HELSINKI UNIVERSITY OF TECHNOLOGY 30.10.2002
T-61.181 Time In Self-Organizing Maps
Autumn 2002

T-61.181 Time In Self-Organizing Maps

Review of two temporal Self Organizing Map methods

Mikko Honkala
42753D

Mikko.Honkala@hut.fi

1 INTRODUCTION

In this paper, two research papers are reviewed. The first paper suggests a unsupervised network model, called Recurrent Self-Organizing Map (RSOM), which is a modified version of the original SOM. The second papers uses similar approach as RSOM to model the self-organization in the visual cortex of the brain.

1.1 Temporal sequence processing

Temporal sequence processing (TSP) is an active research area, with many applications in fieds, such as weather forecasting, and speech recognition. In order to create a model for such temporal process, data is gathered by measuring different variables sequentially in time. It must be noted that the measured data is often incomplete, erroneous and contains noise.

Traditional way of using neural networks in TSP to covert the temporal sequence of input vectors into a longer vector, which is a concatenation of some of the original vectors. This vector is then fed into a normal nonlinear neural network. This is called a *time-delay* network.

2 TEMPORAL SOM MODELS WITH INTERNAL SHORT-TERM MEMORY

Since SOM itself is not designed to be used with temporal data, a number of extensions have been suggested. One of the most popular way to extend the basic SOM is to use some kind of *internal short term memory* (STM). In internal short-term memory models, the map units somehow remember the previous activation or input values.

2.1 Temporal Kohonen Map

Temporal Kohonen Map (TKM) is an modification of the original SOM [TKM]. In TKM, the involvement of the earlier input vectors in each unit is represented by using a recursive difference equation, which defines the current unit activity as a function of the previous activations and the current input vector. The SOM outputs in the TKM are replaced with leaky integrator outputs, which, once activated, gradually lose their activity. Therefore the STM in TKM is in the *output* of each neuron. The modeling of the outputs in the TKM is close to the behavior of natural neurons, which retain an electrical potential on their membranes with decay. In the TKM, this decay is modeled with the difference equation [TKM]:

$$a_i(t) = \lambda a_i(t-1) - (1/2)\|\mathbf{x}(t) - \mathbf{w}_i(t)\|^2, \quad (1)$$

where 0 < λ < 1 is a time constant, $a_i(t)$ is the activation of the unit *i* at step *t*, $\mathbf{w}_i(t)$ is the reference of the weight vector in the unit *i* and $\boldsymbol{x}(n)$ is the input pattern.

The best-matching unit (BMU) is the unit with highest activity.

3 RECURRENT SOM (RSOM)

Some of the problems of the original TKM have a convenient solution in simply moving the leaky integrators from the unit outputs into the inputs. This leads to a model called Recurrent SOM (RSOM), described by Koskela, et.al. Moving the leaky integrators from outputs into the inputs yields: [RSOM]

$$\mathbf{y}_i(t) = (1-\lambda)\mathbf{y}_i(t-1) + \lambda(\mathbf{x}(t) - \mathbf{w}_i(t)), \tag{2}$$

for the temporally leaked difference vector at each map unit. $0 < \lambda <= 1$ is the leaking coefficient analogous to the TKM, $\mathbf{y}_i(t)$ is the leaked difference vector, and $\mathbf{w}_i(t)$ and ***x(t)*** have same meaning as in equation (1). Neurons in RSOM and TKM can be viewed as discrete time filters. Schematic pictures of TKM and RSOM units are shown in Figure 1 (a) and (b).

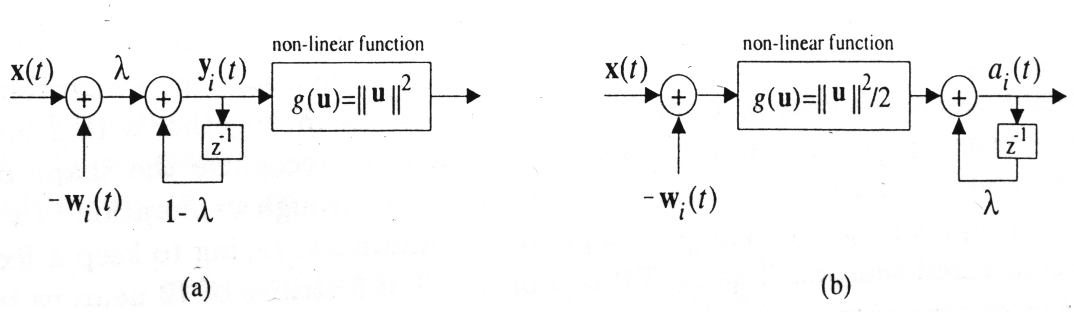

Figure 1. Schematic picture of (a) RSOM unit (b) TKM unit

This means that the feedback term is an vector, instead of an scalar, as in TKM. Because of this, the direction of the error can be used in updating the weights in the training stage. In TKM, only the amount of error could be used, and the update was always done to the direction of the last input vector.

BMU is searched by: [RSOM]

$$y_b = \min_i \{\|y_i(t)\|\}, \tag{3}$$

The learning rule of the normal SOM is modified so that the difference vector *(x(t)-w(t))* is replaced by y_i. This means that the unit is moved toward the linear combination of the sequence of input patterns captured in y_i. The equation for the update is shown below [TSOM].

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + h_{bmu,i}(t)\mathbf{y}_i(t), \tag{4}$$

4 APPLICATIONS OF RSOM

4.1 Temporal sequence classification

The authors test RSOM with a synthetic case, which aims to underline the differences between TKM and the RSOM. Both RSOM and TKM were trained with five one dimensional input patterns {1,6,11,16,21}, with additive approximately Gaussian noise.

Figure 2 shows the results of the expreriment where the sequences of length 2 were used.

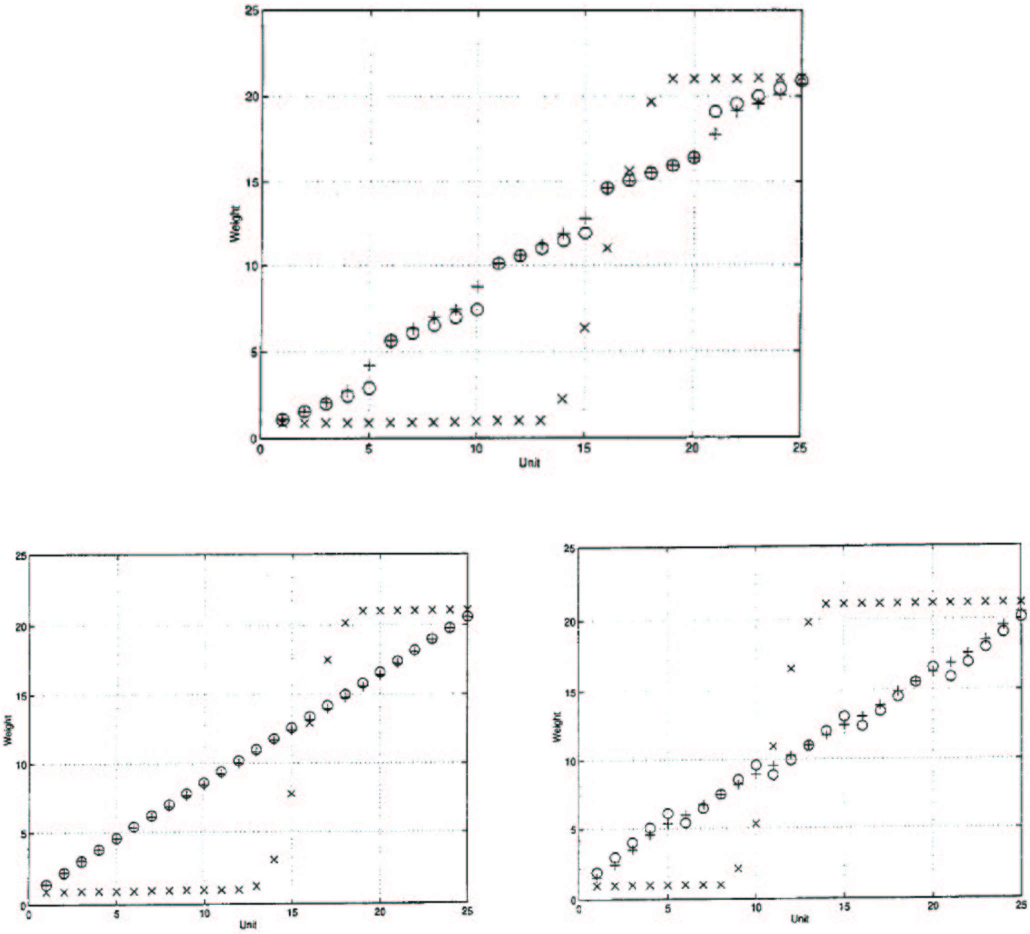

Figure 2. The weights with λ ={0.9; 0.8; 0.7} (optimal = 0.8). Optimal weights with 'o', RSOM weights with '+' and TKM weights with 'X'.

4.2 Other experiments with RSOM

The authors show two other experiments with RSOM. The first does a clustering of EEG Patterns, while the latter focuses on time series prediction.

The clustering of EEG patterns is somewhat better with RSOM compared to normal SOM. The results are not that convincing, though.

In the time series prediction, RSOM is compared to a auto regressive model (AR) and Multilayer Perceptron (MLP) model. The tested timeseries is data from measurements of the intensity of an infrared laser in a chaotic state. The MLP is almost 7 times better in prediction compared to RSOM, which in turn, is 4 times better than the AR model. The authors claim that MLP benefits from noise-free data, but they do not show an example with noisy data.

5 USING RECURRENT SOM TO MODEL SELF-ORGANIZATION IN VISUAL CORTEX

Farkas and Miikkulainen have used a similar model as RSOM, to model the self-organization of directional selectivity in the primary visual cortex [FARKAS]. It also includes a leaky integrator in the inputs of each neuron and they claim that the model is similar to RSOM.

In their model, the input comes from a square retina of *R x R* receptors and the SOM lattice is a square of *N x N* neurons. Every neuron gets its input from a receptive field (RF), which is a circular area in the retina, centered on its projection. The diameter *s* of RF is usually $s \approx (1/2)R$, and the RFs of neighboring neurons overlap significantly. The RFs of neurons near to a boundary of the lattice are not circular, but cut on one or two sides, as shown in Figure 3.

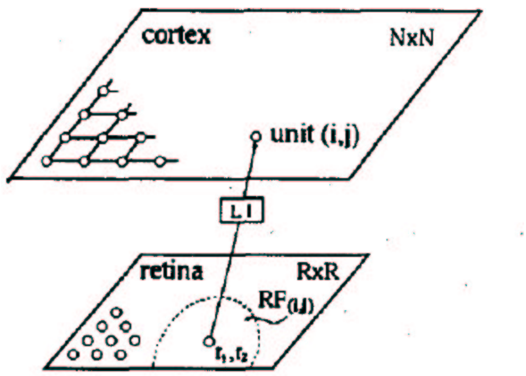

Figure 3. The architecture of the visual cortex model [FARKAS]

In their model, the activation *a* of each neuron (i,j) is at discrete time t is computed as:

$$a_i(t) = (1-\lambda)a_i(t-1) + \lambda\left(\sum_{r1,r2} I_{r1,r2} w_{i,r1r2}\right), \qquad (5)$$

where $0 < \lambda \le 1$ is the memory- or leaking coefficient parameter and $I_{r1,r2}$ is the input from input unit (*r1,r2*). The output y_i of neuron I is defined by the standard sigmoid function:

$$y_i(t) = \sigma(.) = \frac{1}{1 + \exp(-\mu(a_i(t) - \theta_i(t)))}, \tag{6}$$

where $\mu > 0$ is the slope of the sigmoid, and $\theta_i(t)$ is a threshold, that must be properly set to achieve the right amount of activaton in the map. The authors have specified that $\theta_i(t)$ is updated at every time step with $\theta_i(t) = ½ a_i^{\max}(t)$, where $a_i^{\max}(t) = \max_{\tau \le t}\{a_i(\tau)\}$.

The BMU is the one with the strongest accumulated response after the whole sequence has been presented to the network:

$$y_{bmu}(T) = \max_i \, y_i(T), \tag{7}$$

where T is the length of the sequence. After resolving the BMU, the winner and its neighborhood are updated using the Oja's rule (while they say that the standard SOM rule would also work):

$$\Delta w_{i,r1r2}(t) = \alpha(t) y_i(t) \left[I_{r1r2}^{ac} - w_{i,r1r2}(t) y_i(t) \right], \tag{8}$$

where $\alpha(t) > 0$ is the learning rate and I_{r1r2}^{ac} is the accumulated input computed at the end of sequence presentation as follows:.

$$I_{r1r2}^{ac} = \lambda \sum_{t=1}^{T} (1-\lambda) I_{r1r2}^{T-t}(t), \tag{9}$$

So in this model, all inputs are taken into account when updating the weights of the neurons. The learning is based on the accumulated inputs, and neurons are said to generate responses throughout the sequence.

5.1 Experiment

In the experiments, the input consists of moving normalized Gaussian bars, whose direction of motion is perpendicular to their orientation. Sequences have a fixed length that covers a part of the retina, and they start at randomly chosen positions in the retina. The direction is also randomly chosen from 16 possible directions with corresponding 8 orientations. The bars move in a constant speed (1 receptor per time step).

The final direction and orientation map is shown in Figure 4. It shows that almost all units are orientation selective, and most of these are also direction selective. Most of the units have orientation preference perpendicular to its direction preference.

The authors claim that the result has most of the features found in biological counterparts, including:

- Most of the orientation-selective neurons are also direction selective
- Neurons preference to a direction of motion is perpendicular to its preferred orientation

- The map contain discontinuities that have similar shape as biological conterpart

They also list some differences between the model and its biological counterpart, namely some features of the discontinuity lines and the quite high response of a neuron to direction opposite to the preferred one.

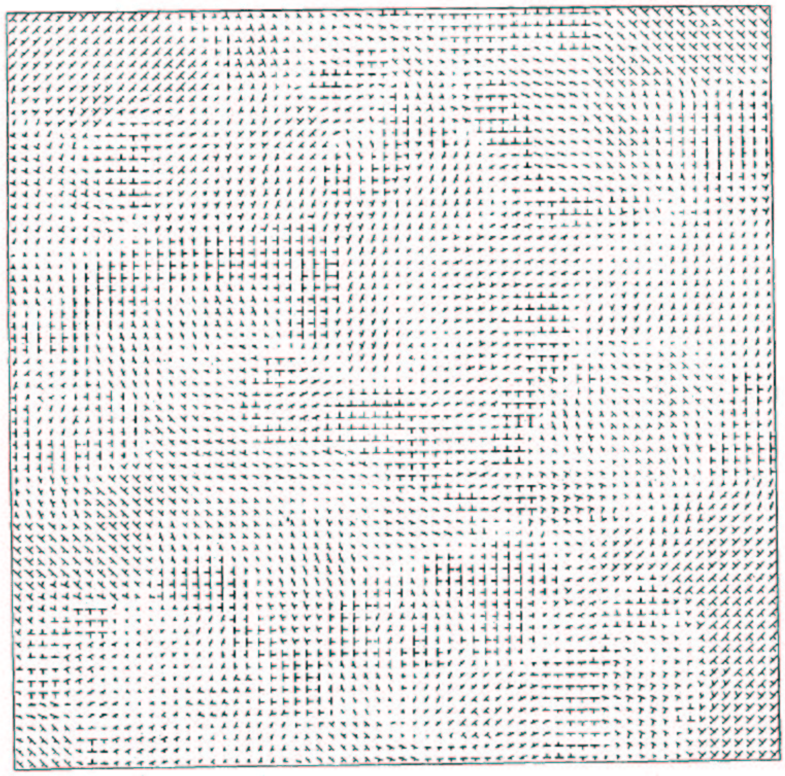

Figure 4. The self-organized direction and orientation map. [FARKAS]

6 CONCLUSIONS

RSOM is a modified TKM that seems to fix some problems in the model. The main difference to TKM is the form of the recurrent term. In RSOM the term is a vector, which enables the use of direction of error when updating the neuron weights.

A modified RSOM by Farkas et. al. has been used quite successfully to model the self-organization of the biological visual cortex.

See Table 1 for comparison of features in the TKM, RSOM and Farkas models.

Table 1: Comparison of features in TKM, RSOM, and FARKAS models.

	TKM	RSOM	FARKAS
Activity	Match between weight and input + earlier matches	Vector of accumulated error	Match between weight and input + earlier matches
Output	Activity	Length of error vector	Sigmoid of activity
Memory	Scalar	Vector	Scalar + Vector
BMU search	Highest activation	Minimum error vector	Highest activation
Update	Normal SOM	Neuron is moved towards the linear combination of inputs	Neuron is moved towards the linear combination of inputs
Update freq.	Every step	Every step	After whole seq.
Input	All neurons see all inputs	All neurons see all inputs	Only input from RF

REFERENCES

[TKM] Chappell, G.J. & Taylor, J.G. The temporal Kohonen map. Neural Networks, Vol 6, 1993, pp. 441-445

[TSOM] G de A. Barreto and A. F. R. Araujo, Time in Self-organizing maps: An overview of models. International Journal of Computer Research vol. 10, no. 2, pp. 235-259, 2000.

[RSOM] Koskela, T., Varsta M., Heikkonen, J. & Kaski, K. Time Series prediction using recurrent SOM with local linear models. International Journal of Knowledge-based Intelligent Engineering Systems, Vol. 2, No.1, 1998, pp. 60-68.

[FARKAS] Farkas, I. & Miikkulainen, R. Modeling the self-organization of directional selectivity in the primary visual cortex. Proc. 9th Int. Conf. On Arificial Neural Networks, (ICANN'96), Bochum, Germany, 1996, pp. 457-462.