SELF-REFRESHING SOM AS A SEMANTIC MEMORY MODEL

Matti Pöllä, Tiina Lindh-Knuutila and Timo Honkela

Neural Networks Research Centre, Helsinki University of Technology,
P.O.Box 5400, FI-02015 TKK, FINLAND, {matti.polla,tiina.lindh-knuutila,timo.honkela}@tkk.fi

ABSTRACT

Natural and artificial cognitive systems suffer from forgetting information. However, in natural systems forgetting is typically gradual whereas in artificial systems forgetting is often catastrophic. Catastrophic forgetting is also a problem for the Self-Organizing Map (SOM) when used as a semantic memory model in a continuous learning task in a nonstationary environment. Methods based on rehearsal and pseudorehearsal have been successfully applied in feedforward networks to avoid catastrophic interference. A novel method based on pseudorehearsal for avoiding catastrophic forgetting in the SOM is presented. Simulations comparing the performance of a self-refreshing SOM compared to a standard SOM are presented in the task of learning three separate sets of data adjacently with results showing that the use of pseudorehearsal can effectively decrease catastrophic forgetting.

1. INTRODUCTION

Computational cognitive systems are typically constructed by imitating natural cognitive systems. One of the research objectives relating to computational cognitive systems is to mimic the way humans store information. This memory can be roughly divided into an episodic memory – remembering the contents of a past event – and a semantic memory – storing general world knowledge. In particular, research on language learning requires a model for the conceptual system of an agent.

Memory is closely related to another characteristic of all natural cognitive systems, namely forgetting. Artificial neural networks also have the tendency of forgetting previously learned data, which can be seen as a manifestation of the general 'stability-plasticity' dilemma (Grossberg, 1982). However, the mechanism of forgetting differs severely in natural and artificial system: while humans forget gradually, connectionist networks forget catastrophically as a result of new input information overwriting old information (French, 1999; French, 1994).

The SOM is typically used as a static mapping of a predetermined training data set. In this case no forgetting occurs as the order of the input patterns is irrelevant to the resulting projection. In a continuous learning task – such as the one described in section 2.2 – the problem of catastrophic forgetting has to be taken into account in order to have a mapping that is dynamic enough to learn new information but at the same time stabile enough to remember previous inputs just as the cognitive system of a human is able to preserve representations of concepts that have no recent 'activations'.

Several methods for avoiding catastrophic forgetting in feedforward networks have been developed with good results (Ans et al., 2004). Typically these schemes avoid catastrophic interference by presenting the network reminders or 'rehearsal' data of previous inputs. Often the actual previous inputs are not available, which has led to the idea of pseudorehearsal (Robins, 1995) based on presenting the network with internally generated representations of previous inputs. In section 3 the idea behind pseudorehearsal is extended to the domain of unsupervised learning in a SOM network.

It has been suggested that pseudorehearsal could, in fact, be the mechanism that the human brain uses to consolidate old and new information. Robins has viewed the pseudorehearsal method as an equivalent to the sleep consolidation hypothesis, which essentially explains the function of sleep as a means of integrating newly acquired information into existing long-term memory (Robins, 1996).

2. SELF-ORGANIZING MAP AS A SEMANTIC MEMORY MODEL

2.1. Self-Organizing Map

The Self-Organizing Map algorithm (SOM hereafter) is a neural network algorithm for creating a topologically correct nonlinear projection of high dimensional data into a neuron lattice of lower dimensionality $\mathbb{R}^n \rightarrow \mathbb{R}^m, m < n$ (Kohonen, 2001). The SOM is typically used to cluster and visualize multidimensional data sets that would be hard to analyze in their original form. Because of the unsupervised nature of the learning algorithm the SOM can be used to model the semantic memory of an autonomous agent (cf., e.g., Honkela and Winter, 2003). Basic framework and motivation for using the SOM in learning and representing conceptual information has been presented, e.g., by Ritter and Kohonen, 1989; Miikkulainen, 1997; MacWhinney, 1998; Gärdenfors, 2000.

The SOM projection is typically a two-dimensional neuron lattice with each node corresponding to a weight vector of the same dimension as the input space. In the resulting projection the SOM codebook vectors are adjusted such that two input vectors that reside close to each other in the input space have their corresponding best matching units on the SOM space near each other.

The SOM algorithm combines competitive learning and Hebbian learning by first selecting a best matching unit (BMU) neuron for each input pattern according to some distance metric and then altering the BMU neuron and its neighboring neurons closer to the input pattern according to some neighborhood function. For a complete review of the SOM algorithm see (Kohonen, 1982).

2.2. Application of the SOM as a Semantic Memory

The unsupervized and adaptive nature of the SOM makes it suitable as a model for semantic memory. For example, the SOM has been used as a basis for a model for concept acquisition (Schyns, 1991), where a self-organizing map was given a task of learning competing categories with a prototype structure.

The SOM has also been used as a model for implementing the individual semantic memory systems of autonomous agents (Honkela and Winter, 2003). In these experiments the SOM was used to associate visual perceptions to linguistic objects represented as continuous-valued vectors rather than discrete symbols. The semantic memory became gradually organized in a meaningful way and thus enabled a shared conceptual system and a language for agent-to-agent communication. Other examples on the use of the SOM include the formation of conceptual spaces (Gärdenfors, 2000).

In an online learning situation such as the one described above the effect of new input disrupting previously learned information should be considered in order to implement a semantic memory with a sufficient balance between the capability to acquire new information and remembering old information. Especially when the SOM is trained continuously – as opposed to batch training – the choice of the time-dependent parameters of the SOM algorithm are critical.

In the following section a self-refreshing mechanism based on pseudorehearsal is presented to alleviate the effect of catastrophic forgetting when using the SOM as a model for semantic memory.

3. SELF-REFRESHING SOM

The described phenomenon of new inputs overwriting learned data is a consequence of the lack of representation of previous input data in the training data set of the SOM. This kind of forgetting could be avoided by training the map each time with a data set that contains not only the new patterns but also every previously input pattern. The discouraging outcome of this approach is the requirement of an infinite short term memory.

However, before each new training round the codebook vectors of the SOM can be used as a representation of previously learned data. This information can be used to present a set of the codebook vectors of the map as reminders or 'rehearsal' data of the old input vectors together with new data patterns (Figure 1).

Mixing new data vectors with codebook vector representation of *all* previously learned data would still require a short term memory of infinite length. This can be avoided by selecting a set of M rehearsal patterns which can be selected randomly or by selecting M patterns that have the smallest quantization error.

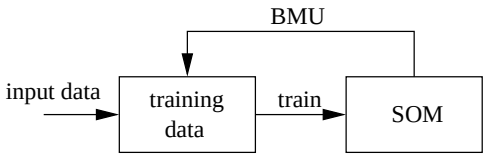

Figure 1. The self-refreshing pseudorehearsal mechanism uses the SOM codebook vectors recursively as a representation of previous inputs.

The self-refreshing pseudorehearsal scheme for the SOM is defined by the following five steps.

1. Create a SOM projection from input space $\mathbb{R}^n$ to the SOM space $\mathbb{R}^m$.

2. Use N input patterns as the initial data set

$$\mathbf{X} = \mathbf{X}_0 = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\} \in \mathbb{R}^n \qquad (1)$$

3. Train the SOM with the current data set $\mathbf{X}$.

4. Map all patterns of the current data set $\mathbf{X}$ to the SOM lattice and replace the current data set with a set of M corresponding codebook vectors[1] combined with N new input patterns.

$$\mathbf{X} = \{\mathbf{m}_1, ..., \mathbf{m}_M\} \cup \{\mathbf{x}_{1,\text{in}}, ..., \mathbf{x}_{N,\text{in}}\} \qquad (2)$$

5. Go to step 3.

The balancing between a fast adoption on new input data and the capability to store previous infomation can be modified by selecting proper values M and N as parameters of the self-refreshing mechanism with $M >> N$ corresponding to a slow adoption rate for new inputs.

In the extreme ends $M = 0$ would be equivalent to a standard SOM with no pseudorehearsal and $N = 0$ to a completely recurrent system with no capability of adopting new information.

4. SIMULATIONS

The proposed method was tested with three sets of generated random data to compare the performance of a self-refreshing SOM and a standard SOM. The data sets $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ each contained 30 samples of six-dimensional normally distributed random vectors

$$\begin{aligned} \mathbf{A} &= \{\mathbf{x_1}, ..., \mathbf{x_{30}}\} \sim N_6(\mu_A, \mathbf{I}) \\ \mathbf{B} &= \{\mathbf{x_1}, ..., \mathbf{x_{30}}\} \sim N_6(\mu_B, \mathbf{I}) \\ \mathbf{C} &= \{\mathbf{x_1}, ..., \mathbf{x_{30}}\} \sim N_6(\mu_C, \mathbf{I}) \end{aligned} \qquad (3)$$

[1]The selection can be done either randomly or by selecting M best matches according to the quantization error (4).

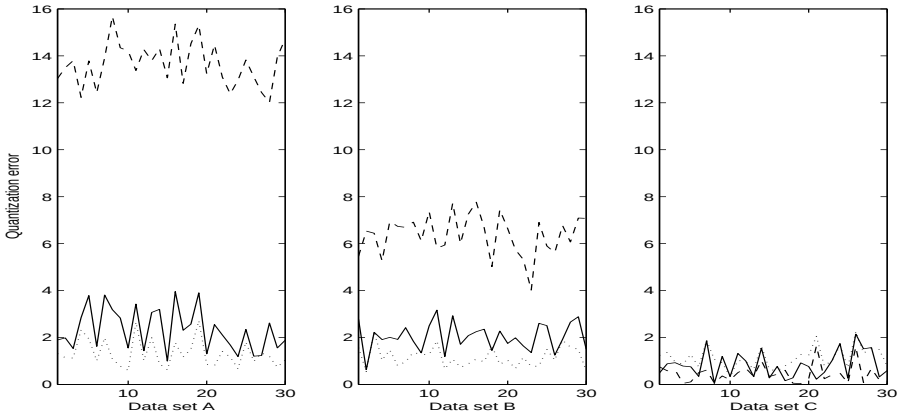

Figure 2. Quantization errors for a standard SOM (dashed line) and a SOM with self-refreshing (solid line) after presenting the network three separate sets of data adjacently. The dotted line represents the performance of a SOM with an infinite short term memory.

where $\mathbf{I}$ is a 6×6 identity matrix, $\mu_A = -3$, $\mu_B = 0$ and $\mu_C = 3$.

A normal SOM and a self-refreshing version were both trained with data sets $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ after which the learning result was examined. In these simulations, 'remembering' and 'forgetting' was measured by the resulting quantization error i.e. Euclidian distance between an input pattern $\mathbf{x} \in \mathbf{X}$ and its closest equivalent $\mathbf{m}_c$ in the SOM codebook vector set $\mathbf{M}$.

$$||\mathbf{x} - \mathbf{m}_c|| \tag{4}$$

As expected, both types of networks resulted in a good representation of the latest data set ($\mathbf{C}$) which can be seen as a small quantization error in Figure 2c. However, the results for remembering the previous data sets ($\mathbf{A}$ and $\mathbf{B}$) in Figure 2a and 2b have a drastic difference. The self-refreshing SOM has only a slightly larger quantization error for the first data set whereas a standard SOM has 'forgotten' it completely.

The same result can be seen in Figures 3a and 3b where the codebook vectors are shown together with the input data patterns as a two-dimensional projection. The codebook vectors of a standard SOM are focused solely on the final data set (centered at 3,3) and no representation for the previous data exists. The self-refreshing SOM has its codebook vectors spread across all three data sets and thus no significant forgetting has occurred.

Using the codebook vectors as representations of previous inputs have the disadvantage of possessing small amounts of noise due to quantization errors. Recurrent feeding of the noisy codebook vectors to the SOM will eventually result in a cumulating effect of the error which is the price for not being able to use the original (undistorted) inputs. This effect can be seen as slow drifting of the codebook vectors as in Figure 4. This effect can be decreased by selecting a sufficiently large SOM lattice combined with strict requirements for the total quantization error

$$\sum_{\mathbf{x}_i \in \mathbf{X}} ||\mathbf{x_i} - \mathbf{m}_c|| \tag{5}$$

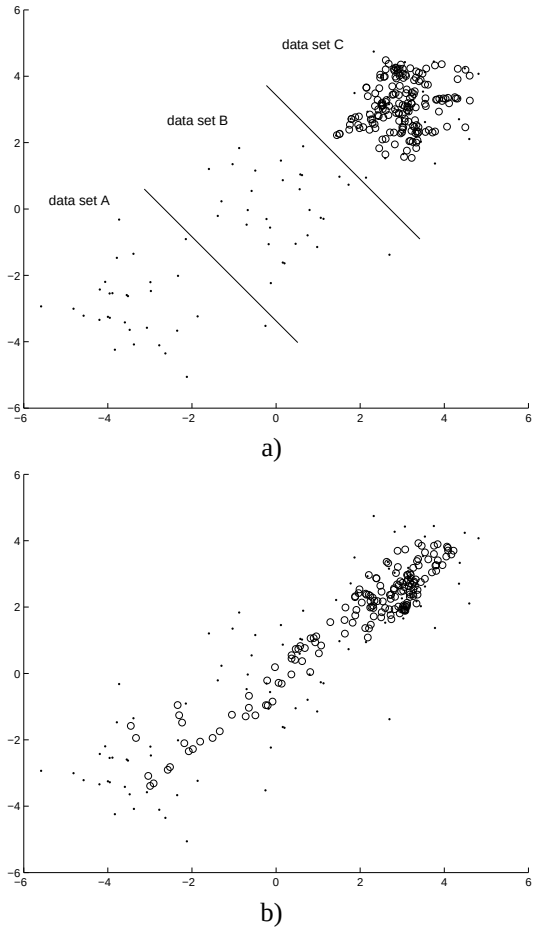

Figure 3. Distribution of the a) training data (dots) and the codebook vectors (circles) of a a) regular and a b) self-refreshing SOM.

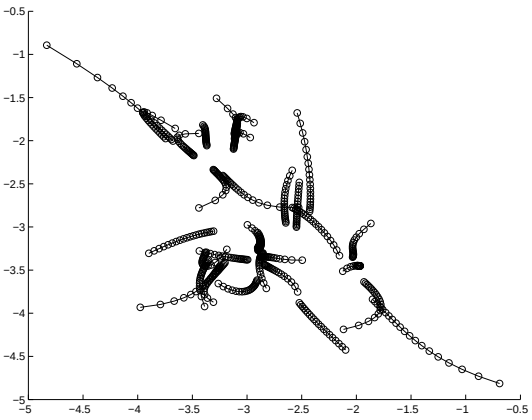

Figure 4. Two-dimensional projection of the codebook vectors of a fully recursive pseudorehearsal system ($N = 0$) during 20 training rounds. The cumulating error shown as divergence of the vectors can be avoided by selecting a sufficiently large SOM.

5. CONCLUSIONS AND DISCUSSION

In this article the Self-Organizing map was discussed as a model of semantic memory. It potentially suffers from the problem of forgetting previously learned information in an online learning situation. Applications of the use of the SOM as a semantic memory were introduced including language learning in a multi-agent environment.

Catastrophic forgetting was discussed as a problem of artificial neural networks. Methods based on rehearsal and pseudorehearsal were discussed as a solution to this problem in feedforward networks.

A modification to the SOM algorithm based on recursive use of previously learned data was presented to avoid catastrophic forgetting.

Simulations of learning three separate data sets each generated from different distributions were presented to study the performance of a self-refreshing SOM compared to a standard SOM. The simulation results show that a self-refreshing SOM was able to learn new information without significant forgetting of older information and thus avoiding catastrophic forgetting.

In the scope of this article, the self-refreshing rehearsal method was introduced and simulated with discrete batch-like input vector sets. Future research efforts include extending the self-refreshing scheme into a more general case of continuous learning where new data is learned one vector at a time instead of fixed-length data sets.

Although catastrophic interference is generally regarded as a problem of artificial neural networks, in some cases or contexts it could be claimed that it is useful to forget semantic memory contents that are based on earlier experiences. Our model supports this by making the forgetting gradual (as in natural systems). Additionally, in our model the rate of forgetting can be adjusted to construct a network with a sufficient sensitivity for new input.

References

Ans, B., Rousset, S., French, R. M., and Musca, S. (2004). Self-refreshing memory in artificial neural networks: learning temporal sequences without catastrophic forgetting. *Connection Science*, 16(2):71–99.

French, R. M. (1994). Catastrophic interference in connectionist networks: Can it be predicted, can it be prevented? In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6, pages 1176–1177. Morgan Kaufmann Publishers, Inc.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Science*, 3(4):128–135.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. A Bradford Book. MIT Press.

Grossberg, S. (1982). Prosessing of expected and unexpected events during conditioning and attention: A psychophysical theory. *Psychologigal Rewiev*, 89:529–572.

Honkela, T. and Winter, J. (2003). Simulating language learning in community of agents using self-organizing maps. Technical Report A71, Helsinki University of Technology. Publications in Computer and Information Science.

Kohonen, T. (1982). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69.

Kohonen, T. (2001). *Self-Organizing Maps*. Springer-Verlag, Heidelberg, 3rd edition.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, (49):199–227.

Miikkulainen, R. (1997). Self-organizing feature map model of the lexicon. *Brain and Language*, (59):334–366.

Ritter, H. and Kohonen, T. (1989). Self-organizing semantic maps. *Biological Cybernetics*, 6(1):241–254.

Robins, A. (1995). Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Science*, 7(2):123–147.

Robins, A. (1996). Consolidation in neural networks and in the sleeping brain. *Connection Science*, 8(2):259–275.

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science*, 15(4):461–508.