Detection of Anomalies and Novelties in Time Series with Self-Organizing Networks

Leonardo Aguayo† and Guilherme A. Barreto‡
† Nokia Institute of Technology (INdT)
SCS Bl. 1, Camargo Corrêa Bld., 6th floor, CEP 70397-900, Brasília, Brazil.
‡ Department of Teleinformatics Engineering, Federal University of Ceará
Av. Mister Hull, S/N, Center of Technology, Campus of Pici, Fortaleza, Ceará, Brazil
emails: leonardo.aguayo@indt.org.br, guilherme@deti.ufc.br

Keywords: Operator map, anomaly detection, time series, local linear models, adaptive filtering.

***Abstract*—** This paper introduces the DANTE project (**D**etection of **A**nomalies and **N**ovelties in **T**ime s**E**ries with self-organizing networks), the goal of which is to evaluate several self-organizing networks in the detection of anomalies/novelties in dynamic data patterns. In this paper, we first describe three standard clustering-based approaches which use well-known self-organizing neural architectures, such as the SOM and the Fuzzy ART algorithms, and then present a novel approach based on the Operator Map (OPM) network [1]. The OPM is a generalization of the SOM where neurons are regarded as temporal filters for dynamic patters. The OPM is used to build local adaptive filters for a given nonstationary time series. Nonparametric confidence intervals are then computed for the residuals of the local models and used as decision thresholds for detecting novelties/anomalies. Preliminary simulations suggest that the proposed approach consistently outperforms standard clustering-based algorithms.

1 Introduction

Anomaly detection[1] methods comprise computational procedures developed to handle the complex problem of finding data samples which are inconsistent with the already modeled set of data. Recently, it has been observed an increasing number of applications of the Self-Organizing Map (SOM) to such a problem [2, 3, 4, 5], most of them dealing with static data only, i.e. data for which the temporal dimension is an unimportant source of information.

However, several real-world applications provide data in a time-ordered fashion, usually in the form of successive measurements of the magnitude of one or several variables of interest, giving rise to time series data. In industry, for example, many process monitoring procedures involves measuring various sensor readings continuously in time to track the state of the monitored system [6, 7, 8]. In financial market modelling, as another example, stock time series may present patterns (e.g. changes in regime) that can guide an investor in his/her investment decisions in short- or long-term horizons [9].

Anomaly detection in time series is particularly challenging due to the usual presence of deterministic features, such as trend and seasonality, that can mask the patterns of novelty that may be present in data. Inherent non-stationary processes, such as regime-switching time series, also impose additional limitations on time series modeling. Furthermore, some time series may have relatively few samples, restricting the amount of data available to extract information about its behavior. Finally, time-critical applications, such as fault detection and surveillance, requires on-line anomaly detection procedures.

Traditional approaches, such as statistical parametric modeling and hypothesis testing [10], can be successfully used to model static (i.e. memoryless) patterns, as these techniques assume some degree of stationarity of the data. On the one hand, linear stationary dynamic processes can be handled by standard Box-Jenkins ARMA time series models. On the other hand, highly nonlinear and nonstationary dynamic patterns, such as chaotic or regime-switching time series, require a more powerful approach in terms of learning and computational capabilities.

At this point the use of artificial neural networks (ANNs) have shown to be useful due to their capability to act as general purpose nonlinear system identifier, generalizing the acquired knowledge to unknown data. Most of the ANN-based methods rely on supervised ANN models, such as MLP and RBF architectures [11, 12]. However, a major drawback of such models in performing anomaly detection in time series is the asymmetry on the size of training data: labeled data for training may not be always available or may be costly to collect. A plausible solution relies on the use of clustering algorithms to find subsets of data with similar temporal structure [13].

However, few clustering-based algorithms for anomaly detection have been proposed to date. In particular, considering the usage of SOM algorithm as a clustering tool for anomaly detection systems, the former assertion is even stronger. Most of the SOM-based approaches usually converts the time series into a non-temporal representation

[1]Depending on the research field, anomaly detection also comes under several designations, namely, novelty detection, outlier detection, fault detection and condition monitoring.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

(e.g. spectral features computed through Fourier transform) and use it as input to the usual SOM [14]. Another common approach is to use fixed-length tapped delay lines at the input of the SOM, again converting the time series into a spatial representation [15].

Since the early 1990's, several temporal variants of the SOM algorithm have been proposed with the aim of performing better than static clustering methods when dealing with time series data (see [16] for a review). However, to the best of our knowledge, such temporal SOMs have never been used for anomaly/novelty detection purposes. Thus, the aim of this paper is to understand how effective are these dynamical self-organizing networks in detecting anomalies or novelties in time series.

The preliminary results to be described fits the scope of the DANTE project, the goal of which is to evaluate several self-organizing networks in the detection of anomalies/novelties in dynamic data patterns. For this purpose, we first describe three standard clustering-based approaches based on well-known self-organizing neural architectures, such as the SOM and the Fuzzy ART algorithms, and then present a novel approach based on the Operator Map (OPM) network, introduced by Lampinen and Oja [1] in the late 1980's. The OPM is a generalization of the Self-Organizing Map whose neurons are regarded as temporal filters for dynamic patters.

The OPM model is used to build local adaptive filters for a given nonstationary time series. Non-parametric confidence intervals are then computed for the residuals (prediction errors) of the local models and used as decision thresholds for detecting novelties/anomalies. We compare the proposed approach with three other self-organizing networks, two of them based on the SOM (including a temporal version of it) and the Fuzzy ART network. All these algorithms are trained on-line and computer simulations are carried out to compare their performances.

The remainder of the paper is organized as follows. In Section 2 we describe the self-organizing algorithms used in this work to perform anomaly/novelty detection in time series. In this section, we also present in detail the decision-support methodology used to run the simulations. In Section 4 the numerical results and comments on the performance of all the simulated algorithms are reported. The paper is concluded in Section 5.

2 Time Series Clustering

There are many approaches to time series clustering, but we limit the scope of our description to prototype-based clustering algorithms. In what concerns the anomlay detection task, we assume that the following algorithms are trained on-line as the data is collected. The input vectors are built through a fixed-length window, sliding over the time series of interest. Thus, at time step t, the input vector is given by

$$\mathbf{x}^{+}(t) = [x(t)\ x(t-1)\ \cdots\ x(t-p+1)]^{T}, \qquad (1)$$

where $p \geq 1$ is the memory-depth parameter. Weight updating is allowed for a fixed number of steps, T_{max}.

The first two algorithms to be described are based on the SOM algorithm, while the third one belongs to the family of ART (Adaptive Resonance Theory) architectures. Once the networks are trained, decision thresholds are computed based on the quantization errors for the SOM-based methods. ART-based models have an intrinsic novelty-detection mechanism, which can also be used for anomaly detection purposes.

2.1 The Standard SOM

Usual SOM training is carried out using the vector $\mathbf{x}^{+}(t)$ as input. Thus, the winning neuron, $i^{*}(t)$, is given by

$$i^{*}(t) = \arg\min_{\forall i} \|\mathbf{x}^{+}(t) - \mathbf{w}_i(t)\|, \quad i = 1, \ldots, Q, \qquad (2)$$

where Q is the number of neurons and t denotes the current iteration of the algorithm. Accordingly, the weight vectors are updated by the following learning rule:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \eta(t)h(i^{*}, i; t)[\mathbf{x}^{+}(t) - \mathbf{w}_i(t)], \qquad (3)$$

where $h(i^{*}, i; t)$ is a gaussian function which control the degree of change imposed to the weight vectors of those neurons in the neighborhood of the winning neuron:

$$h(i^{*}, i; t) = \exp\left(-\frac{\|\mathbf{r}_i(t) - \mathbf{r}_{i^{*}}(t)\|^2}{\sigma^2(t)}\right), \qquad (4)$$

where $\sigma(t)$ defines the radius of the neighborhood function at iteration t, and $\mathbf{r}_i(t)$ and $\mathbf{r}_{i^{*}}(t)$ are the coordinates of neurons i andi^{*} in the output array, respectively. The learning rate, $0 < \eta(t) < 1$, should decay with time to guarantee convergence of the weight vectors to stable states. In this paper, we use the 1-D SOM topology, and set $\eta(t) = \eta_0 \left(\eta_T/\eta_0\right)^{-(t/T_{max})}$, where η_0 is the initial value of η, and η_T is its final value after T_{max} training iterations. The variable $\sigma(t)$ should decay in time in a similar fashion.

2.2 The Kangas' Model

Several SOM-based algorithms for time series clustering have been proposed but they have not been used for anomaly detection purposes yet. Kangas' model [17] is one of the simplest temporal SOM algorithms available, whose underlying idea consists in performing a first-order IIR filtering on the input vector $\mathbf{x}^{+}(t)$ as follows:

$$\overline{\mathbf{x}}(t) = (1-\lambda)\overline{\mathbf{x}}(t-1) + \lambda\mathbf{x}^{+}(t), \qquad (5)$$

where $0 < \lambda < 1$ is a memory decay parameter. The filtered vector $\overline{\mathbf{x}}(t)$ is then presented to the standard SOM algorithm, which is trained as described in Section 2.1.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

2.3 The Fuzzy ART Algorithm

Due to its simplicity of implementation and low computational cost, this paper also evaluates the performance of the Fuzzy ART algorithm [18] on anomaly detection in time series. The input vector $\mathbf{x}^+(t)$ is presented to a competitive layer of Q neurons. The winning neuron i^* is selected according to the following rule:

$$i^*(t) = \arg\max_{\forall i} \{T_i(t)\}, \tag{6}$$

where the choice function of neuron i, $T_i(t)$, is defined as

$$T_i(t) = \frac{|\mathbf{x}^+(t) \wedge \mathbf{w}_i(t)|}{\varepsilon + |\mathbf{w}_i(t)|}, \tag{7}$$

where $0 < \varepsilon \ll 1$ is a very small constant, and $|\mathbf{u}|$ denotes the L_1-norm of the vector $\mathbf{u}$. The symbol $\wedge$ denotes the component-wise minimum operator, i.e.

$$x_j^+(t) \wedge w_{ij}(t) \equiv \min\left\{x_j^+(t), w_{ij}(t)\right\}. \tag{8}$$

The winning neuron is then tested for *resonance*. If its weight vector satifies the following condition

$$\frac{|\mathbf{x}^+(t) \wedge \mathbf{w}_{i^*}(t)|}{|\mathbf{x}^+(t)|} \geq \rho, \tag{9}$$

then the weight vector is updated as follows:

$$\mathbf{w}_{i^*}(t+1) = \beta\left(\mathbf{x}^+(t) \wedge \mathbf{w}_{i^*}(t)\right) + (1-\beta)\,\mathbf{w}_{i^*}(t) \tag{10}$$

where the constants $0 < \rho < 1$ and $0 < \beta < 1$ are the *vigilance parameter* and the learning rate, respectively.

If the resonance test for the current winning neuron $i^*(t)$ fails, then another neuron is selected as the winner, usually the one with the second highest value for $T_i(t)$. If this neuron also fails, then the one with the third highest value for $T_i(t)$ is selected, and so on until one of the selected winning neurons $i^*(t)$ matches Eq. (9). If none of the existing prototype vectors resonates with the current input vector, then the input vector is declared *novel* and turned into a new prototype vector.

The parameter ρ controls the sensitivity of the Fuzzy ART algorithm to new input vectors. If $\rho \to 1$, more prototypes are created in the competitive layer, increasing the number of false alarms (false positives). If $\rho \to 0$, the number of prototypes decreases, increasing the number of missed detection (false negatives).

2.4 Novelty Detection Methodology

Unlike the Fuzzy ART algorithm, the SOM-based methods previously described do not have an intrinsic mechanism to detect novel or anomalous data. However, it has become common practice [2, 3, 8] to use the quantization error

$$e_q(\mathbf{x}^+, \mathbf{w}_{i^*}; t) = \|\mathbf{x}^+(t) - \mathbf{w}_{i^*}(t)\|, \tag{11}$$

as a measure of the degree of proximity of $\mathbf{x}^+(t)$ to a statistical representation of normal behavior encoded in the weight vectors of the SOM.

Once the SOM (or the Kangas' model) has been trained, we present the training data vectors once again to this network. From the resulting quantization errors $\{e_q(\mathbf{x}^+, \mathbf{w}_{i^*}; t)\}_{t=1}^N$, computed for all training vectors, we compute decision thresholds for the anomaly detection tests. For a successfully trained network, the sample distribution of these quantization errors should reflect the 'known' or 'normal' behavior of the input variable whose time series model is being constructed.

Several procedures to compute decision thresholds have been developed in recent years, most of them based on well-established statistical techniques [19]), but we apply the nonparametric method recently proposed in [3]. For a given significance level α, we are interested in an interval within which we can certainly find a percentage $100(1-\alpha)$ (e.g. $\alpha = 0.05$) of normal values of the quantization error. Hence, we compute the lower and upper limits of this interval as follows:

- **Lower Limit** (τ^-): This is the $100\frac{\alpha}{2}$th percentile[2] of the distribution of quantization errors associated with the training data vectors.
- **Upper Limit** (τ^+): This is the $100(1-\frac{\alpha}{2})$th percentile of the distribution of quantization errors associated with the training data vectors.

Once the decision interval $[\tau^-,\ \tau^+]$ has been computed, any anomalous behavior of the time series can be detected on-line by means of the simple rule:

$$\begin{array}{ll} \text{IF} & e_q(\mathbf{x}^+, \mathbf{w}_{i^*}; t) \in [\tau^-,\ \tau^+] \\ \text{THEN} & \mathbf{x}^+(t)\ \text{ is } \textbf{NORMAL} \\ \text{ELSE} & \mathbf{x}^+(t)\ \text{ is } \textbf{ABNORMAL} \end{array} \tag{12}$$

3 The Proposed Approach

The main component of the proposed method is the OPM architecture. Neurons in the OPM are regarded as mathematical *operators*, denoted generically by $G(\cdot)$, representing some kind of filters for temporal patterns. Such operators usually contain adjustable parameters, which can be tuned in an adaptive, self-organized fashion. Thus, a given operator may eventually become specialized to a certain dynamical range of the input time series.

More specifically, let us assume that at discrete time step t a given time series can be described by the following global model

$$x(t) = H\left(\mathbf{x}^-(t)\right) + \varepsilon(t) \tag{13}$$

where $\mathbf{x}^-(t) = [x(t-1)\ x(t-2)\ \cdots\ x(t-p)]^T$ is a vector comprised of p last samples of a time series, $H(\cdot)$

[2] The percentile of a distribution of values is a number N_α such that a percentage $100(1-\alpha)$ of the sample values are less than or equal to N_α.

is an unknown (possibly nonlinear) mapping, and $\varepsilon(t)$ is a random sample from a gaussian white noise process with zero mean and variance σ_ε^2. Let us also assume that the global model $H(\cdot)$ can be approximated with arbitrary accuracy by a set of Q local linear models G_i, $i = 1, \ldots, Q$ associated to the neurons in the OPM model.

Since our target application is anomaly detection in time series, we are interested in providing a good estimate $\hat{x}(t)$ of the current state of the system being monitored, $x(t)$, given $\mathbf{x}^-(t)$ and the local models $G_i(\cdot)$. Let $\hat{x}_i(t)$ be the estimate of the current state of the system computed by neuron i. Then,

$$e_i(t) = x(t) - \hat{x}_i(t), \tag{14}$$

is the *prediction error* due to neuron i. If the system is working normally, then one should expect a small value for the prediction error. Otherwise, something anomalous may be occurring.

A common choice for the local filter G_i is the linear autoregressive (AR) model. In this case, the estimate due to neuron i of the current value of the time series is given by:

$$\hat{x}_i(t) = \mathbf{w}_i^T(t)\mathbf{x}^-(t) = \sum_{j=1}^{n} w_{ij}(t)x^-(t-j), \tag{15}$$

where $\mathbf{w}_i(t) = [w_{1i}(t)\ w_{2i}\ \cdots\ w_{pi}]^T$ is the coefficient (weight) vector associated to neuron i. The winning neuron $i^*(t)$ is the one providing the best estimation of $x(t)$. In other words, the winning filter at time t is the one with the smallest absolute value for the prediction error:

$$\begin{aligned} i^*(t) &= \arg\min_{\forall i}\{|e_i(t)|\}, && (16)\\ &= \arg\min_{\forall i}\{|x(t) - \hat{x}_i(t)|\}, && (17)\end{aligned}$$

where $|u|$ denotes the absolute value of the scalar u. The quantity $e_{i^*}(t) = x(t) - \hat{x}_{i^*}(t)$ is the prediction error produced by the current winning neuron.

The learning rule for the weight vector of neuron i is a LMS-like equation, slightly modified by the inclusion of a neighborhood function:

$$\begin{aligned} \mathbf{w}_i(t+1) &= \mathbf{w}_i(t) + \eta(t)h(i^*, i; t)e_i(t)\mathbf{x}(t), && (18)\\ &= \mathbf{w}_i(t) + \eta(t)h(i^*, i; t)[x(t) - \hat{x}_i(t)]\mathbf{x}(t), \end{aligned}$$

where $h_{i^*,i}(t)$ is the neighborhood function as defined in Eq. (4). A successfully trained OPM network should fit Q local autoregressive models to a given nonstationary time series. Note that an OPM with one single neuron (i.e. $Q = 1$) is equivalent to a linear AR model.

3.1 Anomaly Detection with the OPM

In order to use the OPM for anomaly detection purposes we need to define a decision interval $[\tau^-, \tau^+]$. The computation of the lower/upper limits of this interval follows the same logic of the technique presented in Section 2.4, except for the fact that now we use the distribution of the prediction errors of the winning neurons:

- **Lower Limit** (τ^-): This is the $100\frac{\alpha}{2}$th percentile of the distribution of prediction errors $\{e_{i^*}(t)\}$.
- **Upper Limit** (τ^+): This is the $100(1-\frac{\alpha}{2})$th percentile of the distribution of prediction errors $\{e_{i^*}(t)\}$.

The decision rule for the proposed approach method is then written as follows:

$$\begin{array}{rl} \text{IF} & e_{i^*}(t) \in [\tau^-,\ \tau^+], \\ \text{THEN} & x(t)\ \text{ is } \textbf{NORMAL} \\ \text{ELSE} & x(t)\ \text{ is } \textbf{ABNORMAL} \end{array} \tag{19}$$

4 Simulations

The feasibility of the proposed approach is evaluated using input signals derived from four different dynamic systems, three of them realizations of chaotic series. The first one is composed by the x component of the Lorenz system of equations

$$\dot{x} = \sigma_L(y - x),\ \dot{y} = x(\alpha_L - z) - y,\ \dot{z} = xy - \epsilon_L z, \tag{20}$$

which exhibits chaotic dynamics for $\sigma_L = 10$, $\alpha_L = 28$ and $\epsilon_L = 8/3$. The second and third signals are generated from the Mackey-Glass differential equation

$$\dot{x} = Rx(t) + \frac{Px(t-\tau)}{(1 + x(t-\tau)^{10})}, \tag{21}$$

for two distinct values of the delay τ. We set $P = 0.2$, $R = -0.1$, $\tau = 17$ (second signal) and $\tau = 35$ (third signal). The fourth signal is a linear $AR(2)$ process:

$$x(n+1) = 1.9x(n-1) - 0.99x(n-2) + n(t), \tag{22}$$

with $n(t)$ is a random sample from a gaussian white noise process with zero mean and variance $\sigma_n = 10^{-3}$. Figure 1 depicts 300 samples of each signal.

The novelty detection experiment is designed to perform the on-line detection of an anomalous signal, after training the networks with a sequence considered **NORMAL**. This role is assigned to the Lorenz series, leaving the two Mackey-Glass sequences and the AR process as representatives of **ABNORMAL** time series. For the sake of clarity, all different testing sequences are presented sequentially, i.e. a set of k samples from each series is used as input to the four networks, followed by k samples of the next series.

Figure 2 shows the prediction error $e_{i^*}(t)$ collected from the winning neurons i^* of the OPM network, when the first $k = 1000$ samples of the training set consisted of samples generated by the Lorenz equations. It is worth noting the low prediction errors for the first k samples of the time series, revealing the suitability of the OPM in modelling normal behavior. Note also that when a different time series

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

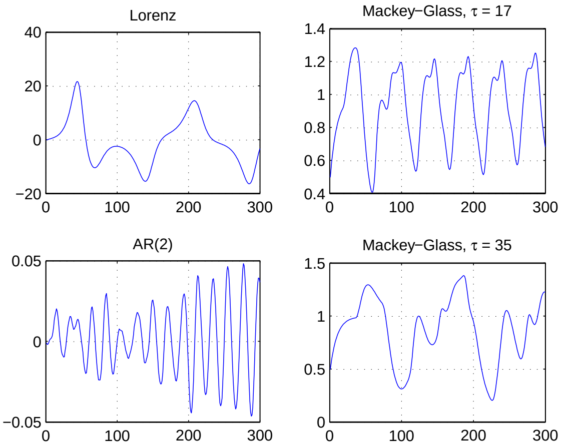

Figure 1: Samples of time series used in the simulations.

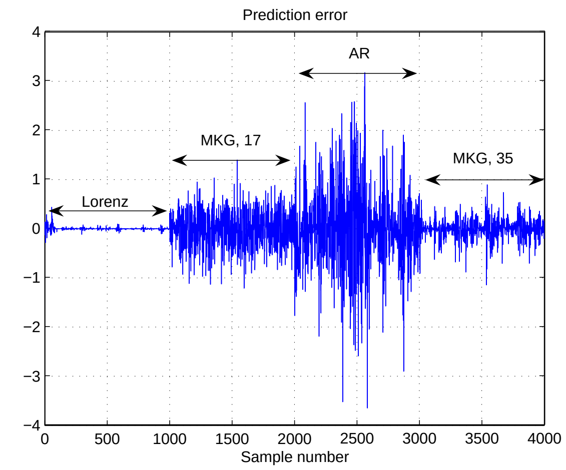

Figure 2: Prediction error $e_{i^*}(t)$. Each testing sequence has $k = 1000$ samples.

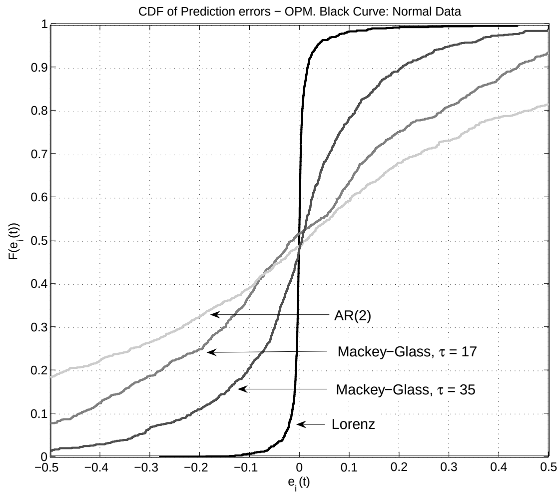

Figure 3: Cumulative distributions for prediction errors.

pattern is presented, the prediction errors are considerably higher.

Before applying the methodology described in Section 3, it is illustrative to observe the cumulative distribution function (CDF) of the prediction errors for the OPM network. Figure 3 depicts the CDFs for $e_{i^*}(t)$ obtained from all the different testing sequences, where it is possible to verify that **ABNORMAL** behavior results in distributions with higher variance.

A comparative analysis of the performances of the OPM, SOM, Kanga's model and Fuzzy-ART models can be achieved straightforwardly using the percentages of true positive and false positive rates. A true positive (TP) is the correct detection of an **ABNORMAL** sample $x(t)$ when the testing signal belongs to novel/abnormal pattern. A false positive (FP) is the incorrect detection of novelty when a testing sample belongs to the training set (i.e. it has an already modeled normal dynamics). The point with coordinates (FP, TP) is a point in the *Receiver Operating Characteristic* (ROC) space, and can be used to visually identify good and bad classifiers. For instance, a perfect binary classifier should achieve the (1,0) point at ROC space. Now, if we change the percentile N_α, the decision interval $[\tau^-, \ \tau^+]$ is modified, and a set of points in ROC space can be derived, allowing the performance comparison to be evaluated under different degrees of tolerance for the prediction error.

Additionally, different configurations for the input signals and network setup are analyzed, using variations on general parameters, such as the size of memory-depth p, noise variance σ_ε^2, number of neurons Q, as well specific parameters, such as Fuzzy-ART ρ and β and Kanga's memory decay λ. Figure 4 shows a typical result of the performance comparison among the networks, obtained for $Q = 40$ and $p = 30$.

The best performance is achieved by the proposed method, followed closely by Kanga's model and the Fuzzy ART network. It is worth noting that the simple filtering procedure implemented by Eq. (5) improves considerably the performance of the Kanga's model, when compared to the standard SOM model. The Fuzzy ART also performed closely to OPM and Kanga's models, even with no explicit mechanism to process time series data. However, its inherent novelty detection procedure, controled by the vigilance test shown in Eq. (9) can explain its good performance. Finally, the excellent performance of the OPM model can in part be explained by the neighborhood structure it inherited from the SOM. Instead of updating the weights of a single filter per input vector as in standard bank of filters, the OPM makes use of the SOM's cooperative-competitive philosophy to jointly update the winning filter and its neighboring filters. As time goes by, the neighborhood radius is decreased in order to stabilize learning.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

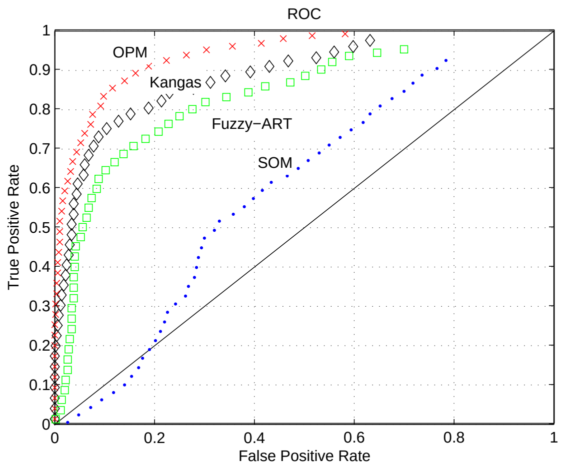

Figure 4: ROC curves for the SOM variations.

5 Conclusions

This paper presented preliminary results of the DANTE project, whose goal is to devise and evaluate self-organizing models for detecting novelties or anomalies in univariate time series. We introduced a novel approach based on the Operator Map (OPM) network, a generalization of the SOM whose neurons are regarded as temporal filters for dynamic patters. The proposed approach uses the OPM model to build local adaptive filters for a given nonstationary time series. Non-parametric confidence intervals are then computed for the residuals (prediction errors) of each local model and used as decision thresholds for detecting novelties/anomalies. We compared the proposed approach with standard clustering-based approaches which are based on other well-known self-organizing neural architectures, such as the SOM and the Fuzzy ART algorithms, to the same problem. Simulations suggested that the proposed approach consistently outperforms standard clustering-based algorithms.

Acknowledgments - The authors thank CAPES/PRODOC and INdT for their financial support.

References

[1] J. Lampinen and E. Oja. Self-organizing maps for spatial and temporal AR models. In *Proceedings of the 6th Scandinavian Conference on Image Analysis (SCIA'89)*, pages 120–127, Helsinki, Finland, 1989.

[2] S. T. Sarasamma and Q. A. Zhu. Min-max hyperellipsoidal clustering for anomaly detection in network security. *IEEE Transactions on Systems, Man and Cybernetics*, B-36(4):887–901, 2006.

[3] G. A. Barreto, J. C. M. Mota, L. G. M. Souza, R. A. Frota, and L. Aguayo. Condition monitoring of 3G cellular networks through competitive neural models. *IEEE Transactions on Neural Networks*, 16(5):1064–1075, 2005.

[4] H.-J. Lee and S. Cho. SOM-based novelty detection using novel data. *Lecture Notes on Computer Science*, 3578:359–366, 2005.

[5] S. Singh and M. Markou. An approach to novelty detection applied to the classification of image regions. *IEEE Transactions on Knowledge and Data Engineering*, 16(4):1041–4347, 2004.

[6] F. Zorriassatine, A. Al-Habaibeh, R.M. Parkin, M.R. Jackson, and J. Coy. Novelty detection for practical pattern recognition in condition monitoring of multivariate processes: a case study. *International Journal of Advanced Manufacturing Technology*, 25(9-10):954–963, 2005.

[7] S. L. Jamsa-Jounela, M. Vermasvuori, P. Enden, and S. Haavisto. A process monitoring system based on the Kohonen self-organizing maps. *Control Engineering Practice*, 11(1):83–92, 2003.

[8] E. Alhoniemi, J. Hollmén, O. Simula, and J. Vesanto. Process monitoring and modeling using the self-organizing map. *Integrated Computer Aided Engineering*, 6(1):3–14, 1999.

[9] M. Olteanu, J. Rynkiewicz, and B. Maillet. Non-linear analysis of shocks when financial markets are subject to changes in regime. In *Proceedings of the European Symposium on Neural Networks (ESANN'2004)*, pages 87–92, 2004.

[10] M. Markou and S. Singh. Novelty detection: a review - part 1: Statistical approaches. *Signal Processing*, 83:2481–2497, 2003.

[11] M. Markou and S. Singh. Novelty detection: A review - part 2: Neural network based approaches. *Signal Processing*, 83:2499–2521, 2003.

[12] C. L. Fancourt and J. C. Principe. On the use of neural networks in the generalized likelihood ratio test for detecting abrupt changes in signals. In *Proceedings of the 2000 IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)*, volume 3, pages 243–248, 2004.

[13] T. W. Liao. Clustering of time series data - a survey. *Pattern Recognition*, 38(11):1857–1874, 2005.

[14] M.L.D. Wong, L.B. Jack, and A.K. Nandi. Modified self-organising map for automated novelty detection applied to vibration signal monitoring. *Mechanical Systems and Signal Processing*, 20(3):593–610, 2006.

[15] T. C. Fu, F. L. Chung, V. Ng, and R. Luk. Pattern discovery from stock time series using self-organizing maps. In *Workshop Notes of KDD2001 Workshop on Temporal Data Mining*, pages 27–37, 2001.

[16] G. A. Barreto and A. F. R. Araújo. Time in self-organizing maps: An overview of models. *International Journal of Computer Research*, 10(2):139–179, 2001.

[17] J. Kangas. Time-delayed self-organizing maps. In *Proceedings of the International Joint Conference on Neural Networks*, volume 2, pages 331–336, 1990.

[18] G. A. Carpenter, S. Grossberg, and D. B. Rosen. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4(6):759–771, 1991.

[19] V. J. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.