# Recursive Self-organizing Map
# as a Contractive Iterative Function System

Peter Tiňo[1], Igor Farkaš[2], and Jort van Mourik[3]

[1] School Of Computer Science, University Of Birmingham
Birmingham B15 2TT, UK
[2] Faculty of Mathematics, Physics and Informatics, Comenius University
Mlynská dolina, 842 48 Bratislava, Slovak Republic
[3] Neural Computing Research Group
Aston University
Aston Triangle, Birmingham B4 7ET, UK

**Abstract.** Recently, there has been a considerable research activity in extending topographic maps of vectorial data to more general data structures, such as sequences or trees. However, the representational capabilities and internal representations of the models are not well understood. We rigorously analyze a generalization of the Self-Organizing Map (SOM) for processing sequential data, Recursive SOM (RecSOM [1]), as a non-autonomous dynamical system consisting of a set of fixed input maps. We show that contractive fixed input maps are likely to produce Markovian organizations of receptive fields on the RecSOM map. We derive bounds on parameter $\beta$ (weighting the importance of importing past information when processing sequences) under which contractiveness of the fixed input maps is guaranteed.

## 1 Introduction

Several modifications of the self-organizing map (SOM) [2] to sequences and/or tree structures have been proposed in the literature. For comprehensive reviews, see [3, 4]. Modified versions of SOM that have enjoyed a great deal of interest equip SOM with *additional feed-back connections* that allow for natural processing of recursive data types. Typical examples of such models are Temporal Kohonen Map [5], recurrent SOM [6], feedback SOM [7], recursive SOM [1], merge SOM [8] and SOM for structured data [9]. However, the representational capabilities and internal representations of the models are not well understood [3, 10, 11].

In this paper we concentrate on the Recursive SOM (RecSOM) [1], because RecSOM transcends the simple local recurrence of leaky integrators of earlier models and it has been demonstrated that it can represent much richer dynamical behavior [11]. We propose to study the RecSOM model as a non-autonomous dynamical system with internal dynamics driven by a stream of external inputs. It is argued that contractive fixed input maps are likely to produce Markovian organizations of receptive fields on the RecSOM map.

## 2  Recursive Self-organizing Map (RecSOM)

In the RecSOM model [1], Each neuron $i \in \{1, 2, ..., N\}$ in the map has two weight vectors associated with it:

- $\mathbf{w}_i \in \mathbb{R}^n$ – linked with an $n$-dimensional input $\mathbf{s}(t)$ feeding the network at time $t$
- $\mathbf{c}_i \in \mathbb{R}^N$ – linked with the context

$$\mathbf{y}(t-1) = (y_1(t-1), y_2(t-1), ..., y_N(t-1))$$

containing map activations $y_i(t-1)$ from the previous time step.

The output of a unit $i$ at time $t$ is computed as $y_i(t) = \exp(-d_i(t))$, where[1]

$$d_i(t) = \alpha \cdot \|\mathbf{s}(t) - \mathbf{w}_i\|^2 + \beta \cdot \|\mathbf{y}(t-1) - \mathbf{c}_i\|^2. \tag{1}$$

In eq. (1), $\alpha > 0$ and $\beta > 0$ are model parameters that respectively influence the effect of the input and the context upon neuron's profile. Both weight vectors can be updated using the same form of learning rule [1]:

$$\Delta \mathbf{w}_i = \gamma \cdot h_{ik} \cdot (\mathbf{s}(t) - \mathbf{w}_i), \tag{2}$$
$$\Delta \mathbf{c}_i = \gamma \cdot h_{ik} \cdot (\mathbf{y}(t-1) - \mathbf{c}_i), \tag{3}$$

where $k$ is an index of the best matching unit at time $t$, $k = \operatorname{argmin}_{i \in \{1,2,...,N\}} d_i(t)$, and $0 < \gamma < 1$ is the learning rate. Neighborhood function $h_{ik}$ is a Gaussian (of width $\sigma$) on the distance $d(i,k)$ of units $i$ and $k$ in the map:

$$h_{ik} = e^{-\frac{d(i,k)^2}{\sigma^2}}. \tag{4}$$

The 'neighborhood width' $\sigma$ decreases in time to allow for forming topographic representation of input sequences.

## 3  Contractive Fixed-Input Dynamics in RecSOM

Under a fixed input vector $\mathbf{s} \in \mathbb{R}^n$, the time evolution of (1) becomes

$$d_i(t+1) = \alpha \cdot \|\mathbf{s} - \mathbf{w}_i\|^2 + \beta \cdot \left\| \left( e^{-d_1(t)}, e^{-d_2(t)}, ..., e^{-d_N(t)} \right) - \mathbf{c}_i \right\|^2. \tag{5}$$

After applying a one-to-one coordinate transformation $y_i = e^{-d_i}$, eq. (5) reads

$$y_i(t+1) = e^{-\alpha \|\mathbf{s} - \mathbf{w}_i\|^2} \cdot e^{-\beta \|\mathbf{y}(t) - \mathbf{c}_i\|^2}, \tag{6}$$

or, in the vector form:

$$\mathbf{y}(t+1) = \mathbf{F_s}(\mathbf{y}(t)). \tag{7}$$

---

[1]  $\|\cdot\|$ denotes the Euclidean norm

Given a fixed input $\mathbf{s}$, we aim to study the conditions under which the map $\mathbf{F_S}$ becomes a contraction. Then, by the Banach Fixed Point theorem, the autonomous RecSOM dynamics $\mathbf{y}(t+1) = \mathbf{F_S}(\mathbf{y}(t))$ will be dominated by a unique attractive fixed point $\mathbf{y_S} = \mathbf{F_S}(\mathbf{y_S})$.

A mapping $\mathbf{F} : \mathbb{R}^N \to \mathbb{R}^N$ is said to be a contraction with contraction coefficient $\rho \in [0, 1)$, if for any $\mathbf{y}, \mathbf{y}' \in \mathbb{R}^N$,

$$\|\mathbf{F}(\mathbf{y}) - \mathbf{F}(\mathbf{y}')\| \leq \rho \cdot \|\mathbf{y} - \mathbf{y}'\|. \tag{8}$$

$\mathbf{F}$ is a contraction if there exists $\rho \in [0, 1)$ so that $\mathbf{F}$ is a contraction with contraction coefficient $\rho$.

We denote the Gaussian kernel of inverse variance $\eta > 0$, acting on $\mathbb{R}^N$, by $G_\eta(\cdot, \cdot)$, i.e. for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^N$,

$$G_\eta(\mathbf{u}, \mathbf{v}) = e^{-\eta\|\mathbf{u}-\mathbf{v}\|^2}. \tag{9}$$

Denote by $\mathbf{G}_\alpha(\mathbf{s})$ the collection of activations coming from the feed-forward part of RecSOM,

$$\mathbf{G}_\alpha(\mathbf{s}) = (G_\alpha(\mathbf{s}, \mathbf{w}_1), G_\alpha(\mathbf{s}, \mathbf{w}_2), ..., G_\alpha(\mathbf{s}, \mathbf{w}_N)). \tag{10}$$

Then we have the following theorem:

**Theorem 1.** *Consider an input $\mathbf{s} \in \mathbb{R}^M$. If for some $\rho \in [0, 1)$,*

$$\beta \leq \rho^2 \frac{e}{2} \|\mathbf{G}_\alpha(\mathbf{s})\|^{-2}, \tag{11}$$

*then the mapping $\mathbf{F_S}$ (7) is a contraction with contraction coefficient $\rho$.*

*Sketch of the proof:* The proof is rather lengthy and complicated. Due to space limitations, we refer the reader to [12]. The proof follows the worst case analysis of the distances $\|\mathbf{F_S}(\mathbf{y}) - \mathbf{F_S}(\mathbf{y}')\|$ between the $\mathbf{F_S}$-images of $\mathbf{y}, \mathbf{y}'$, under the constraint $\|\mathbf{y} - \mathbf{y}'\| = \delta$:

$$D_\beta(\delta) = \sup_{\mathbf{y}, \mathbf{y}'; \|\mathbf{y}-\mathbf{y}'\|=\delta} \|\mathbf{F_S}(\mathbf{y}) - \mathbf{F_S}(\mathbf{y}')\|.$$

The analysis is quite challenging, because $D_\beta(\delta)$ can be expressed only implicitly. Nevertheless, one can prove that, for a given $\beta > 0$, the function $D_\beta : \mathbb{R}^+ \to (0, 1)$ has the following properties:

1. $\lim_{\delta \to 0+} D_\beta(\delta) = 0$,
2. $D_\beta$ is a continuous monotonically increasing concave function of $\delta$.
3. $\lim_{\delta \to 0+} \frac{dD_\beta(\delta)}{d\delta} = \sqrt{\frac{2\beta}{e}}$.

Therefore, we have the following upper bound:

$$D_\beta(\delta) \leq \delta\sqrt{\frac{2\beta}{e}}. \tag{12}$$

Writing (6) as

$$y_i(t+1) = G_\alpha(\mathbf{s}, \mathbf{w}_i) \cdot G_\beta(\mathbf{y}, \mathbf{c}_i),$$

we get that if

$$\delta^2 \; \frac{2\beta}{e} \; \sum_{i=1}^{N} G_{2\alpha}(\mathbf{s}, \mathbf{w}_i) \le \rho^2 \; \delta^2, \tag{13}$$

then $\mathbf{F_s}$ will be a contraction with contraction coefficient $\rho$. Inequality (13) is equivalent to

$$\frac{2\beta}{e} \; \|\mathbf{G}_\alpha(\mathbf{s})\|^2 \le \rho^2. \tag{14}$$

$$Q.E.D.$$

**Corollary 1.** *Consider a RecSOM fed by a fixed input* $\mathbf{s}$. *Define*

$$\Upsilon(\mathbf{s}) = \frac{e}{2}\|\mathbf{G}_\alpha(\mathbf{s})\|^{-2}. \tag{15}$$

*Then, if* $\beta < \Upsilon(\mathbf{s})$, $\mathbf{F_s}$ *is a contractive mapping.*

## 4   Experiments

We illustrate our results on natural language data used to demonstrate RecSOM in [1]. The data is a corpus of written English, the novel "Brave New World" by Aldous Huxley. In the corpus we removed punctuation symbols, upper-case letters were switched to lower-case and the space between words was transformed into a symbol '-'. The complete data set (after filtering) comprised 356606 symbols. Letters of the Roman alphabet were binary-encoded using 5 bits and presented to the network one at a time. RecSOM with $20 \times 20 = 400$ neurons was trained for two epochs using the following parameter settings: $\alpha = 3$, $\beta = 0.7$, $\gamma = 0.1$ and $\sigma : 10 \to 0.5$. Radius $\sigma$ reached its final value at the end of the first epoch and then remained constant to allow for fine-tuning of the weights.

We constructed a map of the neurons' receptive fields (RFs) (shown in figure 1). Following [1], RF of a neuron is defined as the common suffix of all sequences for which that neuron becomes the best-matching unit. Note that the RF map contains regions in which RFs are topographically ordered with respect to the most recent symbols.

Let us consider, how the character of the RecSOM fixed-input dynamics (7) for each individual input symbol $s \in \mathcal{A}$ (coded as $n$-dimensional input vector $\mathbf{s} \in \mathbb{R}^n$) shapes the overall organization of the map. In order to investigate the fixed-input dynamics (7), we initialized context activations $\mathbf{y}(0)$ in 10,000 different positions within the state space $(0, 1]^N$. For each initial condition $\mathbf{y}(0)$, we checked asymptotic dynamics of the fixed input maps $\mathbf{F}_s$ by monitoring $L_2$-norm of the activation differences $(\mathbf{y}(t) - \mathbf{y}(t-1))$ and recording the limit set (after 1000 iterations).

Figure 2 illustrates asymptotic regimes of the fixed-input RecSOM dynamics (7) in terms of map activity differences between consecutive time steps. We observed a variety of behaviors. For some symbols, the activity differences converge to zero (attractive fixed points); for other symbols, the differences level at
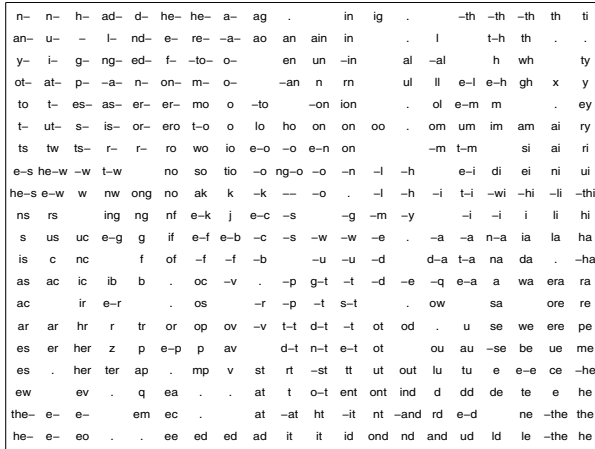
| n– | n– | h– | ad– | d– | he– | he– | a– | ag | . | | in | ig | . | | –th | –th | –th | th | ti |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| an– | u– | – | l– | nd– | e– | re– | –a– | ao | an | ain | in | | . | l | | t–h | th | . | . |
| y– | i– | g– | ng– | ed– | f– | –to– | o– | | en | un | –in | | al | –al | | h | wh | | ty |
| ot– | at– | p– | –a– | n– | on– | m– | o– | | –an | n | rn | | ul | ll | e–l | e–h | gh | x | y |
| to | t– | es– | as– | er– | er– | mo | o | –to | | –on | ion | | . | | ol | e–m | m | . | ey |
| t– | ut– | s– | is– | or– | ero | t–o | o | lo | ho | on | on | oo | . | | om | um | im | am | ai | ry |
| ts | tw | ts– | r– | r– | ro | wo | io | e–o | –o | e–n | on | | | –m | t–m | | si | ai | ri |
| e–s | he–w | –w | t–w | | no | so | tio | –o | ng–o | –o | –n | –l | –h | | e–i | di | ei | ni | ui |
| he–s | e–w | w | nw | ong | no | ak | k | –k | –– | –o | . | –l | –h | –i | t–i | –wi | –hi | –li | –thi |
| ns | rs | | ing | ng | nf | e–k | j | e–c | –s | | –g | –m | –y | | –i | –i | i | li | hi |
| s | us | uc | e–g | g | if | e–f | e–b | –c | –s | –w | –w | –e | . | –a | –a | n–a | ia | la | ha |
| is | c | nc | | f | of | –f | –f | –b | | –u | –u | –d | | d–a | t–a | na | da | . | –ha |
| as | ac | ic | ib | b | . | oc | –v | . | –p | g–t | –t | –d | –e | –q | e–a | a | wa | era | ra |
| ac | | ir | e–r | | . | os | | –r | –p | –t | s–t | | . | | ow | | sa | | ore | re |
| ar | ar | hr | r | tr | or | op | ov | –v | t–t | d–t | –t | ot | od | . | u | se | we | ere | pe |
| es | er | her | z | p | e–p | p | av | | d–t | n–t | e–t | ot | | | ou | au | –se | be | ue | me |
| es | . | her | ter | ap | . | mp | v | st | rt | –st | tt | ut | out | lu | tu | e | e–e | ce | –he |
| ew | | ev | . | q | ea | . | . | at | t | o–t | ent | ont | ind | d | dd | de | te | e | he |
| the– | e– | e– | | em | ec | . | | at | –at | ht | –it | nt | –and | rd | e–d | | ne | –the | the |
| he– | e– | eo | . | . | ee | ed | ed | ad | it | it | id | ond | nd | and | ud | ld | le | –the | he |

**Fig. 1.** Receptive fields of RecSOM trained on English text. Dots denote units with empty RFs
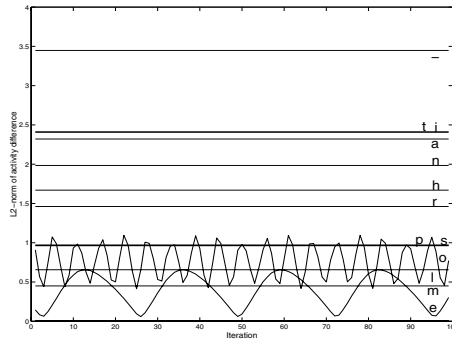


**Fig. 2.** Fixed-input asymptotic dynamics of RecSOM after training on English text. Plotted are $L_2$ norms of the differences of map activities between the successive iterations. Labels denote the associated input symbols (for clarity, not all labels are shown)

nonzero values (periodic attractors of period two, e.g. symbols 'i', 't', 'a', '-'). Fixed input RecSOM dynamics for symbols 'o' and 'e' follows a complicated a-periodic trajectory.

For each input symbol $s$, the autonomous dynamics $\mathbf{y}(t) = \mathbf{F}_s(\mathbf{y}(t-1))$ induces a dynamics of the winner units on the map:

$$i_s(t) = \underset{i \in \{1, 2, \ldots, N\}}{\operatorname{argmax}} y_i(t) \tag{16}$$

The dynamics (16) is illustrated in figure 3[2].

---

[2]  For each of the 10,000 initial conditions $\mathbf{y}(0)$, we first let the system (7) settle down by preiterating it for 1000 iterations and then mark the map position of the winner units $i_s(t)$ for further 100 iterations

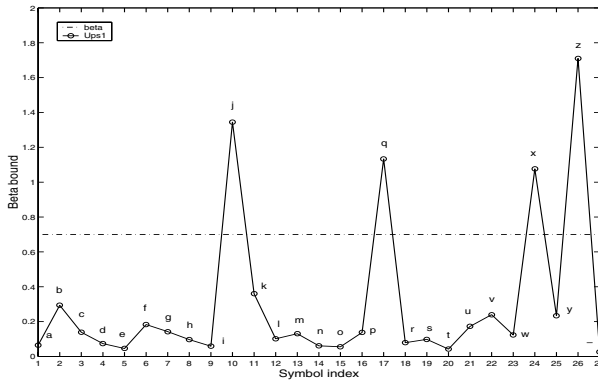**Fig. 3.** Dynamics of the winning units on the RecSOM map induced by the fixed-input dynamics. The map was trained on a corpus of written English ("Brave New World" by Aldous Huxley)

When the fixed-input dynamics for $s \in \mathcal{A}$ is dominated by a unique attractive fixed point $\mathbf{y}_s$, the induced dynamics on the map, (16), settles down in neuron $i_s$, corresponding to the mode of $\mathbf{y}_s$, $i_s = \mathrm{argmax}_{i \in \{1,2,\ldots,N\}} y_{s,i}$. The neuron $i_s$ will be most responsive to input subsequences ending with long blocks of symbols $s$. Receptive fields of neurons on the map will be organized with respect to closeness of neurons to the fixed input winner $i_s$. Assuming a unimodal character of the fixed point $\mathbf{y}_s$, as soon the symbol $s$ is seen, the mode of the activation profile $\mathbf{y}$ will drift towards the neuron $i_s$. The more consecutive symbols $s$ we see, the more dominant the attractive fixed point of $\mathbf{F}_s$ becomes and the closer the winner position is to $i_s$. In this manner, a Markovian suffix-based RF organization is created.

As evident in figure 3, for symbols $s$ with dynamics $\mathbf{y}(t) = \mathbf{F}_s(\mathbf{y}(t-1))$ dominated by a single fixed point $\mathbf{y}_s$, the induced dynamics on the map settles down in the mode position of $\mathbf{y}_s$. However, some autonomous dynamics $\mathbf{y}(t) = \mathbf{F}_s(\mathbf{y}(t-1))$ of period two (e.g. $s \in \{n, h, r, p, s\}$) induce a trivial dynamics on the map driven to a single point (grid position). In those cases, the points $\mathbf{y}^1, \mathbf{y}^2$ on the periodic orbit ($\mathbf{y}^1 = \mathbf{F}_s(\mathbf{y}^2)$, $\mathbf{y}^2 = \mathbf{F}_s(\mathbf{y}^1)$) lie within the representation region (Voronoi compartment) of the same neuron. Interestingly enough, the complicated dynamics of $\mathbf{F}_o$ and $\mathbf{F}_e$ translates into aperiodic oscillations between just two grid positions. Still, the suffix based organization of RFs in figure 1 is shaped by the underlying collection of the fixed input dynamics of $\mathbf{F}_s$ (illustrated in figure 3 through the induced dynamics on the map).

Theoretical upper bounds on $\beta$ (eq. (15)) are shown in figure 4. Whenever for an input symbol $s$ the bound $\Upsilon(s)$ is above $\beta = 0.7$ (dashed horizontal line) used to train RecSOM (e.g. symbols 'j', 'q', 'x'), we can be certain that the fixed input dynamics given by the map $\mathbf{F}_s$ will be dominated by an attractive fixed point. For symbols $s$ with $\Upsilon(s) < \beta$, there is a possibility of a more complicated dynamics driven by $\mathbf{F}_s$. Indeed, the theoretical bounds $\Upsilon(s)$ for all symbols $s$ with asymptotic fixed-input dynamics that goes beyond a single stable sink are below $\beta = 0.7$. Obviously, $\Upsilon(s) < \beta$ does not necessarily imply more complicated fixed input dynamics on symbol $s$.

**Fig. 4.** Theoretical bounds on $\beta$ for RecSOM trained on the English text

## 5    Discussion

Assume that for each input symbol $s \in \mathcal{A}$, the fixed-input RecSOM mapping $\mathbf{F}_s$ (7) is a contraction with contraction coefficient $\rho_s$. Set $\rho_{max} = \max_{s \in \mathcal{A}} \rho_s$. For a sequence $s_{1:n} = s_1...s_{n-2}s_{n-1}s_n$ over $\mathcal{A}$ and $\mathbf{y} \in (0,1]^N$, define

$$\mathbf{F}_{s_{1:n}}(\mathbf{y}) = \mathbf{F}_{s_n}(\mathbf{F}_{s_{n-1}}(...(\mathbf{F}_{s_2}(\mathbf{F}_{s_1}(\mathbf{y})))...))$$
$$= (\mathbf{F}_{s_n} \circ \mathbf{F}_{s_{n-1}} \circ ... \circ \mathbf{F}_{s_2} \circ \mathbf{F}_{s_1})(\mathbf{y}). \tag{17}$$

Then, if two prefixes $s_{1:p}$ and $s_{1:r}$ of a sequence $s_1...s_{p-2}s_{p-1}s_p...s_{r-2}s_{r-1}s_r...$ share a common suffix of length $L$, we have

$$\|\mathbf{F}_{s_{1:p}}(\mathbf{y}) - \mathbf{F}_{s_{1:r}}(\mathbf{y})\| \leq \rho_{max}^L \sqrt{N}, \tag{18}$$

where $\sqrt{N}$ is the diameter of the RecSOM state space $(0,1]^N$.

For sufficiently large $L$, the two activations $\mathbf{y}^1 = \mathbf{F}_{s_{1:p}}(\mathbf{y})$ and $\mathbf{y}^2 = \mathbf{F}_{s_{1:r}}(\mathbf{y})$ will be close enough to have the same location of the mode[3],

$$i_* = \underset{i \in \{1,2,...,N\}}{\operatorname{argmax}} \ y_i^1 = \underset{i \in \{1,2,...,N\}}{\operatorname{argmax}} \ y_i^2,$$

and the two subsequences $s_{1:p}$ and $s_{1:r}$ yield the same best matching unit $i_*$ on the map, *irrespective of the position of the subsequences in the input stream.* All that matters is that the prefixes share a sufficiently long common suffix. We say that such an *organization of RFs on the map has a Markovian flavour,* because it is shaped solely by the suffix structure of the processed subsequences, and it does not depend on the temporal context in which they occur in the input stream. Obviously, one can imagine situations where **(1)** locations of the modes of $\mathbf{y}^1$ and $\mathbf{y}^2$ will be distinct, despite a small distance between $\mathbf{y}^1$ and $\mathbf{y}^2$, or where **(2)** the modes of $\mathbf{y}^1$ and $\mathbf{y}^2$ coincide, while their distance is quite

---

[3]   Or at least mode locations on neighboring grid points of the map

large. This is the price to be paid for discontinuity of the best-matching-unit operation. However, in our extensive experimental studies, we have registered only a negligible number of such cases.

We suggest the theory of non-autonomous dynamical systems as a possible framework for studying representations of temporal structures in SOMs endowed with recursive processing mechanism. Contractive fixed input maps are likely to produce Markovian organizations of receptive fields on the RecSOM map. Periodic or aperiodic dynamics of $\mathbf{F}_s$ can result in a 'broken topography' of RFs and embody a potentially unbounded memory structure.

# References

1. Voegtlin, T.: Recursive self-organizing maps. Neural Networks **15** (2002) 979–992
2. Kohonen, T.: Self–organizing formation of topologically correct feature maps. Biological Cybernetics **43** (1982) 59–69
3. de A. Barreto, G., Araújo, A., Kremer, S.: A taxanomy of spatiotemporal connectionist networks revisited: The unsupervised case. Neural Computation **15** (2003) 1255–1320
4. Hammer, B., Micheli, A., Strickert, M., Sperduti, A.: A general framework for unsupervised processing of structured data. Neurocomputing **57** (2004) 3–35
5. Chappell, G., Taylor, J.: The temporal kohonen map. Neural Networks **6** (1993) 441–445
6. Koskela, T., znd J. Heikkonen, M.V., Kaski, K.: Recurrent SOM with local linear models in time series prediction. In: 6th European Symposium on Artificial Neural Networks. (1998) 167–172
7. Horio, K., Yamakawa, T.: Feedback self-organizing map and its application to spatio-temporal pattern classification. International Journal of Computational Intelligence and Applications **1** (2001) 1–18
8. Strickert, M., Hammer, B.: Neural gas for sequences. In: Proceedings of the Workshop on Self-Organizing Maps (WSOM'03). (2003) 53–57
9. Hagenbuchner, M., Sperduti, A., Tsoi, A.: Self-organizing map for adaptive processing of structured data. IEEE Transactions on Neural Networks **14** (2003) 491–505
10. Schulz, R., Reggia, J.: Temporally asymmetric learning supports sequence processing in multi-winner self-organizing maps. Neural Computation **16** (2004) 535–561
11. Hammer, B., Micheli, A., Sperduti, A., Strickert, M.: Recursive self-organizing network models. Neural Networks **17** (2004) 1061–1085
12. Tiňo, P., Farkaš, I., van Mourik, J.: Topographic organization of receptive fields in recursive self-organizing map. Technical Report CSRP-05-06, University of Birmingham, School of Computer Science,
http://www.cs.bham.ac.uk/∼pxt/PAPERS/tino.tr05.pdf (2005)