errors. This is perhaps the most surprising and interesting result of this research. It may simulate human reaction under mental stress or attention deficit, where the automatic functions, such as the visual recognition tasks, are functioning but the higher functions are disrupted. The temperature limitation shows that a diffuse (nonlocalized) damage that affects the network's convergence process may lead to specific semantic errors, which are well-known symptoms of dyslexia.

The results described here have relevance for evaluating the robustness of the Boltzmann network, as well as for understanding the convergence process in both normal and damaged operation. For the artificial network, we can use the information in the error distribution to correct errors and to accelerate the convergence process by eliminating its final stage.

The Boltzmann machine can be viewed as an information channel with a specific input and output coding. We have demonstrated that in respect to the temperature limitation, the system is much less sensitive to its input than to its output mapping. For the cognitive modeling of the reading system, the temperature limitation may represent mental stress or learning disruptions, and the unique error distribution may be analogous to the degradation of higher mental skills. In all cases, extreme lesions lead to total failure of the system while modest damage leads to dyslexia-like symptoms, which fits the "minimal damage" assumption in dyslexia. In summary, this stage of the research can only give us limited insight into the neural mechanism of dyslexia; the model should be adjusted and tested according to further results of neurophysiological and neuropsychological experiment in order to achieve a deeper understanding of the disorder.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press

REFERENCES

Ackley, D.H., Hinton, G.E., & Sejnowski, T.J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9*, 147–169.

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration in images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 721–741.

Hinton, G.E. & Sejnowski, T.J. (1983). Optimal perceptual inference. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 448–453). New York: IEEE.

Hinton, G.E. & Sejnowski, T.J. (1986). Learning and relearning in Boltzmann machines. *Parallel Distributed Processing*, *1*, 282–317.

Hinton, G.E., Plaut, D.C., & Shallice, T. (1993, October). Simulating brain damage. *Scientific American*.

Kohonen, T., Barna, G., & Chrisley, R. (1988). Statistical pattern recognition with neural networks: Benchmarking studies. *Proceedings of the IEEE International Conference on Neural Networks*, San Diego, I, 61–68, New-York: IEEE.

Lippmann, R.P. (1989). Review of neural networks for speech recognition. *Neural Computation*, *1*, 1–38.

Parks, M. (1987). Characterization of the Boltzmann Machine Learning Rate. In M. Caudill & C. Butler, III (Eds.), *IEEE First International Conference on Neural Networks* (pp. 715–719). New York: IEEE.

Plaut, D.C. & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology*, *10(5)*.

Appendix

MATHEMATICAL ANALYSIS OF THE TEMPERATURE LIMITATION ERRORS

From the description of the Boltzmann machine basics it is clear that when the temperature is higher than zero and the system is at a global minimum of the energy function, there is still a chance for a flip of any unit's state that will increase the total energy. So, for a case where the system's output is sampled at a low, positive temperature, an error is possible[3].

Two points in the state space that have the same energy seem to have also the same Boltzmann probability, so this substantial difference of probabilities is not trivial. It's only when we take into consideration the dynamics of the convergence process that we can justify this behavior. Observing the Boltzmann function we see that in the continuing convergence process toward the global minimum there is an increasing probability (along with temperature decline) of the global minimum state, so when dealing with the final stages of the process, it can be assumed that to reach every other state, different from the minimum, the system must go through a path of flips from the global minimum to that different state. In this way, the time parameter, together with the distance from the minimum state, plays an important role in the probability calculation.

Consider an ensemble of states with an energy that is ΔE higher than the minimum (as shown in Figure 8). The path of N flips that has to be taken from the minimum state will have a probability of:

$$P\left(N,\{\Delta E_i\}_{i=1}^{N} \,\middle|\, \sum_{i=1}^{N} \Delta E_i = \Delta E\right) = \prod_{i=1}^{N} \frac{1}{1+e^{\frac{\Delta E_i}{T}}} \tag{11}$$

It can be shown that for all paths $\{\Delta E_i\}_{i=1}^{N}$ with total energy ΔE of length N are always more probable then a paths of length M where $M > N$—shorter paths are more probable. When we examine the probability expression it is easy to show with Lagrange multipliers that the minimal probability will be achieved when all the energy deltas of all flips are equal, and that the maximal probability will be achieved when all energy deltas but one equal zero (and one that equals ΔE):

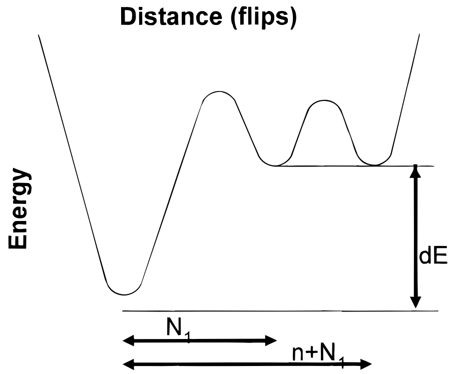

Fig. 8. Two points of equal energy but different distance from global minimum.

$$\min\{P(N,\{\Delta E_i\}_{i=1}^{N})\} = \left(\frac{1}{1+e^{\frac{\Delta E}{NT}}}\right)^{N} \tag{12}$$

$$\max\{P(N,\{\Delta E_i\}_{i=1}^{N})\} = \left(\frac{1}{2}\right)^{N-1}\left(\frac{1}{1+e^{\frac{\Delta E}{T}}}\right) \tag{13}$$

Replacing $C = e^{\frac{\Delta E}{NT}}$ and comparing the minimal probability of an N flips route and maximal probability of an $N+1$ flips route we get:

$$\frac{\min(P_N)}{\max(P_{N+1})} = \frac{2^N(1+C^N)}{(1+C)^N} > 1 \tag{14}$$

$$\min\{P(N,\{\Delta E_i\}_{i=1}^{N})\} > \max\{P(N+1,\{\Delta E_j\}_{j=1}^{N+1})\} \quad \forall N,\{\Delta E_i\}_{i=1}^{N},\{\Delta E_j\}_{j=1}^{N+1} \tag{15}$$

which means that output errors (shorter paths) are more probable than error types that require longer paths.

[3] Note, however, that some flips from the global minimum state do not necessarily mean an error, as the flips can all take place in the hidden units; moreover, a small distance from the correct result will not be considered a mistake.

https://doi.org/10.1017/S135561770065511X Published online by Cambridge University Press