Selective enhancement of recall through plasticity modulation in an autoassociative memory

A. Sandberg[a,*], A. Lansner[a], K.M. Petersson[b]

[a]KTH Nada, NADA/SANS, Royal Institute of Technology, Lindstedvägen 3, 100 44 Stockholm, Sweden
[b]Cognitive Neurophysiology, Karolinska Institutet, 171 77 Stockholm, Sweden

Abstract

The strength of a memory trace is modulated by a variety of factors such as arousal, attention, context, type of processing during encoding, salience and novelty of the experience. Some of these factors can be modeled as a variable plasticity level in the memory system, controlled by arousal or relevance-estimating systems. We demonstrate that a Bayesian confidence propagation neural network with learning time constant modulated in this way exhibits enhanced recall of an item tagged as salient. Proactive and retroactive inhibition of other items is also demonstrated as well as an inverted U-shape response to overall plasticity. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Memory modulation; von Restorff effect; BCPNN

1. Introduction

Long-term memory (LTM) formation in everyday life often occurs incidentally without explicit intention to remember the information processed. It has been suggested that memory formation is the dynamic consequence of information processing and system plasticity [10]. Research indicates that specific kinds of information processing contributes to LTM-formation, including meaning-based, context and relational processing and factors like emotional significance and attentional allocation (for recent reviews see e.g. [15]).

* Corresponding author. Tel.: + 46-8-790-69-09; fax: + 46-8-790-09-30.
E-mail address: asa@nada.kth.se (A. Sandberg).

Endogenous processes activated by experience can modulate memory strength in terms of recall probability [7]. For example, emotionally arousing [1] or humorous [13] experiences are generally better remembered than less affective experiences, and hormones and neuromodulators can affect how strongly experiences are retained [6]. The novelty of a stimulus also plays an important role. The von Restorff effect consists of improved recall or recognition of an item (the isolate) that is distinct or different from the others in a set, while the other items are less well recalled (retroactive and proactive inhibition) [14]. While this has mainly been studied in human list recall, a similar effect has been observed in rats [11] and monkeys [8].

Some of these factors can be interpreted in the framework of memory consolidation as a relevance modulation of the "print-now" signal by regulating memory encoding and synaptic plasticity. During ordinary events the plasticity is lower than when salient, arousing or motivational stimuli increase the learning rate, for example through a modulating mechanism. Such a mechanism relates closely to neuromodulation in the brain, e.g., the effect of dopamine [16] and acetylcholine in synaptic plasticity [17,3].

We have previously studied the properties of an autoassociative Bayesian confidence propagation neural network (BCPNN) [5,4] with incremental learning [12]. The BCPNN is based on a probabilistic view of learning and retrieval, with input and output unit activities roughly representing confidence of feature detection and posterior probabilities of outcomes, respectively. Besides in ANN models, it has also been used for simulation of cortical dynamics and cell assemblies [2].

This learning rule has a time constant of learning that determines how quickly it will adapt to new information. By modulating this time constant, we can model the modulatory regulation of the print-now signal on associative encoding of information into for example LTM.

This paper describes a simple model of an autoassociative network with plasticity modulation for one item, and shows that it can produce the enhanced recall of the isolate, proactive and retroactive inhibition and an inverted U-shape response curve to overall plasticity similar to the one commonly observed in arousal-performance or dose-response plots.

2. Network

We have used the BCPNN with incremental learning and hypercolumns [12] which is based on a Hebbian learning rule derived from Bayes rule [5]. It is an autoassociative network with a learning rule where weights are updated depending on estimates of unit activity regulated by a learning time constant. Units are combined into hypercolumns with normalized output activities (corresponding to mutually exclusive or combinations of otherwise dependent features [4]).

The activity update and equations are

$$\frac{\mathrm{d}h_{ii'}(t)}{\mathrm{d}t} = \beta_{ii'}(t) + \sum_{j}^{N} \log\left(\sum_{j'}^{M_j} w_{ii'jj'}(t)o_{jj'}(t)\right) - h_{ii'}(t),$$

$$o_{ii'}(t) = \frac{\mathrm{e}^{h_{ii'}}}{\sum_j \mathrm{e}^{h_{ij}}},$$

$$\frac{\mathrm{d}\Lambda_{ii'}(t)}{\mathrm{d}t} = \kappa(t)\alpha([(1-\lambda_0)o_{ii'}(t) + \lambda_0] - \Lambda_{ii'}(t)),$$

$$\frac{\mathrm{d}\Lambda_{ii'jj'}(t)}{\mathrm{d}t} = \kappa(t)\alpha([(1-\lambda_0^2)o_{ii'}(t)o_{jj'}(t) + \lambda_0^2] - \Lambda_{ii'jj'}(t)),$$

$$\beta_{ii'}(t) = \log(\Lambda_{ii'}(t)),$$

$$w_{ii'jj'}(t) = \frac{\Lambda_{ii'jj'}(t)}{\Lambda_{ii'}(t)\Lambda_{jj'}(t)},$$

where $h_{ii'}$ is the potential of the ith unit in hypercolumn i, $o_{ii'}$ the output of the unit, β the bias, w the weights, Λ are rate estimates, λ_0 a low background activity (set to 10^{-4} throughout this paper), N the number of hypercolumns, M_j the number of neurons in hypercolumn i and α the inverse of the learning time constant. Low α correspond to slow learning and forgetting, high α fast learning and forgetting.

$\kappa(t)$ is the relevance signal, assumed to be sent from a modulator system as a response to the current experience. In the following experiments, it was kept at $\kappa(t) = 1$ except for one pattern in the training set, the isolate, where $\kappa(t)$ was set to κ_i during training.

3. Results

A 100 neuron BCPNN network with 10 hypercolumns of 10 neurons each was trained by clamping unit activity to each training pattern ξ^p for one unit of time, allowing the weights to adapt. Retrieval was tested by activating the neurons with a trained pattern where the activity in three hypercolumns of ten had been randomized, and then allowed to relax for one unit of time (no learning was used during testing). Performance was measured by the overlap $\xi^p \cdot \mathbf{x}/|\xi^p|\,|\mathbf{x}|$.

Fig. 1(a) shows the mean overlap for all patterns as a function of the learning time constant α, with no isolate element ($\kappa_i = 1$). At low α, the network learns too slowly to learn the patterns, while at high α the network learns quickly but forgets the oldest patterns.

Fig. 2 shows the selective enhancing effect on recall when an isolate pattern occurs. Note the inhibition of recall of other patterns in the high κ_i condition. α was set to 10^{-8} in this and the following plots.
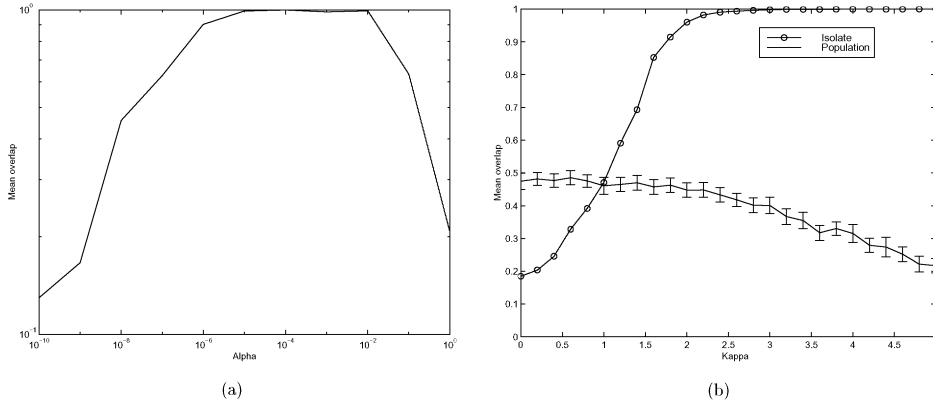
(a)   (b)

Fig. 1. (a) Mean overlap after convergence from noisy cue as a function of the inverse learning time constant α. Below $\alpha = 10^{-6}$ the network learns too slowly to store the shown patterns, and above $\alpha = 10^{-2}$ it forgets too quickly to store all patterns. (b) Mean overlap after convergence from noisy cue of the isolate pattern and the other patterns as a function of the increase in plasticity κ_i.
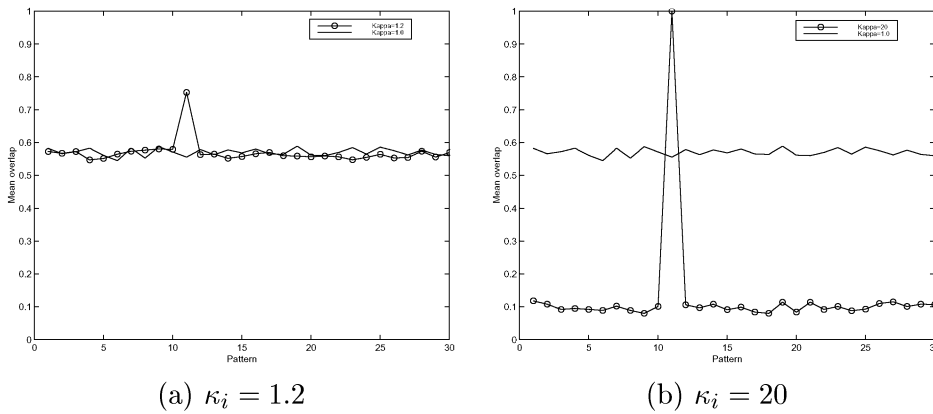

(a) $\kappa_i = 1.2$ (b) $\kappa_i = 20$

Fig. 2. Mean overlap after convergence from noisy cue with or without an isolate pattern. In the dotted run, κ_i was set to 1.2 (left) or 20 (right) for pattern 11.

Fig. 1(b) shows the mean overlap as a function of the modulation strength for the isolate and for the normal items. As κ_i increases the recall of the isolate becomes better and better, while there is an inhibition effect on the other items. However, it is possible to avoid inhibition in this model for low levels of the relevance signal while still observing a recall enhancement for the isolate.

When the network is stimulated by a random pattern it will converge to a given attractor state with a probability depending on the relative volume of the basin of attraction to the volume of the state space. Again the isolate is more likely to be

(a) $\kappa_i = 2,\ \alpha = 10^{-2}$ (b) $\kappa_i = 20,\ \alpha = 10^{-1.5}$
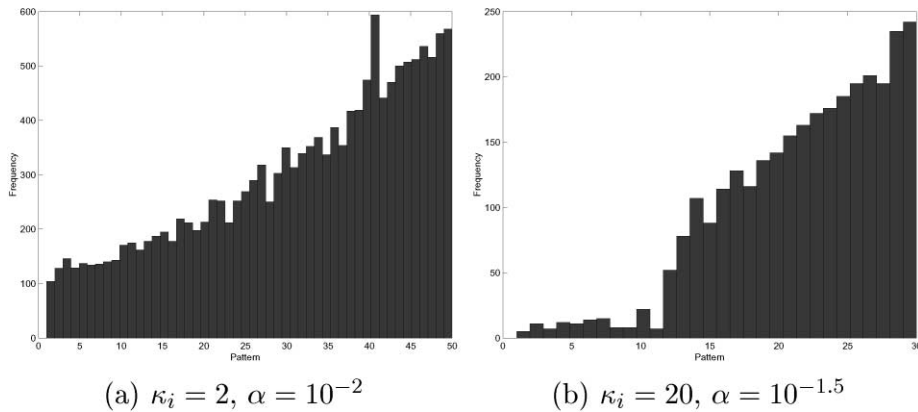
Fig. 3. Frequency of ending up in the attractors corresponding to the learned patterns when activating the network with a random pattern (one randomly activated unit per hypercolumn) for a high level of plasticity. The isolate is pattern 41 (left) and 11 (right).

recalled, suggesting that for these parameters the increased plasticity has enlarged its basin of attraction relative to the other attractors.

If the network has a shorter learning time constant as in Fig. 3, there will be a memory gradient due to fast forgetting. The increase in plasticity caused by a strong modulatory input can prevent encoding of the isolate and retroactively interfere with the early patterns without affecting the subsequent storage of the network, as can be seen in the right subfigure.

4. Discussion

We have described a simple model of an autoassociative memory where the learning rate is modulated by a relevance signal and shown that it can exhibit selective enhancement of items associated with a relevance signal.

We have shown that a model of an autoassociative memory with modulated learning rate can exhibit selective enhancement of items associated with a relevance signal. When the signal is strong, inhibition of other items occurs as well as a relative expansion of the isolate item basin of attraction.

The inverted-U curve in Fig. 1 has similar properties as the Yerkes–Dodson law or the observed inverted-U dose-response relationship seen for many memory enhancing drugs [6,9].

If there is a need to encode relevant experiences at different levels of trace strength, for example to distinguish between everyday and important stimuli, the baseline plasticity needs to be in a range where the increase in plasticity due to the relevance signal produces a positive change in encoding success. If the memory exhibits a U-shaped response to plasticity, this suggests that the baseline plasticity is not close

to the maximum of the overlap curve but below it, and that an increased level of baseline arousal would decrease the distinction between relevant and irrelevant memories.

Interestingly, a strong modulatory input increases the plasticity in this model to the extent that the overall effect becomes a memory impairment rather than enhancement.

Acknowledgements

References

[1] S.-A. Christianson (Ed.), Handbook of Emotion and Memory: Current Research and Theory, Erlbaum, Hillsdale, NJ, 1992.

[2] E. Fransén, A. Lansner, A model of cortical associative memory based on a horizontal network of connected columns, Network 9 (1998) 235–264.

[3] M.E. Hasselmo, B.P. Wyble, G.V. Wallenstein, Encoding and retrieval of episodic memories: role of cholinergic and gabaergic modulation in the hippocampus, Hippocampus 6 (6) (1996) 693–708.

[4] A. Holst, The use of a Bayesian Neural Network Model for Classification Tasks, Ph.D. Thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, Sweden, September 1997, TRITA-NA-P9708.

[5] A. Lansner, Ö. Ekeberg, A one-layer feedback artificial neural network with a Bayesian learning rule, Int. J. Neural Systems 1 (1) (1989) 77–87.

[6] J.L. Martinez, G. Schulteis, S.B. Weinberger, Learning and Memory: A Biological View, 2nd Edition, Academic Press, New York, 1991, pp. 149–198 (Chapter 4).

[7] J.L. McGaugh, Memory—a century of consolidation, Science 287 (2000) 248–251.

[8] A. Parker, E. Wilding, C. Akerman, The von restorff effect in visual object recognition memory in humans and monkeys: the role of frontal/perirhinal interaction, J. Cogn. Neurosci. 10 (6) (1998) 691–703.

[9] M. Parsons, P. Gold, Glucose enhancement of memory in elderly humans: an inverted-u dose-response curve, Neurobiol. Aging 13 (3) (1992) 401–404.

[10] K. Petersson, C. Elfgren, M. Ingvar, Dynamic changes in the functional anatomy of the human brain during recall of abstract designs related to practice, Neuropsychologia 37 (1999) 567–587.

[11] P. Reed, A. Richards, The von restorff effect in rats (*Rattus norvegius*), J. Comp. Psychol. 110 (2) (1996) 193–198.

[12] A. Sandberg, A. Lansner, K.-M. Petersson, Ö. Ekeberg, A palimpsest memory based on an incremental Bayesian learning rule, Neurocomputing 32–33 (2000) 987–994.

[13] S. Schmidt, The effects of humor on sentence memory, J. Exp. Psychol. Learning, Memory Cogn. 20 (1994) 953–967.

[14] H. von Restorff, Analyse von vorgängen in spurenfeld, Psychol. Forsch. 18 (1933) 299–342.

[15] A. Wagner, W. Koutstaal, D. Schacter, When encoding yields remembering: insights from event-related neuroimaging, Phil. Trans. R. Soc. London B 121 (1999) 1307–1324.

[16] J. Wickens, R. Kötter, Cellular models of reinforcement, in: J.C. Houk, J.L. Davis, D.G. Beiser (Eds.), Models of Information Processing in the Basal Ganglia, MIT Press, Cambridge, MA, 1995, pp. 187–214.

[17] N.J. Woolf, The critical role of cholinergic basal forebrain neurons in morphological change and memory encoding: a hypothesis, Neurobiol. Learn Mem. 66 (3) (1996) 258–266.

Anders Sandberg has a M.Sc. in Computer Science from Stockholm University. He is pursuing a Ph.D. at the SANS group at the Department of Numerical Analysis and Computing Science of the Royal Institute of Technology in Stockholm. His main research interest involves the interplay between the medial temporal lobe and neocortex in memory consolidation, and how this relates to modulation of neural plasticity.

Anders Lansner received his M.Sc. in Engineering Chemistry and Biochemistry in 1974 and his Ph.D. in Computer Science from the Royal Institute of Technology in 1986. He was appointed associate professor and professor in Computer Science in 1992 and 1999 respectively. He is project manager for the SANS group (Studies of Artificial Neural Systems), which he founded in 1987. His research interests range from neural computation and control to detailed computational models of specific biological systems and functions.

Karl Magnus Petersson received his B.Sc. in Mathematics and Physics in 1988 at the University of Stockholm and his M.D. at the Karolinska Institutet in 1996. He is currently pursuing a Ph.D. at the Cognitive Neurophysiology group at the Department of Clinical Neuroscience of the Karolinska Institute. His research interests range from artificial neural computation to the functional organization of the human brain. Current work focuses on memory and learning as well as language function in the human brain.