

Neural detection of parameter changes in a dynamic system using time-frequency transforms

E. Swiercz

Electrical Faculty, Bialystok Technical University, Poland

Abstract

In this paper a neural detector of internal parameter changes in a stationary, non-linear SISO dynamic system, represented by a discrete model of the NARX type, is considered. The system analysed in this paper is described by the nonlinear difference equation $y(n) = f(y(k-1), \dots, y(k-p), u(k), \dots, u(k-q), \Theta)$, where f is a non-linear function, $y(k), \dots, y(k-p)$ are the output samples, $u(k), \dots, u(k-q)$ are the input samples and Θ is a vector of internal parameters of the system. The values of the vector Θ can be changed in random moments of time, but these values belong to the finite set, so that detection of parameter changes can be considered as classification of signals acquired for different values of changeable parameters. Such a formulation of the problem can be suitable in industrial applications where the change of parameters can model selected faults (or changes of an operating point) of an industrial dynamic system. To decrease dimensionality of classified data, extraction of specific characteristics from a time-frequency transform of an output signal, produces a vector of features Φ , which constitutes the decision space for classification. As the intelligent approach to such a complex problem is justified, the extracted signal features are the inputs of a neural network. The LVQ (Learning Vector Quantisation) neural network has been chosen because of its ability of learning data classification, where the similar input vectors are grouped into a region represented by a so-called coded vector (CV). Such an approach corresponds to pointing out the most probable values of the vector Θ . The detection ability of the LVQ network, both in a non-noisy and noisy environment, has been examined in detail in the paper.

Keywords: time-frequency transforms, detection of abnormal states of dynamic system, LVQ neural classifier.



1 Introduction

The complexity of the systems that we use on a day-to-day basis is constantly growing. Hence an abnormal work of a system can occur more frequently, so that a variety of computer aided methods for evaluation of system reliability is strongly required. Automatic detection of changes in a system state is an important subject, because such a change can reflect undesired faults of a system. Early detection of these changes allows one to protect a system, for example by changing a control algorithm. The concept of detection of a current system state presented in this paper assembles some methods and algorithms of signal processing, feature extraction and feature classification with the use of neural networks. Mathematically, a state of a discrete SISO system can be described by a set of difference state equations and an output equation with a specific collection of parameters Θ

$$\begin{aligned} x(k+1) &= \Lambda(x(k), u(k), \Theta) \\ y(k) &= \Gamma(x(k), u(k), \Theta) \end{aligned} \quad (1)$$

or a direct input–output difference equation with another equivalent set of parameters, Ξ

$$y(k) = f(\Delta^p y(k), \dots, y(k), \Delta^q u(k), \dots, u(k), \Xi), \quad (2)$$

where f is a nonlinear function.

From a signal point of view, tracking parameter changes could be understood as detection of non-stationarity of a system in a long time horizon. In this paper a nonlinear dynamic model with changes of parameters in random instances of time is analysed. It can be assumed that each change of parameter values creates a new non-nominal model of a dynamic system. It is also assumed that there is a finite number N of changes hence the same number of corresponding collection of models (classes) $\{\omega_1, \omega_2, \dots, \omega_N\}$ is formed. Thus the detection of parameter changes can be formulated as a multi-model classification. To categorise a current system state into classes, a vector capturing unique features of a system has to be created, based on observation of the output signal $y(n)$. Afterwards a classifier is applied to feature vectors to assign data to one of several classes.

Most of classification algorithms require probabilistic information: $P(\omega_i)$ - class prior probability, $p(y|\omega_i)$ - class-conditional density and $P(\omega_i|y)$ - posterior probability, which is rarely given a priori [1]. The stochastic classification rules use the most frequently following approaches: discriminant functions, the optimal Nearest Neighbour classifiers. Each of these rules has some advantages and disadvantages. The other approaches utilise expert systems or other artificial intelligence techniques, such as neural networks and fuzzy logic [2].

In the present study, in order to classify the signals, a feature vector Φ is formulated by a time-frequency processing of an output signal. Time-frequency representations originated from the Wigner-Ville transformation, require a troublesome optimisation of a transformation kernel that leads to the minimum classification error. Due to the great number of data possible to obtain after such a processing, an approach resulting in compressed data is required. Feature



extraction is a method of data dimension reduction. Both the discrete wavelet transform (DWT) and the continuous wavelet transform (WT), as affine time-frequency transformations, are frequently used for this task [3]. The main advantages of wavelets is that they have a varying window size, being wide for slow frequencies and narrow for the fast ones. Thus it leads to an optimal time-frequency resolution in all frequency ranges. Furthermore, owing to the fact that windows are adapted to the transients of each scale, wavelets are able to process non-stationary signals.

A lot of papers address DWT as a discrete decomposition with multi-scale wavelet transforming of signals for features extraction. Unlike to continuous wavelet algorithms, discrete algorithms are represented by a collection of a finite number of decomposition coefficients what is a compressed form for a signal representation. A vector Φ could be for example: mean of the absolute values of the wavelet decomposition coefficients, maximum of the absolute values of the wavelet coefficients, average power of the wavelet coefficients, standard deviation of the wavelet, the absolute sum of the wavelet coefficients at each resolution level, ratio of the absolute mean values of adjacent sub-bands, distribution distortion of the coefficients [4–7]. By signal processing methods we could reduce the signal (i.e., the original waveform) to a lower dimension represented by a vector Φ . Next a vector Φ has to be entered to an input of a neural LVQ classifier, performing an appropriate classification and pointing out the most probable values of vector parameters Θ .

2 The formulation of the problem

The approach proposed in the paper consists of four steps:

- exciting of a system by non stationary signals;
- transforming an output signal through chosen time-frequency transforms;
- extracting the feature vector Φ from characteristic points of a time-frequency transformation;
- detection and classification of the vector Θ by an LVQ network.

An important problem in the successful detection, treated here as the classification, is a choice of an excitation signal of a dynamic system which allows to enhance unique properties of a system. An excitation signal should be located in an essential frequency band of a dynamic system. In the presented method, an excitation signal $u(k)$ consists of two Gaussian atoms, well-separated both in time domain and in the time-frequency plane, called non-stationary Gaussian atoms.

The Gaussian atom is a short signal with a Gaussian envelope modulated by the signal $m(t)=e^{j\nu t}$. An excitation signal is non-stationary because it cannot be written as a discrete sum of sinusoids [8].

$$u(t) = \sum_{k \in N} A_k \exp[j(2\pi\nu_k t + \phi_k)]. \quad (3)$$

Frequencies of exciting atoms have to be selected very carefully, according to frequency properties of a system. A choice of atoms (generally not only two) as



the excitation signal allows one to expect that after the non-linear processing performed by a dynamic system, atoms are equally well separated. From this point of view the output signal is also non-stationary. Time-frequency processing of atoms after non-linear processing also preserves separation of localisation in the time-frequency plane and emphasizes even subtle differences resulting from parameter changes, which are invisible in the time domain. Selection in a specific way a finite number of characteristic points of the time-frequency transformation allows to formulate a low dimension feature vector Φ . In this study the wavelet transform $T_x(t, a, \Psi)$ as an example of time-frequency processing is considered

$$T_x(t, a; \Psi) = \int_{-\infty}^{+\infty} y(s) \Psi_{t,a}^*(s) ds. \quad (4)$$

The set of wavelets $\Psi_{t,a}(s)$ are created in the following way

$$\Psi_{t,a}^*(s) = |a|^{-1/2} \Psi\left(\frac{s-t}{a}\right). \quad (5)$$

The variable a represents the scale whereas t is the translation of the mother-wavelet. Calculating wavelet coefficients at every possible scale generates a lot of data. If we choose scales based on powers of two, so-called *dyadic* scales, we can formulate an indexed family of wavelets from the mother wavelet function in the form

$$\Psi_{i,j}(x) = 2^{-i/2} \Psi(2^{-i} x - j). \quad (6)$$

The wavelets defined in such a way play a substantial role in multiresolution decomposition of a signal in different decomposition levels. The mother wavelet function Ψ for $i=0, j=0$ must satisfy the multiresolution condition related to the scale equation $\phi(x)$

$$\phi(x) = \sqrt{2} \sum_k h(k) \phi(2x - k), \quad (7)$$

$$\Psi(x) = \sqrt{2} \sum_k g(k) \phi(2x - k), \quad (8)$$

where h and g can be viewed as filter coefficients of half band low-pass and high-pass filters, respectively. J -level wavelet decomposition can be computed as follows

$$f_0(x) = \sum_k a_{j+1,k} \phi_{j+1,k}(x) + \sum_{j=0}^J d_{j+1,k} \Psi_{j+1,k}(x), \quad (9)$$

where $a_{j,k}$ and $d_{j,k}$ are approximation (AC) and details coefficients (DC). Discrete wavelet transform can be defined by a collection of approximation and details coefficients. In each decomposition level a signal is divided into two sub-bands by two filters: a low-pass and a high-pass. Approximation coefficients correspond to a low band, detail coefficients correspond to a high band



respectively. The selection of an appropriate wavelet and the number of decomposition levels is a key problem, because each scale of decomposition represents a particular coarseness and unique properties of the signal under the study [9]. Multiresolution analysis leads to hierarchical and fast algorithms of computing of DWT coefficients.

In this paper the following quantities were chosen for creating the vector Φ : energies of approximation and details coefficients in sub-bands for selected levels of decomposition, means of wavelet coefficient absolute values in sub-bands for selected levels of decomposition, variances of wavelet coefficients in sub-bands for selected levels of decomposition, average power of decomposition coefficients in sub-bands for selected levels of decomposition, ratios of averages of mean details coefficients in sub-bands. It is also assumed that parameters of nonlinear system from a set $\Theta = (\theta_1, \theta_2, \dots, \theta_M)$ can be changed in a stepped way in random time instances k_i creating non-nominal models corresponding to N classes ω_i , $i = 1, \dots, N$. Between changes the parameter values are constant. The goal of the classifier is the selection of the most probable value of the element from the vector Θ , which corresponds to the selection of the class ω_i . Time interval between changes $\langle k_i, k_{i+1} \rangle$ is enough long to perform classification.

2.1 Neural identifier of parameters values

Learning Vector Quantisation (LVQ) is a supervised version of vector quantisation, similar to Self-Organising Maps (SOM) based on the Kohonen's works [10, 11]. Classes for each input pattern are predefined. The goal of the LVQ algorithm is to define class boundaries based on prototypes, covering the input space of samples in such a way that the boundaries divide the space, creating the best approximation of regions occupied by data belonging to each class. Prototypes are also called 'codebook vectors' (CV); each of them represents a region labelled with a class. They are localised in the centre of a class or a decision region ('Voronoi cell') in the input space. A class can be represented by an arbitrarily number of CVs, but one CV represents one class only. The division (tessellation) of the input space performed by the set of CVs is optimal if all data within one cell belong to the same class.

From the neural networks point of view, the LVQ network is built as a feedforward net with one hidden layer of neurons, fully connected with the input layer. A CV can be described as a hidden neuron ('Kohonen neuron') or a weight vector of the weights between all input neurons and the regarded Kohonen neuron. Coded vectors are constructed during supervised learning of a network. Learning modifies the weights in accordance with adapting rules and changes the position of a CV in the input space. The classification after learning is relied on finding a Voronoi cell, specified by the CV with the smallest distance to the input vector and assigning it to the labelled class. The most frequently the Euclidean distance is used for comparison between an input vector and the class representatives. The node of a particular class which has the smallest distance is declared to be the winner.



3 Numerical experiments

The dynamic system under study is the second order non-linear system with the state equation given by eqn. (10) and a vector of internal parameters $\Theta = (a, b, c, d)$. Nominal values of the system parameters $[a, b, c, d] = [1.0, 1.0, 3.0, 1.5]$.

$$\begin{aligned}x_1(k+1) &= x_2(k) \\x_2(k+1) &= b x_2(k) + c x_1^3(k) + d x_1^2(k) + a u(k). \\y(k) &= x_1(k)\end{aligned}\quad (10)$$

Graphs of the system output equilibrium points for the nominal set of parameters versus a constant input signal u (i.e. static characteristics of the system) and the family of amplitude frequency characteristics (system gain for a harmonic input of a single, changeable frequency) are other forms of system description, fig. 1.

In numerical experiments performed in this study it was assumed that c was the only changeable parameter and could change in the range $[-2; +2]$ around its nominal value. It was also assumed that the current value of that parameter would be assigned to the five-element set, $\Theta_c \in \{1; 2; 3; 4; 5\}$, i.e. the actual value of c "around" 2 will be identified as $c_{des} = 2$. In that way five classes (clusters) were established in the range of variability of c parameter; the values closed to nominal represented one of these classes. The characteristics of the system over the range of parameter changes are presented graphically in fig. (2).

In the series of extensive numerical experiments, the system was excited by two Gaussian atoms with frequencies from the essential frequency band of the system, what can be seen in fig. 1.

From examination of a variety of features of wavelet decomposition of the output signal, it turned out that Daubechies wavelets (*db* family), symlets wavelets (*sym* family) and reverse biorthogonal spline wavelets (*rbio* family) have the best separation properties to be exploited for the classification task considered in this study.

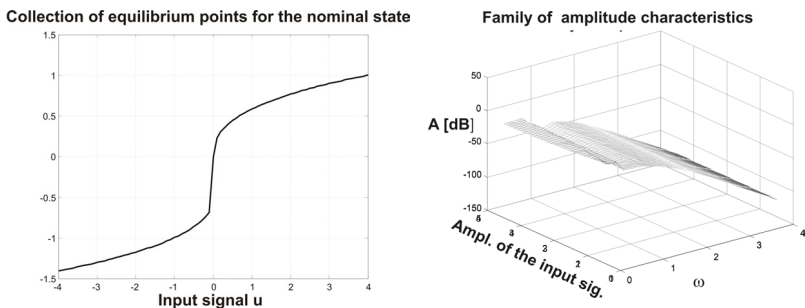


Figure 1: Examples of graphs describing the nonlinear system.

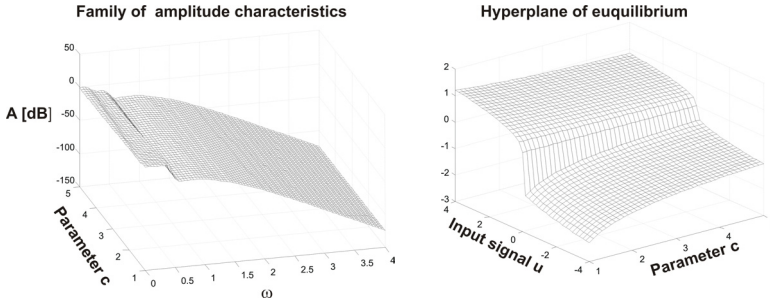


Figure 2: Examples of graphs describing parameter changes of the system.

To prepare training data for the LVQ classifier, the set of 205 system output signals (41 per each class) has been obtained for simulations. To obtain a set of signals representing each class (corresponding to the class centre value $c_{des} \in \Theta_c$), simulations were performed for c changing in the range $[-0.4; +0.4]$ around the centres, with the step equal to 0.02.

Then the discrete wavelet decomposition of the output signals were performed up to the fourth level decomposition for the wavelet families mentioned above: Daubechies of order 3 (*db3*), symlets of order 3 (*sym3*) and reverse biorthogonal spline of order 2.2 (*rbio2.2*).

The feature extraction stage consisted of computing the following sets of values: energies of details coefficients, means of coefficient absolute values, variances of decomposition coefficients, average power of decomposition coefficients, ratios of averages of mean details coefficients. Detailed examination of the above characteristic values revealed that the combinations of selected subsets of them are sufficiently good to create classification space, i.e. the feature vectors to be passed to LVQ inputs. Fig. 3 shows the examples of characteristic variables extracted from wavelet decomposition (with *db3* employed at level 4), which reveal good separation properties.

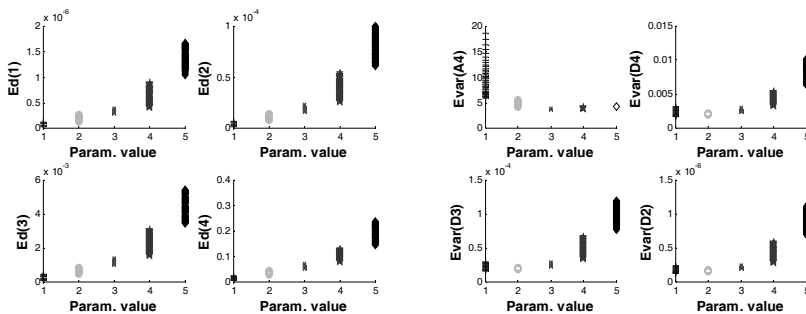


Figure 3: Energies of detail coefficients and variances of coefficients for *db3_lev4* up to fourth level decomposition.

Table 1: Classification accuracy [%] – *db3* wavelets, 4-th decomposition level.

Input data/ Competitive neurons	Input noise variance σ^2								
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.09	0.10
Coefficients energy 25 neurons	99.5	100.0	99.0	98.5	97.6	96.1	93.7	90.2	88.3
Coefficients energy 15 neurons	98.5	98.5	98.1	98.1	98.1	96.1	95.1	91.7	89.8
Coefficients variance 25 neurons	99.5	100.0	99.0	98.5	97.6	96.1	93.7	90.2	88.3
Coefficients variance 15 neurons	98.5	98.5	98.1	98.1	98.1	96.1	95.1	91.7	89.8

Table 2: Classification accuracy [%] – *rbio* wavelets, 4-th decomposition level.

Input data/ Competitive neurons	Input noise variance σ^2								
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.09	0.10
Coefficients energy 25 neurons	99.0	98.5	98.1	97.6	96.6	96.1	96.1	95.1	95.1
Coefficients energy 15 neurons	98.5	97.5	97.6	97.1	96.1	95.6	95.1	94.6	94.1
Coefficients variance 25 neurons	100.0	99.5	99.0	98.5	98.1	97.6	97.1	96.1	96.1
Coefficients variance 15 neurons	99.5	99.0	98.1	98.1	96.6	96.1	95.1	94.1	93.2

Among several LVQ architectures examined, two of them: with 15 neurons (3 per class) and with 25 neurons (5 per class) in the competitive layer shown the best performance during the training. For a variety of feature combinations, classification accuracy obtained by the above networks during the training ranged from 97% to 100%.

For further analysis, the above LVQ structures with the five-elements input vectors created from mean energies of coefficients (at approximation and details levels) and coefficient variances (at the same levels) have been chosen. Tables 1, 2 and 3 show classification accuracies of the networks when the excitation signal is contaminated with white noise (random numbers of zero mean and normal distribution) of increased variance. These cases correspond to the situation, when the desired, exact values of the input flow cannot be provided due to inaccuracies of actuators functioning in the system. As it could be expected, increased noise

Table 3: Classification accuracy [%] – *sym* wavelets, 4-th decomposition level.

Input data/ Competitive neurons	Input noise variance σ^2								
	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.09	0.10
Coefficients energy 25 neurons	99.5	100.0	99.0	98.5	97.6	96.1	93.7	90.2	88.3
Coefficients energy 15 neurons	98.5	98.5	98.1	98.1	98.1	96.1	95.1	91.7	89.8
Coefficients variance 25 neurons	99.5	100.0	99.0	98.5	97.6	96.1	93.7	90.2	88.3
Coefficients variance 15 neurons	98.5	98.5	98.1	98.1	98.1	96.1	95.1	91.7	89.8

Table 4: Classification accuracy [%] for extended range of parameter variability. Wavelet decomposition at 4-th level. Input noise variance $\sigma^2 = 0$.

Wavelet family							
	<i>db3</i>	<i>rbio</i>	<i>sym</i>	<i>db3</i>	<i>rbio</i>	<i>sym</i>	
Number of neurons in Kohonen layer							
	25			15			
Range of parameter changes							
	Classification features – mean energy of coefficients						
(0.40; 0.45]	70.0	70.0	70.0	70.0	70.0	70.0	70.0
(0.45; 0.50]	63.0	60.0	63.0	62.0	60.0	62.0	62.0
	Classification features – mean variance of coefficients						
(0.40; 0.45]	70.0	66.0	70.0	70.0	70.0	70.0	70.0
(0.45; 0.50]	63.0	60.0	63.0	62.0	63.0	62.0	62.0

variance causes deterioration of classification, but even with relatively large noise variance, the classification quality is acceptable (especially for reverse biorthogonal spline wavelets).

The next experiment examined the extrapolation ability of the LVQ classifiers (however good performance of the network outside the training range should not be expected in any case). Table 4 shows the accuracy for values of *c* parameter, which stay outside the interval (i.e. symmetrically on both sides of the central point from \mathcal{O}_c), from which the values for network training have been chosen. The border ± 0.5 means that parameter *c* having such a value can be equally likely assigned to two neighbouring classes. As it can be seen, the classification accuracy for *c* values from outside the training range significantly decreases.

4 Conclusions

Detection of parameter changes in a nonlinear dynamic system and identification of values of a changeable parameter, via classification of features extracted from

the output signal, has been considered. The detection scheme consisted of system excitation with a non-stationary signal, data pre-processing with the use of discrete wavelet transform, feature extraction by aggregation of properties of wavelet coefficients and intelligent classification with the use of LVQ networks. Simulation experiments confirmed that wavelet decomposition has the ability to separate signal features for different ranges of parameter values. Also, that the classification system is robust to noise and has certain extrapolation ability.

Acknowledgements

This work was supported by the Bialystok Technical University research project No. W/WE/3/07.

References

- [1] Duda, R. O., Hart, P. E. & Stork, D. G., *Pattern Classification*, John Wiley & Sons, New York, second edition, 2001.
- [2] DePold, H. R. & Gass, F. D., The application of expert systems and neural networks to gas turbine prognostics and diagnostics. *ASME Journal of Engineering for Gas Turbines and Power*, **121** (4), pp. 607-612, Apr. 1999.
- [3] Wu, J. D. & Chen, J. C., Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines. *NDT&E International*, **39**, pp. 304–311, 2006.
- [4] Guler, I. & Ubeyli, E. D., A modified mixture of experts network structure for ECG beats classification with diverse features. *Engineering Applications of Artificial Intelligence*, **18**, pp. 845–856, 2005.
- [5] Al-Assaf, Y., Recognition of control chart patterns using multi-resolution wavelets analysis and neural networks. *Computers & Industrial Engineering*, **47**, pp. 17–29, 2004.
- [6] Ubeyli, E. D., Wavelet/mixture of experts network structure for EEG signals classification. *Expert Systems with Applications*, **34**, pp. 1954–1962, 2008.
- [7] Subasi, A., EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, **32**, pp. 1084-1093, 2007.
- [8] Auger, F., Flandrin, P., Goncalves, P. & Lemoine, O., *Time-Frequency Toolbox For Use with MATLAB*, CNRS, Rice University, 1995-1996.
- [9] Hanbay, D., Turkoglu, I. & Demir Y., An expert system based on wavelet decomposition and neural network for modeling Chua's circuit. *Expert Systems with Applications*, **34**, 2278–2283, 2008.
- [10] Kohonen, T., New Developments of Learning Vector Quantization and the Self-organizing Map. *Proc. of the Symposium on Neural Networks: Alliances and Perspectives in Senri 1992, (SYNAPSE '92)*, Osaka, 1992.
- [11] Misiti, M., Misiti, Y., Oppenheim, G. & Poggi, J. M., *MATLAB wavelet toolbox. User's Guide*, Version 2.

