

Novelty based feedback regulation in artificial neural networks

Agnieszka Rychwalska¹, Piotr Jabłoński², Michał Żochowski³
and Andrzej Nowak³

¹Department of Psychology; ²Institute for Social Studies, University of
Warsaw, 5/7 Stawki St., 00-183 Warsaw, Poland; ³Warsaw School for Social
Psychology, 19/31 Chodakowska St., 03-815 Warsaw, Poland

Abstract. In this paper we present a theoretical framework for novelty based feedback regulation in artificial neural networks. Novelty is assessed on the basis of monitoring the coherence of network dynamics. The result of novelty detection is dynamically coupled to parameters that control the dynamics of the recognition process. The paper presents a new measure of novelty detection – the strength of the local field – and presents new simulation results concerning novelty detection. It also integrates previously published models and simulation results into a general dynamical model of feedback regulation.

The correspondence should be
addressed to A. Nowak,
Email: nowak@fau.edu

Key words: novelty, feedback regulation, coherence, neural networks, local field, Hopfield model, neuronal flips

INTRODUCTION

Novelty in the environment carries information that needs to be decoded in order to provide the individual with better chances of adaptation. Often it is crucial for survival. It is therefore natural that the nervous system evolved in such a way that novelty detection is fast and reliable. The effects of this process have been extensively researched but little is known about its mechanisms.

On the behavioural level, the effects of novelty are most easily seen as an improved fluency and efficiency of processing of repeatedly presented stimuli (Habib 2001). This means that stimuli, once they become familiar, are processed differently from novel ones. On the neural level, single unit recordings reveal that repeated presentation causes a reduction of neural activity – which again implies that with novel stimuli cell activation is elevated. This effect has been witnessed in many experimental designs (Gilbert et al. 2001, Henson and Rugg 2003, Miller et al. 1991) and may last for different periods of time (Sanchez-Vives et al. 2000, Schoups et al. 2001, Sobotka and Ringo 1994). Therefore it was termed either repetition suppression or adaptation. However, Ranganath and Rainer (2003) propose that in fact these are instantiations of the same cellular mechanism. Reduced neural activity to repeated stimuli has also been observed as a decrease in metabolic changes as witnesses in fMRI studies (Grill-Spector and Malach 2001, Huk et al. 2001).

Another well researched effect of novelty is the P300 component in the event-related potential (ERP) recordings (Sutton et al. 1965), which, at the behavioural level, correlates with an orienting reflex (Sokolov 1963, 1969). The P300 potential is composed of several subcomponents of which the P3a has been directly linked to novelty processing (Friedman et al. 2001, Knight 1984, Squires et al. 1975). The orienting reflex may be thought of as a vivid indication (e.g., orienting of the body) of deploying cognitive resources to process the novel stimulus. Novelty is also directly linked to emotional response – it was shown in mere-exposure paradigm that known stimuli elicit positive affective response while novel stimuli evoke negative affect (Zajonc 1980).

All these cases provide examples of feedback regulation in the nervous system. In all of them the novelty of the stimulus is assessed early in perception – even before the actual recognition – and it changes the way in which the neural system processes information.

The notion of feedback is traditionally related to constant modulation of the process by the results of the process itself and is often conceptualized as a functional link leading from output to input. This notion does not capture however important properties of novelty based feedback regulation, which requires that novelty is detected and the regulation takes place before the output is reached.

Although both the principles of feedback regulation in general and the effects of novelty of stimuli on the processing of these stimuli have been well documented experimentally, the specific model describing feedback regulation based on novelty of the stimulus is still missing. Such a model should describe how novelty of the stimulus is detected in early stages of stimulus processing, and how the detection of novelty influences the functions of the nervous system.

GENERAL APPROACH

Previous work investigated specific measures of dynamical recognition of novelty and specific models of feedback based regulation (Lewenstein and Nowak 1989a, b, Vallacher and Nowak 1999, Żochowski et al. 1994). Another line of research related the measurements to the emotional response (Drogosz and Nowak 1996, Winkielman and Nowak 2005, Winkielman et al. 2002). In the current paper we present a novel criterion for fast novelty detection and a general model of novelty based feedback regulation.

Our model is based on the assumption that pre-recognition of novelty, taking place in the very first moments of stimuli processing, affects the dynamics of recognition. A schematic course of this process is presented in Fig. 1. We assume that early dynamics of recognition carries information that can be decoded and fed back into the system, so that it affects the recognition process itself. As can be inferred from the figure, there are two crucial steps here – monitoring the initial response of the system and modulating the dynamics of the network. The first is responsible for the detection of dynamical response of the system based on type of the incoming stimuli (i.e., familiar vs. novel), the other for implementing this information.

Novelty based feedback regulation may thus be understood as modulating the properties of the system depending on its own state or dynamics during infor-

mation processing. To propose such a feedback regulation model we need to identify what dynamics characteristics indicate processing of a novel stimulus and how they may affect the recognition process.

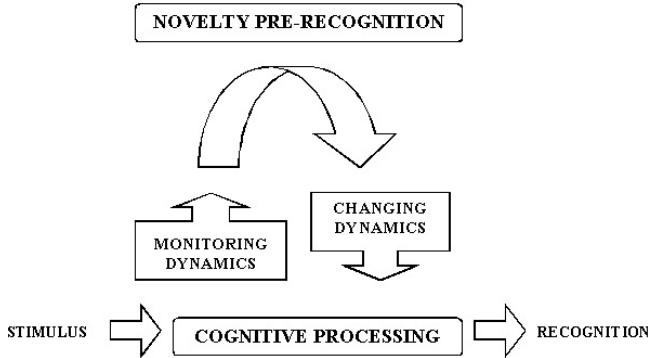


Fig. 1. Schematic course of novelty based feedback regulation.

We propose that novelty pre-recognition is based on dynamical measurement of coherence of the signal in the information processing system: low coherence indicates novelty, whereas high coherence indicates familiarity to the incoming stimulus. The level of coherence in turn, affects global control parameters of the system and therefore forms feedback modulation of its dynamics. We will describe here two possible measures of coherence in artificial neural networks – frequency of state changes and local field – and identify the control parameters of the system they might modulate to qualitatively change the dynamics of the network: signal to noise ratio and connection plasticity. We will present examples of implementation of these measures in attractor neural networks. Our simulations indicate that novelty can indeed be assessed very early during information processing, and it might in turn modulate this process to obtain qualitatively different response of the system. Before we present our models and simulation results, we will briefly discuss the assumptions of attractor networks, as this type of architecture was used in the models.

Attractor neural networks

Simplified network models are usually constructed in very rough analogy to their biological counterparts. Neurons, or in the case of psychological applications blocks of neurons, are modeled as relatively simple input-output elements (McCulloch and Pitts 1943).

Such elements, which we shall call "neurons" for simplicity, are connected with each other through the so called synaptic connections. The efficacy of synaptic connection of the j -th neuron to the i -th neuron is described by a real number J_{ij} , that measures an effective influence of j -th neuron on the i -th neuron. The dynamics of such a net, can be described as a process of adjustment of the state of a given neuron to the influence that it experiences from different sources: other neurons or external stimuli (Amit 1989). The most important property that distinguishes network models from other models is distributed and parallel character of information processing and emergence of complex collective effects.

The class of networks used in models presented here may be termed attractor neural networks (Rumelhart and McClelland 1986). These kind of networks, in contrast to feed-forward ones, are characterized by a massive appearance of feedback loops among its elements and no clear direction of information flow. External stimuli determine the initial state of the network. The evolution of the network is determined by network connectivity and neurons' response function (i.e., specific rules describing neuron's response to incoming signal). Dynamics of the classic Hopfield network is described by

$$S_i^{t+1} = \text{sign} \left(\sum_j S_j^t J_{ij} \right) \quad (1)$$

where S_i^t is the state of i -th neuron at time t .

During its evolution, after some time the network approaches an asymptotic state, usually a stationary (fixed) point where the state of the network does not evolve over time.

For appropriately constructed connections the fixed point represents one of the stored configurations in the network (i.e., memories). This type of networks serves as associative memory model and they can recognize highly distorted patterns. The memory retrieval is very robust with respect to different kinds of perturbations, such as removal of connection links, noise etc. One of the most known learning algorithms is the Hebb's rule, which stores statistically independent configurations and is given by:

$$J_{ij} = \frac{1}{p} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} \quad (2)$$

where J_{ij} is strength of the synapse between i -th and j -th neuron, ξ_i^μ is state of the i -th neuron of the μ -th learning pattern, and p is the number of learning patterns. The algorithm represents a well known neurophysiological notion that coupling is formed between coactive neurons.

All of the known learning rules lead to a finite capacity of the memory i.e., the performance of the model decreases dramatically if the number of stored patterns exceeds some critical value. For example the Hebb's rule allows for storing $p \approx 0.14N$ patterns in the limit of large number of neurons $N \in \infty$. The Hebb's learning rule, implemented in Hopfield networks, was used in the models presented in this paper (for a detailed description of the Hopfield model see Hopfield 1982).

The matrix of connections in a Hopfield network does not have to be symmetric i.e., J_{ij} does not have to be equal to J_{ji} . It has been, however, stressed by Hopfield that the model simplifies enormously if the symmetric connections are used. In such a case the "energy" function may be defined and its dynamics during a recognition process becomes equivalent to the process of minimalization of energy.

The concept of energy minimalization allows for a formulation of a simple intuitive description of the dynamics. Namely, we can imagine that the energy landscape in the space of network configurations is composed of many hills and valleys, as presented in Fig. 2.

The dynamics of the network then corresponds to the strongly damped motion from the initial configuration to the "closest" minimum in the direction of the steepest descent.

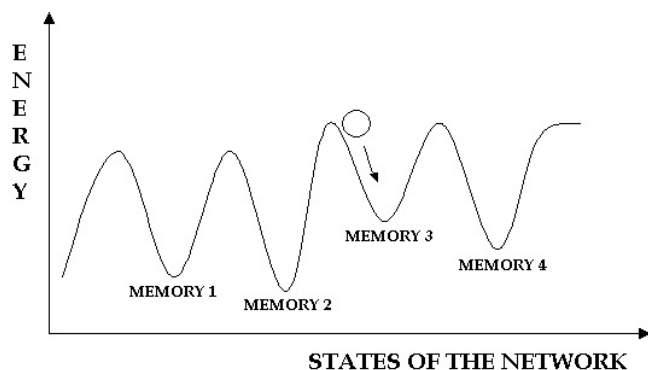


Fig. 2. Energy landscape in an attractor neural network.

As we mentioned before, attractor networks, including the Hopfield system, are hugely simplified model of biological neural networks. Nevertheless, when interpreted properly, they may give much insight into

how the nervous system works. To understand their potential, one has to take into account that attractor networks are qualitative models. They do not seek to precisely mimic real neuronal groups but are rather aimed at simulating simplified and generic brain mechanisms – not dependent on properties of a particular group of cells. Furthermore, nodes in the case of Hopfield model do not represent activity of a single cell. Rather, the state of a node in this type of network is interpreted as an averaged activity of a group of neurons, or their active vs. non-active state. Thus, although attractor neural networks cannot be regarded as realistic models of neural systems, they may provide qualitative understanding on some processes occurring in real neural nets.

RESULTS

In the following sections we will propose two measures that may function as a dynamical novelty check in artificial neural networks: neuronal flip frequency and measurement of the local field. Next, we will show how they might be coupled with control mechanisms, such as adjusting noise level and plasticity of connections, to produce a self-adaptive, dynamical system. We will use two examples to show the workings of those mechanisms. In the first, the self-adjustment of noise levels in the system will allow the network to vary its dynamical state between the retrieval regime and quickly evolving unstable state that could be interpreted as "don't know" answer. In the second model the self-adjustment of the speed of synaptic modifications will drive the network into retrieval or learning phase, depending on stimulus novelty.

Flip frequency

In the standard Monte Carlo simulation (Binder 1979, 1984) of a neural network of the Hopfield type, randomly chosen neurons are checked for their alignment with the signal arriving from other neurons. If neural state agrees with the sign of this signal the neuron does not change state. Conversely if the field has an opposite sign to that describing the neural state, the neuron modifies it. Thus the average flip rate of neurons is a good measure of the agreement of initial state of the network with one of its stable points (stored memories). If the state of the network does not overlap with one of the stored memories (i.e., is novel) the flip

rate in the network is high (flip rate probability is 0.5). On the other hand, if the initial state of the network overlaps with one of the stable points the flip rate is very low or zero.

We have tested flip frequency as novelty indicator in computer simulations (compare also Lewenstein and Nowak 1989a, b). We used a standard Hopfield network with a slightly modified Hebb's rule in which different patterns contribute to connection weights with different strength:

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^p w^{\mu} \xi_i^{\mu} \xi_j^{\mu} \quad (3)$$

where w^{μ} is the weight of the μ -th pattern. This allows for definition of "weakly" and "strongly" memorized patterns.

The net was presented with an input state that was one of the memorized states t^{μ} , partially distorted. The distortion was done by choosing randomly a given fraction of N neurons and changing their states. Such distorted initial patterns S_i^{μ} are very well characterized by their overlap q with their ancestor states

$$q = \frac{1}{N} \sum_j t_j^{\mu} s_j^{\mu} \quad (4)$$

The overlap $q=1$ means that the overlap is complete – none of the elements was changed and the actual pattern is identical to its ancestor. On the other hand, the overlap 0 means that the similarity of an actual pattern and its ancestor is at the level of chance i.e., 50 percent.

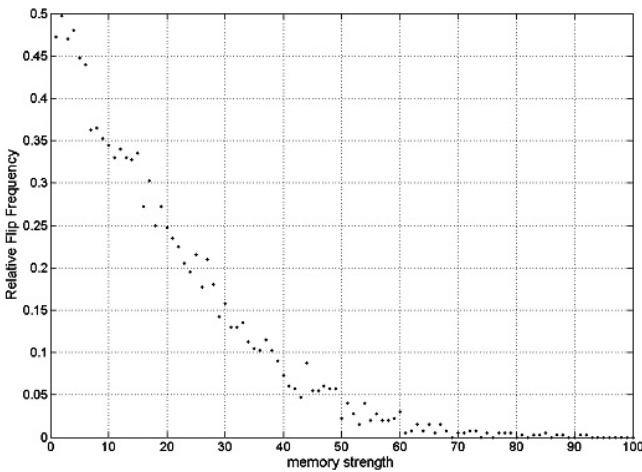


Fig. 3. Flip frequency as a function of memory strength.

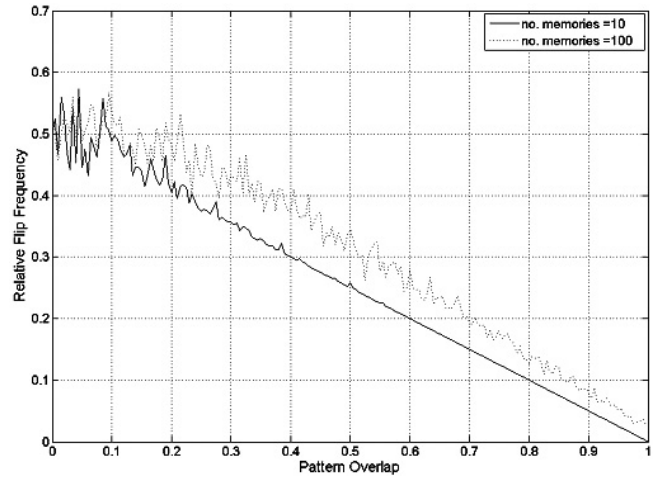


Fig. 4. Flip frequency as a function of pattern overlap.

Figure 3 summarizes the effects of memory strength on flip frequency, while Fig. 4 shows that the relative frequency of flips is a monotonic function of the initial overlap of the input pattern with its ancestor. This measure works even for heavily overloaded network (400 neurons gives maximum network capacity about 56 stored patterns). In the latter simulations stored patterns weights were equal and set to 1 – there was no division into weak and strong memories.

Local field

The above proposed novelty check can hardly be generalized to the case of multilayered neural networks. This can be done, however, if we use an alternative novelty check based on a criterion of local field strength. Let us first consider the case of the standard Hopfield model, and let us look at the so called local field, which is in fact the signal that comes to the neuron i 'th from all the other neurons j i.e.,

$$h_i = \sum_{j \neq i} J_{ij} s_j \quad (5)$$

Let us also assume, that the synaptic connection matrix J_{ij} has a standard Hebbian form (Eq. 2). We may consider two separate cases:

- The configuration S_j is random and is not correlated to any of the memorized patterns t^m . In such a case the local field averaged over stochastic properties of t 's is zero.

- b. The configuration S_i has a non vanishing correlation to one of the stored patterns, t^μ . In this case the mean local field is given by

$$\langle h_i \rangle = q \quad (6)$$

while its standard deviation by

$$\sqrt{\langle h_i - \langle h_i \rangle \rangle^2} = \sqrt{\frac{p}{N}} \quad (7)$$

The above two statements allow for formulation of a novelty check based on the value of the local field.

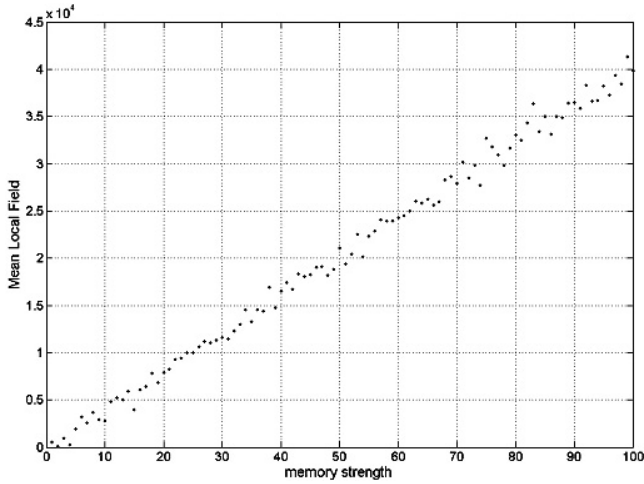


Fig. 5. The same as Fig. 3, but for mean local field instead of flip frequency.

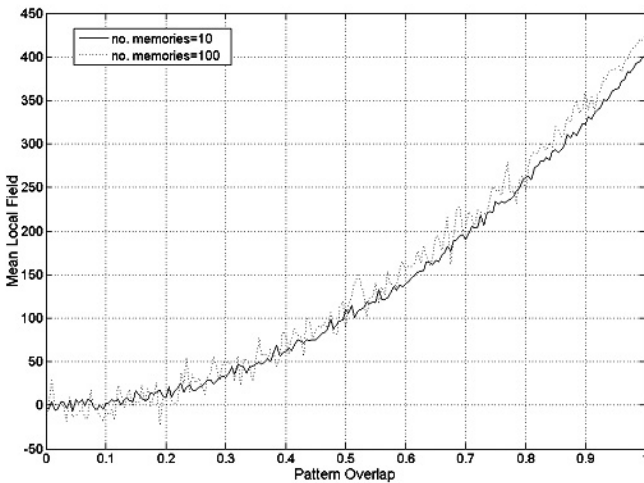


Fig. 6. Strength of mean local field as a function of pattern overlap. It can be observed that this measure works for overloaded networks as well as relative flip frequency.

Namely, if the values of the local field are large it is an indication that the current pattern overlaps with one of the stored memories and thus it is known to the system. Conversely, if the local fields are small, the current pattern does not overlap with the stored configurations and thus is novel.

Again we have studied this measure as a novelty check using computer simulations. The results were similar to these presented above for flip frequency and are summarized in Figs 5 and 6.

Properties of the novelty measures

Novelty detection based on the above measurements shows two very interesting and useful features. First, novelty check may be many times faster than pattern recognition. Judgment is made on the very beginning of the cognitive process. Those findings correspond well with empirical data known as mere exposure effect. Seamon and coauthors (1984) showed that human affective evaluation of the presented pattern is based on the novelty of the stimulus and occurs much faster than pattern recognition. Novel stimulus evokes negative and familiar – positive emotional response. In fact, subjects were able to distinguish between known and unknown stimulus even when they did not recognize the patterns, as they were presented subliminally (Monahan et al. 2000). The ability of flip frequency model to simulate the effects of Seamon and coauthors was shown by Drogosz and Nowak (1996).

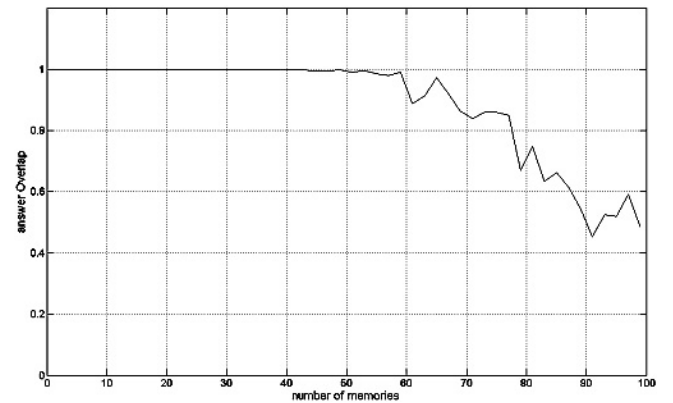


Fig. 7. Effect of overtraining. If number of patterns used during learning phase exceeds 14% of the number of neurons, the network starts to make mistakes during recognition. Plot of the mean answer and training pattern overlap over 20 trials. $N=400$.

Second, novelty detection is much more sensitive and selective than recognition. Hopfield neural networks have estimated memory capacity (number of imprinted patterns) about 14% of the number of neurons.

If one over trains such a network, the memories will merge and will not represent any of the patterns from the learning set. In contrast, novelty detection remains accurate even for a strongly overloaded network.

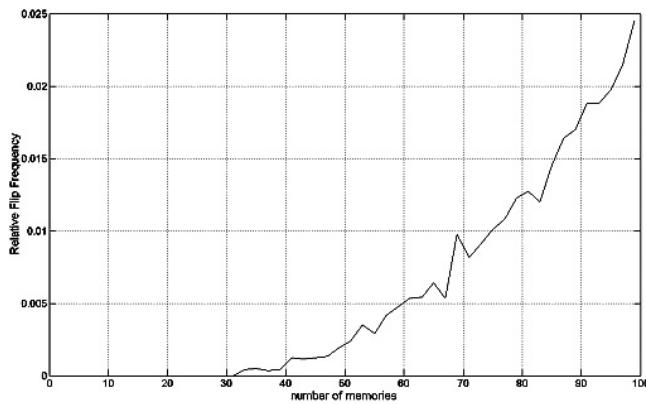


Fig. 8. Flip frequency as a function of number of memories. $N=400$.

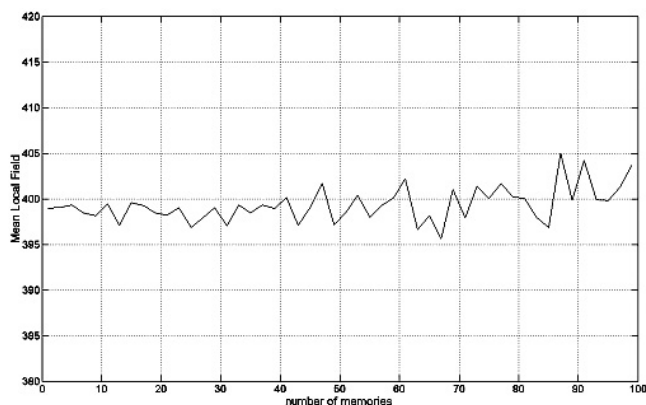


Fig. 9. Mean local field as a function of number of memories. $N=400$.

As we can see from Figs 8 and 9, novelty measures remain accurate for overloaded networks – there is little difference between networks with encoded 10 and 100 memories (0.025 and 0.25 N). Presented stimulus was one of the learned patterns. Flip frequency remains low (below 5%) and local field high for number of learned patterns even 2 times higher than network capacity.

Mechanism of feedback regulation

The essence of novelty based regulation is that once novelty is assessed, the result of this assessment may be used to change the dynamics of the recognition process. The general mechanism is based on dynamical coupling of novelty criteria to parameters that control the recognition. Below we present two examples of such feedback regulation. In the first one, model A, the varied control parameter is the noise level of the network; in the second one – the speed of modifications of neural connections.

Model "A"

One of the main advantages of neural networks is the ability of this class of models to recognize highly distorted patterns. In fact, neural network models are "too good" in recognizing distorted patterns. The problem is that eventually any pattern will be recognized as one of the stored ones. In their standard form networks are not capable of producing "do not know" answers.

To design a network capable of rejecting unknown patterns, the noise level may be used as a control parameter of recognition. The noise level is dynamically linked to a measure of network coherence in such a way that low coherence (i.e., high flip frequency) increases the noise level in the system. Low flip frequency decreases the noise level.

The energy landscape metaphor, discussed previously, allows for intuitive understanding of the role of noise in the dynamics of the network. Namely, the introduction of random terms into the dynamics correspond roughly to the process of "shaking" of the energy landscape. If the "shaking" is weak the network retrieves stored patterns like an ordinary Hopfield model. On the other hand, if the noise is large enough the system may leave the local minimum of energy and "jump" to other one or become random. As we see the probability of such an event depends crucially on the fact whether the noise can provide sufficient energy to overcome the energy barriers that surround any of the local energy minima.

The recognition dynamics of the simulated system (presented in Lewenstein and Nowak 1989a, b) is described by Hopfield equation with additional noise term. This noise term describes probability of i -th neuron changing the state even if its state is in agreement

with the sign of the local field and is given by:

$$P_i = 1 - \exp(-2\beta h_i S_i) \quad (8)$$

where β is the noise level in the system. Thus with the increased noise levels, the neuron has higher probability of flipping even though its state is in agreement with its local field.

The learning patterns were divided into 'weak' and 'strong' ones like in Eq. 3, with $w^1 > 1$ for strong memories and $w^2 = 1$ for weak memories. Earlier simulations indicated that the variation of weights reduces the number of distinctive patterns that such a system may remember. The stronger memories tend to destroy the weaker ones.

In the recognition phase, the effect of flip frequency on noise level was implemented as follows:

- Each neuron that changes state raises the noise level of the system by some quantity
- Each neuron that does not change state, when checked, decreases the noise level.

The recognition process starts at relatively low levels of noise.

BEHAVIOUR OF THE SYSTEM

In our simulations we have observed three types of dynamics of the recognition process.

1. For initial patterns that were highly similar to remembered ones (overlap, as in Eq. 4, close to 1), the initial relative number of neurons changing their states is small. The pattern is categorized as "known" in the prerecognition process. The noise level decreases to zero and the stimulus is perfectly recognized.

2. Processing of highly distorted patterns (overlap close to 0) is characterized by a high frequency of neuron flips. The noise level rapidly increases. At some level of noise the system gets into a positive feedback loop, where the noise increases the frequency of elements changing states, what in turn increases the noise level. The dynamics of the recognition process becomes chaotic and none of the stored patterns can be retrieved in the recognition process. The resulting unstable state can be regarded as a "don't know" answer.

3. For some intermediate levels of novelty (distortion) the noise slowly increases to a level where the weakly memorized patterns can not be recognized. In the process of approaching the weakly remembered

pattern the noise level may also slowly drop, but this change is too slow to allow for recognition. The recognition is characterized by wandering in the energy landscape, approaching the weakly memorized states, but not being able to recognize them. The noise level raises and drops indicating closeness of stored patterns. At some time stronger memory is encountered. It can be recognized at this level of the noise. The frequency of neurons changing states decreases and the system reduces its noise level, ending up in one of the strongly remembered patterns. In effect, it recognizes distorted versions of one of the poorly remembered patterns as one of the strongly remembered ones.

4. When the presented patterns belonged to the class of strongly memorized ones, only two kinds of dynamics were observed. Even at high level of distortion the ancestor patterns were correctly recognized. "Don't know" answers appeared only when the presented stimuli were almost completely random and unfamiliar (overlap < 0.2). For weakly memorized patterns the recognition is also perfect when distortion levels are low. For medium levels of familiarity (overlap < 0.4) the third type of dynamics dominates. The weak, distorted patterns are erroneously recognized as strongly remembered patterns. At still higher levels of distortion the "don't know" answer dominates.

The results of the model are macroscopically comparable to empirical results of Skarda and Freeman (1987), who have demonstrated that brain dynamics in response to unknown stimuli may in fact become chaotic.

Smartnet

One of the main problem of designing neural networks, and artificial intelligence systems in general, is the problem of separation of the learning and recognition phases. During recognition novel input pattern is usually transformed to match some memory, therefore the information which should be memorized is lost. On the other hand, if the pattern is already in the memory, it is undesirable to learn it because increasing the strength of this memory (deepening the attractor) will lead to distortion of other memories.

Żochowski and coauthors (1995) have shown that a system able to pre-recognize novelty on the basis of flip frequency measure is able to switch between two distinct dynamics:

1. recognition – when the input is known, the network tries to match it with the stored configurations (neural dynamics plays a dominant role).

2. learning – when the stimulus is novel, the network tries to accommodate it by changing the connection strengths (proceeding with neural dynamics at the same time).

The dynamics of the synaptic connections J_{ij} can be defined as follows.

$$J_{ij}^{t+1} = J_{ij}^t + F(\gamma^t)(S_i^y S_j^t) \quad (9)$$

Here $F(\gamma^t)$ is a learning force and is an appropriately designed positive function. It increases rapidly when γ (relative flip frequency) goes away from zero and γ tends to 0.5, and approaches zero when $\gamma \cong 0$.

It can be seen that the learning process is significant only when the pattern is unknown to the network. Conversely, when the pattern is recognized as known and γ tends to 0, the dynamics of the synaptic matrix is reduced and the connection strengths change very slowly. The factor $S_i^y S_j^t$ in the second term of the above equation is a standard Hebbian learning term.

Learning/recognition phases interchange: at the beginning of the first study session synaptic interconnections are random and weak. Then we introduce the patterns that are to be memorized. When a learning pattern is presented for the first time, the network of course does not recognize it. There flip frequency is high and the "learning force" $F(\gamma)$ quickly rises – the system is memorizing new information.

After introduction of a known stimulus, the system will quickly reduce the plasticity of the synaptic interconnections, and will prevent undesired enlargement of the basin of attraction of the stimulus presented – it will not overwrite and destroy other stored memories. Simulation results have shown that although Smartnet dynamically switches between learning and recognition, the network capacity is equal to standard Hopfield model.

Summary of results

We have introduced novelty measures in neural networks that enable the net to dynamically control its information processing. Both measures can be applied early in the recognition process and therefore allow for fast and reliable novelty check. Moreover, both are robust and work even for very overloaded networks. It

is reasonable then to state that distinguishing between novel/familiar stimuli may indeed be faster than recognition itself, as experimental studies suggest (Skarda and Freeman 1987).

Our simulation results show that these measures can be successfully implemented in self-adapting systems to produce various phenomena present in cognitive system – such as a "don't know" response or learning and recognizing phases. We think therefore that this novelty check may have a much broader domain of applications. The same novelty check can, in our opinion, be also applied for multilayered perceptrons and networks (see also Rumelhart and McClelland 1986).

In this paper we presented two specific models of feedback regulation. It is worth stressing, though, that the proposed novelty measures are dynamical and therefore many other feedback mechanisms might be implemented by using different network architectures and different parameters that control the dynamics of recognition (e.g., Żochowski et al. 1993, 1994).

In model "A", novelty affects noise level in order to produce a "don't know" response. Adjusting the noise level can lead to major qualitative changes of the dynamics. In particular some or all of the stationary states of the system may become unstable due to the influence of noise. In such a case even small departures from learned states may end up in an escape to other regions of the space of network configurations or in case of larger departures produce random dynamics which may be defined as "don't know" response.

"Smartnet", the second model reviewed here, relies on pre-recognition of familiarity of presented patterns to switch between recognition and learning phases. Because of the "overwriting effect", the storage capacity of a standard attractor network with dynamic connections is significantly reduced. For example, the model discussed by Dong and Hopfield (1992) reveals such limitation. Here, thanks to the novelty check, the system is able to dynamically change the speed of connection changes and therefore its capacity remains unaffected. Broadly speaking, novelty affects the plasticity of neural connections.

CONCLUSIONS

In this paper we presented specific models of feedback regulation introduced in Hopfield type networks. The general principle of feedback regulation may, however, be applied to a much broader class of networks.

Constructing a particular model requires specification of novelty criteria and control parameter that is applicable to the specific network architecture. While the neural flip frequency measure is applicable only to attractor networks with binary neurons, the local field criterion introduced in this paper may be applied to a larger class of models – both to models with neurons with continuous activation function and multi-layered networks. Both the noise level and synaptic plasticity may also function as control parameters for large class of artificial neural networks. A question remains about applicability of the proposed model to biological neural networks. Extension of the proposed models of self-regulation in this direction would depend on the ability to specify how processing of novelty is reflected in the dynamics of natural neural systems and which properties of such systems may act as dynamical control parameters.

REFERENCES

- Amit DJ (1989) *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge: Cambridge University Press.
- Binder K (ed.) (1979) *Monte Carlo Methods in Statistical Physics*, Vol 1. Heidelberg: Springer-Verlag.
- Binder K (ed.) (1984) *Monte Carlo Methods in Statistical Physics*, Vol 2. Heidelberg: Springer-Verlag.
- Dong D, Hopfield JJ (1992) Dynamic properties of neural networks with adapting synapses. *Network: Computation in Neural Systems* 3: 267–283.
- Drogosz M, Nowak A (1996) Simulation theory of the mere exposure effect: The EXAC neural network model (in Polish). *Przegląd Psychologiczny* 38: 65–84.
- Friedman D, Cycowicz YM, Gaeta H (2001) The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neurosci Biobehav Rev* 25: 355–373.
- Gilbert CD, Sigman M, Crist RE (2001) The neural basis of perceptual learning. *Neuron* 31: 681–697.
- Grill-Spector K, Malach R (2001) fMR-adaptation: A tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst.)* 107: 293–321.
- Habib R (2001) On the relation between conceptual priming, neural priming, and novelty assessment. *Scand J Psychol* 42: 187–195.
- Henson RN, Rugg MD (2003) Neural response suppression, haemodynamic repetition effects, and behavioural priming. *Neuropsychologia* 41: 263–270.
- Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational capabilities. *Proc Natl Acad Sci U S A* 79: 2554–2558.
- Huk AC, Ress D, Heeger DJ (2001) Neuronal basis of the motion aftereffect reconsidered. *Neuron* 32: 161–172.
- Knight RT (1984) Decreased response to novel stimuli after prefrontal lesions in man. *Electroencephalogr Clin Neurophysiol* 59: 9–20.
- Lewenstein M, Nowak A (1989a) Fully connected neural networks with self-control of noise levels. *Phys Rev Lett* 62: 225–228.
- Lewenstein M, Nowak A (1989b) Recognition with self-control in neural networks. *Phys Rev A* 40: 4652–64.
- McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in neural nets. *Bulletin of Mathematical Biophysics* 5: 115–124.
- Miller EK, Gochin PM, Gross CG (1991) Habituation-like decrease in the response of neurons in inferior temporal cortex of the macaque. *Vis Neurosci* 7: 357–362.
- Monahan JL, Murphy ST, Zajonc RB (2000) Subliminal mere exposure: Specific, general, and diffuse effects. *Psychol Sci* 11: 462–466.
- Ranganath C, Rainer G (2003) Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci* 4: 193–202.
- Rumelhart DE, McClelland JL (eds.) (1986) *Parallel Distributed Processes: Explorations in Microstructure of Cognition*, Vol 1. Cambridge, MA: MIT Press.
- Sanchez-Vives MV, Nowak LG, McCormick DA (2000) Cellular mechanisms of long-lasting adaptation in visual cortical neurons in vitro. *J Neurosci* 20: 4286–4299.
- Seamon JG, Marsh RL, Brody N (1984) Critical importance of exposure duration for affective discrimination of stimuli that are not recognized. *J Exp Psychol Learn Mem Cogn* 10: 465–469.
- Schoups A, Vogels R, Qian N, Orban G (2001) Practising orientation identification improves orientation coding in V1 neurons. *Nature* 412: 549–553.
- Skarda CA, Freeman WJ (1987) How brains make chaos in order to make sense of the world. *Behav Brain Sci* 10: 161–195.
- Sobotka S, Ringo JL (1994) Stimulus specific adaptation in excited but not inhibited cells in inferotemporal cortex of macaque. *Brain Res* 646: 95–99.
- Sokolov EN (1963) Higher nervous functions: The orienting reflex. *Annu Rev Physiol* 25: 545–580.
- Sokolov EN (1969) The modeling properties of the nervous system. In: *A Handbook of Contemporary Soviet Psychology* (Cole M, Maltzman I, eds). New York: Basic Books, p. 670–704.

- Squires NK, Squires KC, Hillyard SA (1975) Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol* 38: 387–401.
- Sutton S, Braren M, Zubin J, John ER (1965) Evoked potential correlates of stimulus uncertainty. *Science* 150: 1187–1188.
- Winkelman P, Nowak A (2005) Dynamics of cognition-emotion interface: Coherence breeds familiarity and liking, and does it fast. *Behav Brain Sci* 28: 222–223.
- Winkelman P, Schwarz N, Nowak A (2002) Affect and processing dynamics. In: *Emotional Cognition* (Moore S, Oaksford M, eds). John Benjamins Publishing Company, Amsterdam/Philadelphia, p. 111–136
- Vallacher RR, Nowak A (1999) The dynamics of self-regulation. In: *Advances in Social Cognition*, Vol 12 (Wire Jr RS, ed.). Mahwah, NJ, Lawrence Erlbaum, p. 241–259.
- Zajonc RB (1980) Feeling and thinking: Preferences need no inferences. *Am Psychol* 39: 151–175.
- Żochowski M, Lewenstein M, Nowak A (1993) Memory that tentatively forgets. *Journal of Physics A*, 26: 2453–2460.
- Żochowski M, Lewenstein M, Nowak A (1994) Local noise in neural networks with self-control. *Int J Neural Syst* 5: 287–298.
- Żochowski M, Lewenstein M, Nowak A (1995) SMART-NET: A neural network with self-controlled learning. *Network* 6: 93–101.

Received 1 August 2005, accepted 8 November 2005

