Anomaly detection in time series data using a combination of wavelets, neural networks and Hilbert transform

Kanarachos, S. , Mathew, J. , Chroneos, A. and Fitzpatrick, M.E.

Anomaly detection in time series data using a combination of wavelets, neural networks and Hilbert transform

Kanarachos, S.; Mathew, J.; Chroneos, A.; Fitzpatrick, M. E.
Faculty of Engineering and Computing
Coventry University
Coventry, United Kingdom
stratis.kanarachos@coventry.ac.uk

***Abstract*—Real time detection of anomalies is crucial in structural health monitoring applications as it is used for early detection of structural damage and to identify abnormal operating conditions that can shorten the life of operating structures. A new signal processing algorithm for detecting anomalies in time series data is proposed in this study. The algorithm is expressed as a combination of wavelet analysis, neural networks and Hilbert transform in a sequential manner. The algorithm has been evaluated for a number of benchmark tests, commonly used in the literature, and has been found to perform robustly.**

Keywords—anomaly detection; wavelets; neural networks; Hilbert;

I. Introduction

The normal behavior of machines and structures is described by data that follow regular time driven patterns. Conversely, abnormal behavior can lead to deviations from the regular time pattern. Anomaly detection is the set of processes and methods put in place to automatically recognize abnormal patterns. Anomaly detection is relevant to a vast number of domains and with the explosion of sensor utilization has attained paramount importance. For example, in structural health monitoring scenarios, it is used to detect structural damage or in condition monitoring to detect overloads or under loads [1]. Data are collected in the form of sequences or time-series. For example, sequences of observations – accelerations, displacements, strains, etc. – are recorded during operation. A fault results in anomalous readings in sequences collected from one or more of the sensors. Anomaly detection has been investigated by several research communities to address ongoing issues in different application domains [2].

Surprisingly, till date there has not been any formal definition of what anomaly in a time series is. Different terms, like novelty [3], anomaly [4], surprise [5], deviant [6], change and point [7] have been used to describe similar notions. In general, an anomaly can be defined as an outlier; a point that stands out from a series of data points (point anomaly). In another definition an anomaly might be the change in behavior of a sequence of data points (pattern anomaly). In a greater context, anomaly can be described as a change in the response of a set of patterns (series anomaly). In the present study, we are interested in discovering pattern anomalies and not point or series anomalies. A number of methods have been proposed up to now for pattern anomaly detection.

Generally used methods to detect anomalies include statistical analysis ([2],[8]). An example is to employ symbolic time series analysis (STSA) [9] of noise-contaminated responses for feature extraction to detect and localize a gradually evolving deterioration in a structure due to the changes in the statistical behavior of symbol sequences. Specifically in STSA, statistical features of the symbol sequence can be applied to describe the dynamic status of the system under investigation. Symbolic dynamics and the set of statistical measures constitute a solid framework addressing the main challenges of the analysis of non-stationary time data. STSA may allow the capture of the main features of the system under investigation, thereby alleviating the effects of harmful noise. The simulation results under a range of damage conditions confirm the efficacy of the proposed technique for localization of gradually evolving deterioration in the structure.

For anomaly detection using support vector machines (SVMs) ([2],[10]), the semi-supervised variant known as the one-class SVM is predominantly used. In this, only normal data is used for training before anomalies can be detected. Theoretically, the one-class SVM may also be applied in an unsupervised anomaly detection setup, where there is no prior training. On the contrary, one-class SVM can be highly sensitive to outliers in the data. Debruyne [11] developed two modified versions to make one-class SVMs more suitable for unsupervised anomaly detection namely; Robust one-class SVMs and eta one-class SVMs. In both modified versions presence of outliers has least influence on the decision boundary compared to normal instances. Experiments performed on a number of datasets showed that the modified versions are promising. In particular, comparing with other standard unsupervised anomaly detection algorithms, the enhanced one-class SVMs are better in two out of four cases. Overall, the proposed eta one class SVM has shown the most consistent results.

In previous studies hybrid methods were developed to combine the strengths of individual algorithms. For example

Georgoulas *et al.* [11] presented an integrated anomaly detection approach for seeded bearing faults. The approach combined the Empirical Mode Decomposition and the Hilbert Huang transform to extract a compact feature set. Thereafter, a hybrid ensemble detector was trained using data derived only from the normal bearings and successfully employed to detect any deviation from the normal condition.

In this present study, we propose a novel hybrid anomaly detection algorithm based on a combination of wavelet analysis, neural networks and Hilbert transform. The wavelets are employed to denoise the original signal. A nonlinear autoregressive neural network is trained to emulate the output signal under normal operating conditions. The error signal, the difference between the neural network's output and the denoised signal, is then analyzed using the Hilbert transform and the output of the analysis is used to identify anomalous patterns.

The paper is structured as follows. The hybrid anomaly detection algorithm is presented in Section 2. The performance of the algorithm for benchmark tests found in literature is presented and discussed in Section 3. Conclusions and future research perspectives are given in Section 4.

II. ANOMALY DETECTION METHODOLOGY

The proposed algorithm combines wavelet analysis, nonlinear autoregressive neural networks and Hilbert transform in a sequential manner. A schematic of the algorithm consisting of various steps is shown in Fig. 1. Initially, the noise embedded in the signal is filtered using the the wavelet decomposition process. The output of the filtered signal is then predicted for normal operating conditions by training using a nonlinear autoregressive neural network. Hilbert transform is used consequently to analyze the error i.e. difference between the neural network output and the filtered signal. It is then possible to analyze the instantaneous frequency and amplitude following the data extraction. The pattern is classified as anomalous if the amplitude and/or instantaneous frequency vary significantly. A detailed description relevant to the application of the algorithm is given below.

A. Wavelet Analysis

Wavelets are a class of functions used to decompose signals into multiple components and can be reconstructed into the original signal without losing any information present. Using the wavelet transform a signal can be decomposed into different scales with different levels of resolution via the dilation of a single prototype function (the basis wavelet).

The major advantage of analysing a signal with wavelets is that it enables local features of the signal to be investigated with the level of detail matching their characteristic scale. This attribute of the wavelet transform enable to perform a multi-resolution analysis for a given signal. By using an inverse transformation it is possible to separate the original signal from the noise which is also termed as denoising. In the present study the basis function was selected from the Daubechies wavelet family (refer to Fig. 2) with the signal decomposed up to the 8^{th} level.

Fig. 1. Schematic illustration of the proposed algorithm

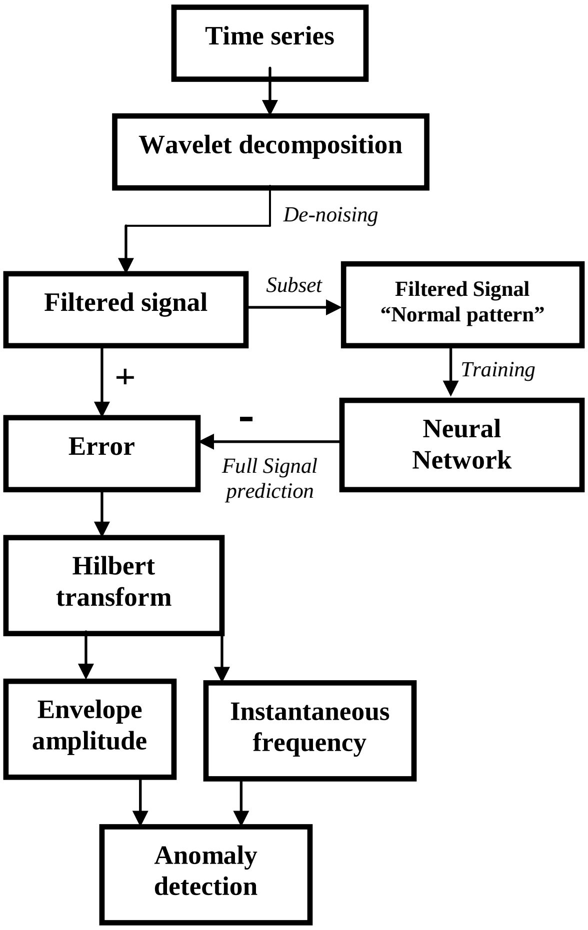
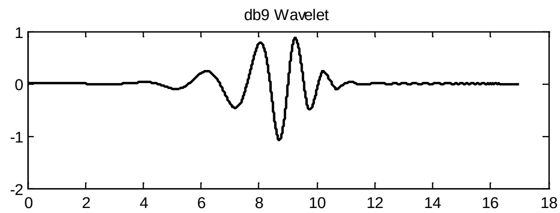


Fig. 2. Daubechies 9 wavelet basis



B. Nonlinear Autoregressive Neural Networks

Neural networks are increasingly being used to classify normal and abnormal signals. There are two ways to model temporal data using neural networks. The first way provides recurrent connections from output nodes to the preceding layer, whereas the second way is to provide buffers on the output of the nodes [13]. In this paper, an autoregressive neural network is implemented for predicting the normal pattern of the filtered signal. The neural network was trained using only a small subset of the filtered signal with careful discrimination, so that no anomalies were included in the training set.

A schematic of the neural network architecture is shown in Fig. 3. The network consisted of 10 neurons in the hidden layer with log-sigmoid activation function. A linear transfer function was used in the output layer. A three layer buffer was used for accumulating past knowledge and predicting the future time series data. The neural network was trained using the Levenberg-Marquardt backpropagation algorithm. In general, Levenberg–Marquardt algorithm (LMA) is used to solve non-linear least squares problems and usually converges within 20 iterations. The LMA interpolates between the Gauss–Newton algorithm (GNA) and the method of gradient descent. The LMA is more robust than the GNA, as it can find an optimal solution even if it starts far away from the final minimum. However, the LMA cannot distinguish between a local minima and a global minima similar to most of the fitting algorithms. The neural network regression statistics for one of the cases studied is shown in Fig. 4.

Fig. 3. Architecture of the Nonlinear autoregressive neural network

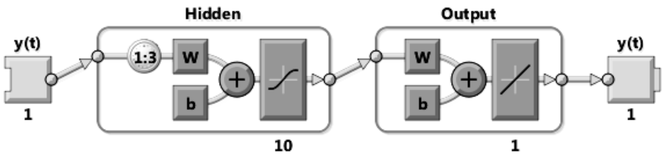
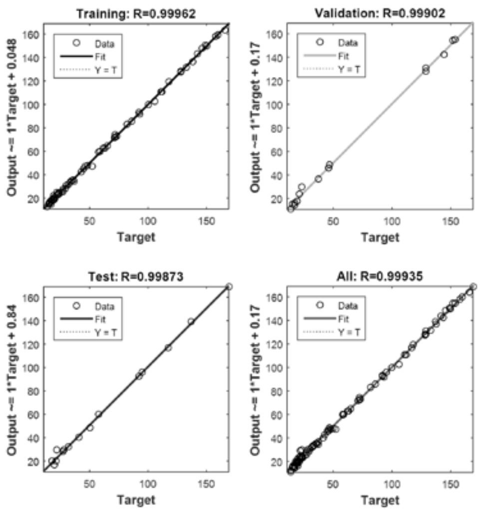


Fig. 4. Regression statistics of the neural network training

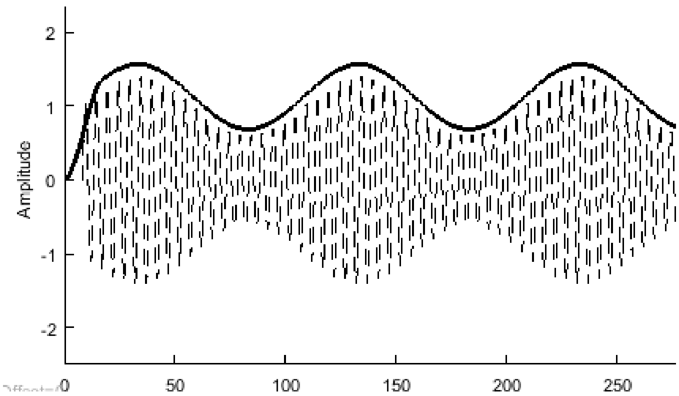


C. Hilbert transform

Hilbert transform (HT) is often used in the field of signal processing to derive the analytical representation of a signal. Using HT, it is plausible to detect the presence of a defect and could be very useful in structural health monitoring and damage detection. The signal's envelope can be considered to be equivalent to its outline and an envelope detector in effect connects all the peaks in this signal. Envelope detection is applicable in signal processing, particularly in amplitude modulation (AM) detection.

Hilbert transform is one of the most popular methods for extracting the envelope of a signal. HT is used to calculate the instantaneous frequency of a signal but restricted to monocomponent signals which are described in the time-frequency plane by a single "ridge." The set of monocomponent signals can include both single sinusoidal signals and signals like chirps. The outcome of an envelope detector is represented in Fig. 5 (shown as thick solid line). In the present study, HT is applied to the error signal rather than the original signal. This is because the error signal defined as the difference between the filtered signal and the output of the autoregressive neural network is characteristically a monocomponent signal. Hence, the issues pertaining to the decomposition and analysis of multicomponent signals are strategically avoided. It also assists in the automatic detection of the anomalies by monitoring the variations in the amplitude or frequency of the error signal.

Fig. 5. Envelope detector using Hilbert transform (Thick solid line represents Signal's envelope)



III. BENCHMARK RESULTS AND ANALYSIS

The hybrid algorithm presented in this study is tested based on two major existing anomaly detection methods. The results analysed using the published experimental data ([15] and [16]) are treated as benchmark conditions to evaluate the performance of the algorithm and are discussed below.

A. Experiment 1: Ma dataset

Chan *et al.* [15] used simulation datasets to test the anomaly detection algorithm. The dataset consists of two time series generated from the stochastic process described as:

$$X_1(t) = sin\left(\frac{40 \cdot \pi}{N} \cdot t\right) + n(t) \quad (1)$$

$$X_2(t) = sin\left(\frac{40 \cdot \pi}{N} \cdot t\right) + n(t) + e_1(t) \tag{2}$$

where t = 1, 2, . . . , N, N = 1200, $n(t)$ is an additive Gaussian noise with zero-mean and a standard deviation of 0.1 and $e_1(t)$ is a novel event, and expressed as:

$$e_1(t) = \begin{cases} e_1(t), & t \in [600,620] \\ 0. & otherwise \end{cases} \tag{3}$$

$X_1(t)$ is the normal time series with 1200 points, $X_2(t)$ is a signal added in the interval [600,620] representing an abnormal event $e_1(t)$ as illustrated in Fig. 6. In Fig. 7 the filtered signal $X_{2d}(t)$ following the wavelet transformation is shown.

Fig. 6. Illustration of anomalous signal X_2(t) of the Ma Data set [15]

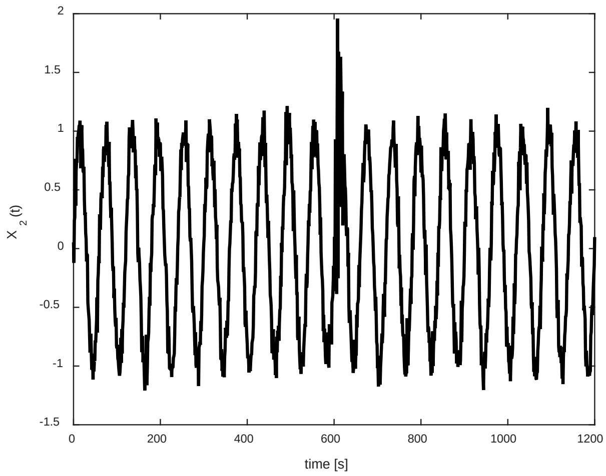
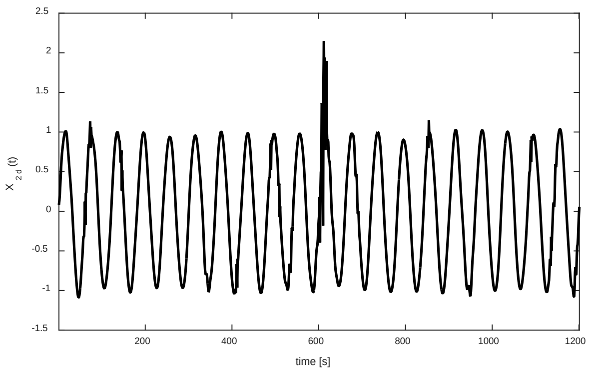


Fig. 7. Illustration of the filtered signal X_{2d}(t) following the wavelet decomposition



In Fig. 8 the error *e(t)* is illustrated; *e(t)* is defined by, *e(t)* = $X_{2d}(t)$- $X_{2nn}(t)$, where $X_{2nn}(t)$ is the output of the autoregressive neural network. It is evident that an anomalous behavior has taken place in the period $t \in [600,620]$ because of the different pattern observed in the envelope's amplitude as shown in Fig 9.

Fig. 8. Illustration of the error *e(t)*

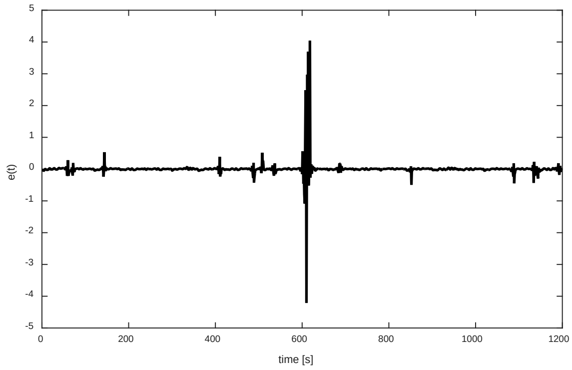
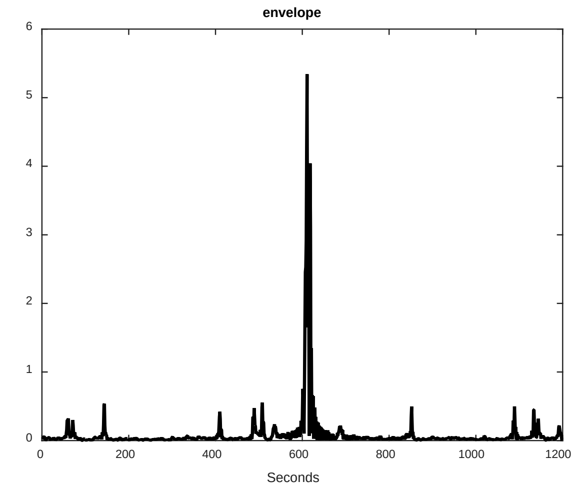


Fig. 9. Illustration of *e(t)* signal's envelope following the Hilbert transform



B. Experiment 2: Keogh dataset

The Keogh dataset is a simulation dataset used to test three anomaly detection algorithms IMM, TSA-Tree and Tarzan by Keogh *et al.* [16]. The dataset is generated by the following expression:

$$Y_1(t) = sin\left(\frac{50 \cdot \pi}{N} \cdot t\right) + n(t) \tag{4}$$

$$Y_2(t) = sin\left(\frac{50 \cdot \pi}{N} \cdot t\right) + n(t) + e_1(t) \qquad (5)$$

where t = 1, 2, . . . , N, N = 800. $n(t)$ is an additive Gaussian noise with zero-mean and a standard deviation of 0.1 and $e_1(t)$ is a synthetic "anomaly", defined as follows:

$$e_1(t) = \begin{cases} sin\left(\frac{75 \cdot \pi}{N} \cdot t\right) - sin\left(\frac{50 \cdot \pi}{N} \cdot t\right) & t \in [400,432] \\ 0. & otherwise \end{cases} \qquad (6)$$

$Y_1(t)$ is a sine wave with a Gaussian noise as given in the above description. $Y_2(t)$ is obtained by adding an anomalous event e_1 (t) in the time series $Y_1(t)$ to the sine wave cycle in the interval [400,432]. Fig. 10 shows the signal $Y_2(t)$ and the output of the wavelet decomposition as represented in Fig. 11. Fig. 12 shows the output of the neural network and the amplitude of the signal's envelope following the Hilbert transformation as illustrated in Fig. 13. It is evident from the results that the deviation in the signal e(t) is only observed in the time interval $t \in [400,432]$.

Fig. 10. Illustration of the anomalous signal $Y_2(t)$[16]

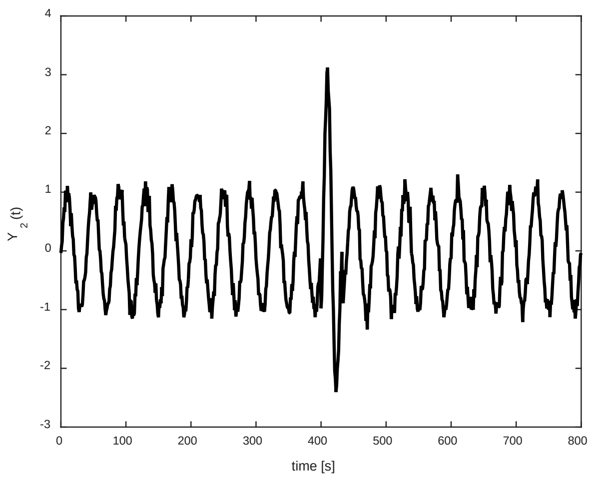


Fig. 11. Illustration of the anomalous signal $Y_2(t)$[16]

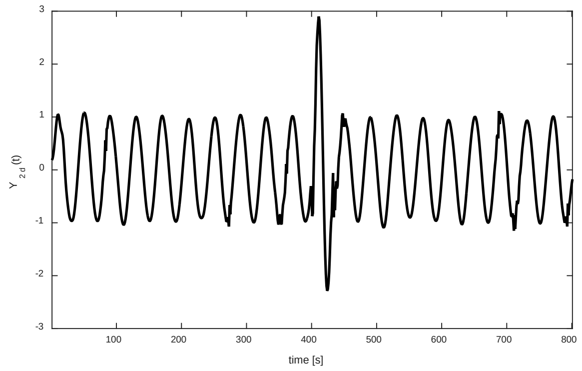


Fig. 12. Illustration of error $e(t)$

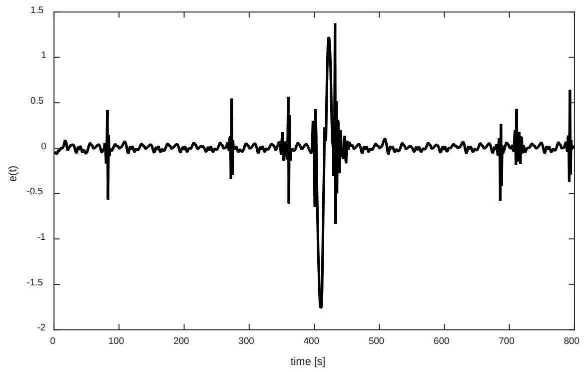
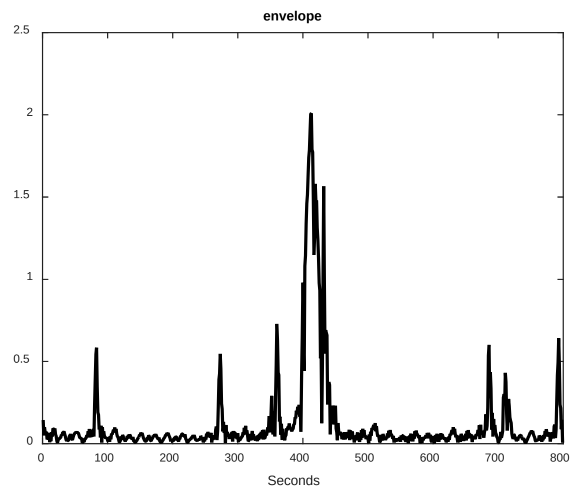


Fig. 13. Illustration of $e(t)$ signal's envelope following the Hilbert transform



IV. CONCLUSIONS

To summarize, a new signal processing algorithm for detecting anomalies in time series data is presented. The algorithm employs a combination of wavelet analysis, neural networks and Hilbert transform in a sequential manner to process the raw signal. The efficacy of this algorithm has been tested for a number of benchmark conditions and shown to perform robustly. This approach can potentially be applied to detect anomalies especially in structural health monitoring applications and for early detection of structural damage scenarios that have a negative impact upon a structure's lifetime.

ACKNOWLEDGEMENTS

The authors are grateful for funding from the Lloyd's Register Foundation, a charitable foundation helping to protect life and property by supporting engineering-related education, public engagement and the application of research.

REFERENCES

[1] S. Kanarachos and G. Demosthenous, "Investigation of a beam crack identification method using the standard finite element formulation," Proceedings of the 5th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing, 2006, pp. 209-214.

[2] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," ACM Computing Surveys, pp. 1–72, 2009.

[3] X.Y. Chen and Y.Y. Zhan, "Multi-scale anomaly detection algorithm based on infrequent pattern of time series," J. Comput. Appl. Math., vol. 214 (1), pp. 227-237, 2008.

[4] D. Dasgupta and S. Forrest, "Novelty detection in time series data using ideas from immunology," in: Proceedings of the Fifth International Conference on Intelligent Systems, Reno, Nevada, June 19–21, 1996.

[5] E. Keogh, S. Lonardi, and W. Chiu, "Finding surprising patterns in a time series database in linear time and space," in: Proceedings of the Eigth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2002, pp. 550–556.

[6] C. Shahabi, X. Tian, and W. Zhao, "TSA-Tree: a wavelet-based approach to improve the efficiency of multi-level surprise and trend queries on time-series data," in: Proceedings of the 12th International Conference on Scientific and Statistical Database Management (SSDBM'00), IEEE Computer Society, Berlin, Germany, pp. 55–68, 2000.

[7] K. Yamanishi and J.-I. Takeuchi, "A unifying framework for detecting outliers and change points from non-stationary time series data," in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, New York, 2002, pp. 676–681.

[8] W. Gu, J. Choi, M. Gu, H. Simon, and K. Wu, "Fast Change Point Detection for electricity market analysis," Proceedings - 2013 IEEE International Conference on Big Data, Big Data, 2013, pp. 50-57.

[9] M.M. Alamdari, B. Samali, and J. Li, "Damage localization based on symbolic time series analysis," Structural Control and Health Monitoring, vol. 22 (2), pp. 374-393, 2015.

[10] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class Support Vector Machines for unsupervised anomaly detection," Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description, ODD 2013, pp. 8-15, 2013.

[11] M. Debruyne, "An outlier map for support vector machine classification," The Annals of Applied Statistics, vol. 3, pp. 1566–1580, 2009.

[12] G. Georgoulas, T. Loutas, C.D. Stylios, and V. Kostopoulos, "Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition," Mechanical Systems and Signal Processing, vol. 41 (1-2), pp. 510-525, 2013.

[13] S. Haykin, "Neural Networks: A Comprehensive Foundation," Prentice Hall, 2 edition. 1999.

[14] G.P. Zhang and M. Qi, "Neural network forecasting for seasonal and trend time series." Eur. J. Operational Res., vol. 160(2), pp. 501–514, 2005.

[15] K.-P. Chan, W.-C. Fu, and C. Yu, "Haar wavelets for efficient similarity search of time-series: with and without time warping," IEEE Trans. Knowledge Data Eng. vol. 15 (3), pp. 686–705, 2003.

[16] E. Keogh, S. Lonardi, and W. Chiu, "Finding surprising patterns in a time series database in linear time and space," in: Proceedings of the Eigth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, NewYork, 2002, pp. 550–556.