

ANOTHER ADAPTIVE APPROACH TO NOVELTY DETECTION IN TIME SERIES

Matouš Cejnek, Peter Mark Beneš, Ivo Bukovsky

Department of Instrumentation and Control Engineering,
Czech Technical University, Prague, Czech Republic
{Matous.Cejnek;PeterMark.Benes;Ivo.Bukovsky}@fs.cvut.cz

ABSTRACT

This paper introduces a novel approach to novelty detection of every individual sample of data in a time series. The novelty detection is based on the knowledge learned by neural networks and the consistency of data with contemporary governing law. In particular, the relationship of prediction error with the adaptive weight increments by gradient decent is shown, as the modification of the recently introduced adaptive approach of novelty detection. Static and dynamic neural network models are shown on theoretical data as well as on a real ECG signal.

KEYWORDS

Novelty Detection, Time Series, Gradient Descent, Neural Networks, ECG

1. INTRODUCTION

In principle, novelty carried through measured samples of data may be evaluated either via probability based approaches as exemplified in [1] or, via learning system based approaches as in the work [2]. The first of these streams, i.e. probabilistic, is represented by the statistical approaches of novelty measures and by probabilistic approaches for evaluation of entropy. The Sample Entropy (SampEn) and the Approximate Entropy (ApEn) are very typical and very relevant examples to be mentioned [3]-[4]. These approaches are closely related to the multi-scale evaluation of fractal measures, where further case studies utilizing SampEn, ApEn, and Multiscale Entropy (MSE) can be found in [5]-[7]. Further to this, probabilistic entropy approach to the concept shift (sometimes the concept drift) detection in sensory data is reported in [8]. The second of the mentioned streams is represented by the utilization of learning systems, such as neural networks and fuzzy-neural systems, and this is also the main area of focus for the presented work in this paper. During the last three decades of 20th century, the works that were focused in regards to learning systems are that of [9]-[12], and for incremental learning approach can be referenced for example also the work [13]. Then, a particularly focused approach toward the utilization of learning systems has been rising with works [14]-[17]. Where, nonlinear estimators and learning algorithm were utilized for the fault detection via the proposed utilization of a fault function that evaluates behaviour of residuals of a learning system. Currently, significant research that shall also be referenced is adaptive concept drift detectors, proposed in [18]. Some readers might also see some analogies of the proposed approach in this paper to the Adaptive Resonance Theory [19]. Another approach to novelty detection is based on utilization of adaptive parameters of incrementally learning models (neural networks), i.e. the Adaptation Plot

[20] that has been recently enhanced with multi-scale approach [21]. A most recent method is the Learning Entropy, i.e., a multi-scale approach to evaluation of unusual behaviour of adaptive parameters of a learning model is introduced in [22]. This paper however introduces another, different approach to novelty detection, that is neither based on statistical approaches, nor is it based on evaluation of error residual. Contrary to [21], this new approach operates only on parameter space of incrementally learning systems, and it does not use the multi-scale approach for detection sensitivity. Here the method introduced shows one of the possibilities how to detect perturbations within measured data in every new sample.

2. USED METHODOLOGY

The method of detection proposed in this paper, utilises an adaptive prediction model. It is interesting to note how promising such method of detection works when used with a static linear adaptive model on non-linear data. Thus, for demonstrational purposes of this method, we will use a linear based adaptive model. The adaptation technique as applied to the predictive model, is based on the classical Gradient Descent algorithm. The employed linear adaptive model initially features a vector of random numbers as initial neural adaptive weights, which over each sample, are incrementally trained to model the real data, and an input vector comprised from previous samples of the real data, as follows;

$$y(k+1) = \sum_{i=0}^n w_i x_i = \mathbf{w} \cdot \mathbf{x} \quad (1)$$

Where, here, \mathbf{w} denotes the row vector of all neural weights and \mathbf{x} represents the vector of inputs. For the scope of this paper, the structure of this input vector for the linear predictive model is as follows;

$$\mathbf{x}(k) = [1 \quad y_r(k) \quad y_r(k-1) \quad \dots \quad y_r(k-10)]^T \quad (2)$$

Where, $x_0=1$ allows for neural bias in case of the linear predictive model and y_r represents measured values of the time series. The input vector was chosen to feature ten previous samples of the measured time series. Prior to prediction, the input vector was standardized via equation (3), in order to acquire better stability of the adaptive algorithm.

$$y_r = \frac{y_r - \text{mean}(y_r)}{3 \times \text{std}(y_r)} \quad (3)$$

For further improvement of the model stability, we provide adaptation of the learning rate μ via Normalized Least Squares method [23]. The equation (4) describes how this adaptation is performed. The Learning Rate adaptation is utilized before calculation of every sample. From equation (4) it is obvious that the normalized learning rate is smaller than the default learning rate in dependency on input $\mathbf{x}(k)$ for every sample;

$$\eta = \frac{\mu}{1 + \mathbf{x}(k)^T \times \mathbf{x}(k)} \quad (4)$$

Where, η serves as a substitution for μ For novelty detection estimation in every new sample in discrete time k , we use the product of absolute values of the prediction error and adaptive weight changes according to equation (5), as follows;

$$ND(k) = [|e(k) \times \Delta w_i(k)|; i = 0, \dots, 10] \quad (5)$$

Equation (5), thus presents the main principle behind this introduced method of novelty detection of every new sample according to temporary system dynamics. This method is simultaneously using prediction error and changes of adaptive weights ($e(k)$ and $\Delta w_i(k)$, respectively). This is the main theoretical result of this work. In the following sections we will investigate functionality of this method on real and artificial ECG with sampling of 256Hz frequency.

3. EXPERIMENTAL ANALYSIS

In this section, we will demonstrate the newly introduced method via computer simulations created in programming language Python 2.7 [24]. The numerical algorithms were realized via python library Numpy [25]. The simulation was performed on a personal computer and was faster than real time measuring of an ECG signal. Thus, this justifies that such implementation is applicable for online realization.

3.1. Artificial ECG

In this subsection, we will demonstrate usage of the newly introduced method on an artificial ECG signal both, with and without noise, before eventually testing on a real ECG signal (of 256 Hz frequency), in the following subsection. For demonstrational purposes in this section, noise was added to the signal to highlight the ability of this introduced method, in detection of unexpected samples within the data. Especially where, the predictive model features lower prediction accuracy, comparative to data without noise. The reason why we test this method on an artificial signal, is to emphasize how well the detection works on perturbed data, if the signal doesn't contain any complicated phenomena. The artificial ECG time series used in our paper was created by a serially repeated pattern of a real ECG signal, sampled over one period. Thus, this artificial time series is an ideally periodic signal.

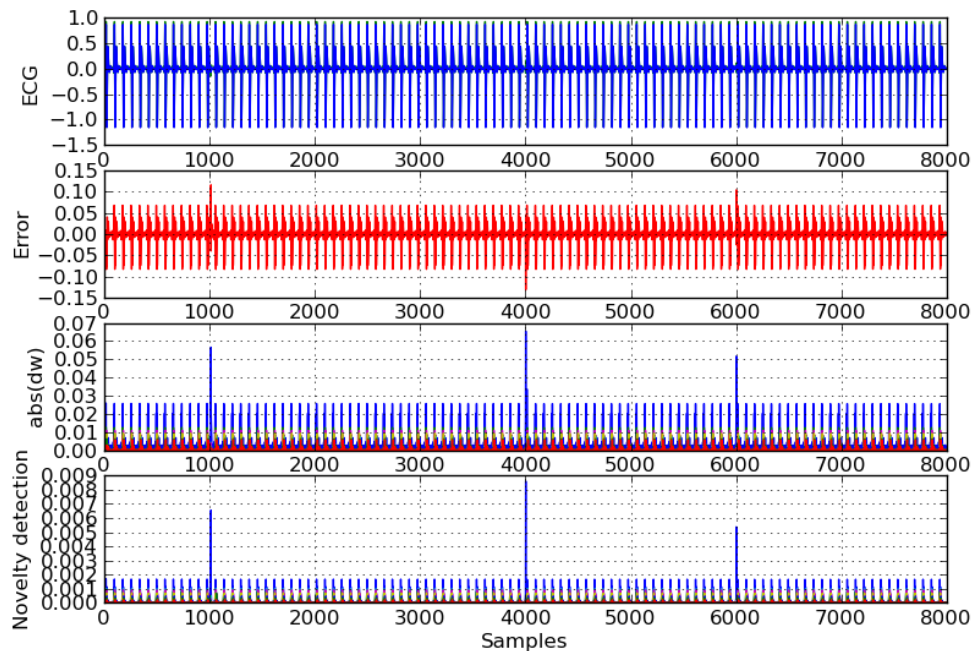


Figure 1. Novelty Detection used on artificial ECG without noise (for details refer to Figure 2)

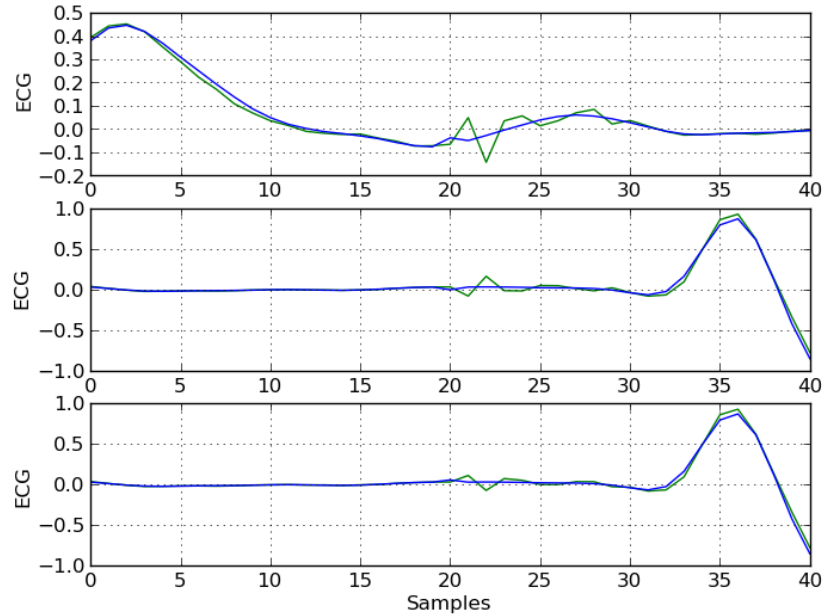


Figure 2. Details of prediction in areas of introduced perturbations in artificial ECG without noise (blue – signal with perturbation, green – predicted signal)

Regarding the length of the pre-training data for the predictive model, we chose to take the first 2000 samples of the artificial data (ref. Figure 1). For sufficient training, 100 epochs was adequate. Figure 2, depicts the details of all perturbations included in the artificial ECG signal without noise. The size of the introduced perturbation is 0.03 mV. As we can see on Figure 2, these perturbations are small in comparison with the amplitude of the signal. Looking at the behaviour of the used predictive model, here we see the model tries to relearn immediately when the prediction error and weight adaptation increases (Figure 1). The return of the predictive model to previous prediction accuracy takes approximately 20 following samples. In Figure 1, it is possible to see the prediction error in specific places of a single period. These errors are caused by insufficient prediction ability of the simple, linear predictive model. Furthermore, in Figure 1, these errors are not detected as new data by the applied novelty detection method. Figure 3, shows the simulation of the artificially created ECG with the addition of noise. Here again, three perturbations were introduced to the data. These perturbations are located on the same positions as the signal without noise. This introduced noise was implemented via a generator of pseudo-random numbers, composed as a vector of random numbers in range from 0 to 0.01, added via the following equation (6);

$$y_r(k) = y_r(k) + rand, rand \in \langle 0, 0.01 \rangle, k = 1000, 4000, 6000. \quad (6)$$

Figure 3 displays the real signal together with the simulated. Here in the first graph, it is possible to see the difference between the signals within the region of peaks of the amplitude. For the used predictive model, it is much more difficult to learn the pattern of the signal with added noise in comparison to that without noise. The errors and absolute values of the weight increments (Figure 3) are not entirely dependent on the periodicity of the signal. In the plot of error on Figure 3, it is not possible to see the perturbations clearly, as in the plot of prediction error without noise on Figure 2. In the graph of absolute error of weight increments, we can more evidently see the location of the perturbations. On the graph of novelty detection (Figure 3) the perturbation locations are even more evidently seen and, this is because a huge part of the models periodic

errors, are filtered. Furthermore, these errors have no impact on the detection of unexpected samples in the data.

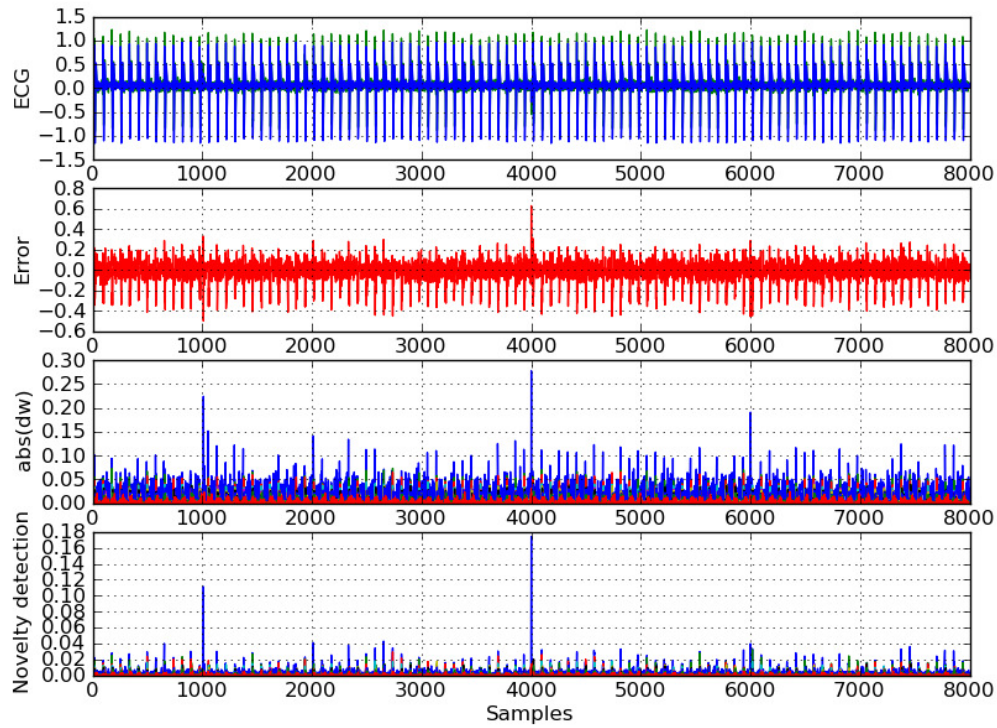


Figure 3. Novelty Detection used on artificial ECG with noise (details on Figure 4)

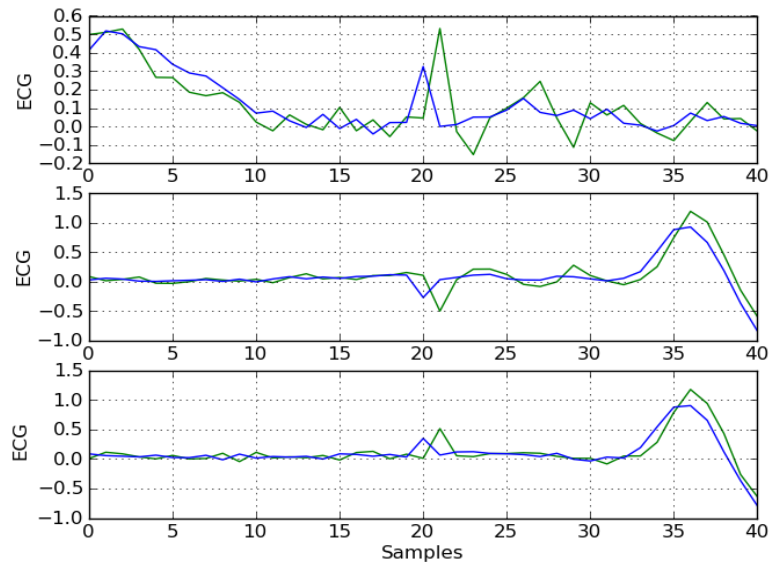


Figure 4. Details of prediction in areas of introduced perturbations in artificial ECG with noise (blue – signal, green – predicted signal)

Figure 4 shows the details of three perturbations included into the artificial signal containing noise. Here, it is clearly shown that the prediction error of the signal with noise and also in the un-

perturbed section of the signal, could reach substantially high values. The adaptive model again immediately reacts to the introduced perturbation and tries to relearn the data signal. The time in which the model needs for regaining normal accuracy after meeting a perturbation is hard to estimate, because the prediction error is strongly dependant on the level of noise present in the data.

3.2. Real ECG Signal

In this paper we also demonstrate the introduced method on an ECG signal. The used time series was measured by an internal cardio-defibrillator with 256Hz frequency (the measured data was obtained courtesy of [26]). This signal was chosen because it contains spontaneous ventricular tachycardia, which is a rare phenomenon to measure. In this subsection, this introduced method of novelty detection will be demonstrated via a real ECG signal, which features a healthy, sinus like structure followed, by a ventricular tachycardia (arrhythmia). Later, we will show that this technique works in both in the sinus like region and also in the arrhythmic region of this signal. For demonstration, we will choose the same linear neural unit as the one previously used for prediction of the artificial ECG signal. The size of data chosen for learning of the predictive model is 1000 samples. In order to achieve sufficient pre-learning of the used neural unit with such amount of data, less than 500 epochs is adequate for achieving optimal results. Using any larger size of epochs doesn't seem to significantly improve the accuracy. Figure. 5 clearly depicts which part is the healthy ECG signal and which part represents the arrhythmia. In the introduced novelty detection, it is possible to detect the start of the arrhythmia signal, approximately 1000 samples before the arrhythmia is introduced (the shape of the period before the arrhythmia looks the same, but the scale of amplitudes starts varying). In the first graph of Figure 5, we can see how the measured signal is practically equivalent with the predicted signal. The included perturbations are not clearly seen in this graph. However, these perturbations are located in the samples of discrete time 1000, 3000, 5000. On the graph of the prediction error, it is possible to see the perturbations quite well and even more so in the graph of absolute values of adaptive weights. However, looking on the graph of novelty detection, these perturbations are even more evidently pronounced. Moreover, the periodic errors are suppressed in region of the healthy ECG and arrhythmia signal. It is important to notice that the suppressing of the periodic error, is not that dominant in the onset of arrhythmia, unlike in other parts of the data. On Figure 6, we can see the prediction models seem to be quite accurate, but in some parts of the ECG period, there is always some small inaccuracy. These inaccuracies are mostly located in the positions of global minima's and maxima's, for a given period on the measured signal. The included perturbations are significantly small in size (0.04 – approximately 2% of the amplitude of healthy ECG signal). Furthermore, the adaptive model immediately reacts with re-learning of the ECG data, where by achieving the previous model accuracy, takes approximately 5-10 following samples.

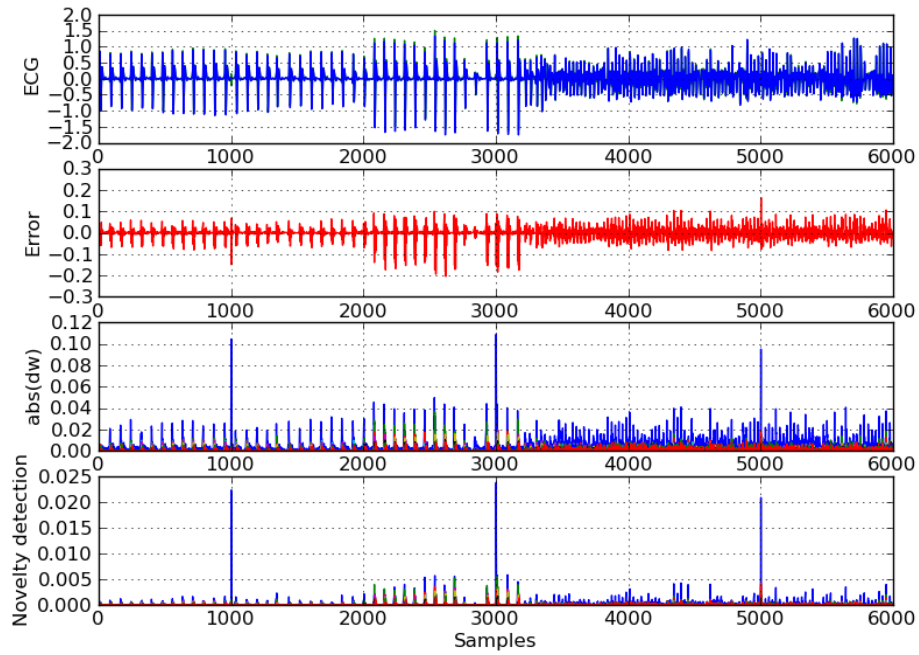


Figure 5. Novelty Detection used on real measured ECG signal

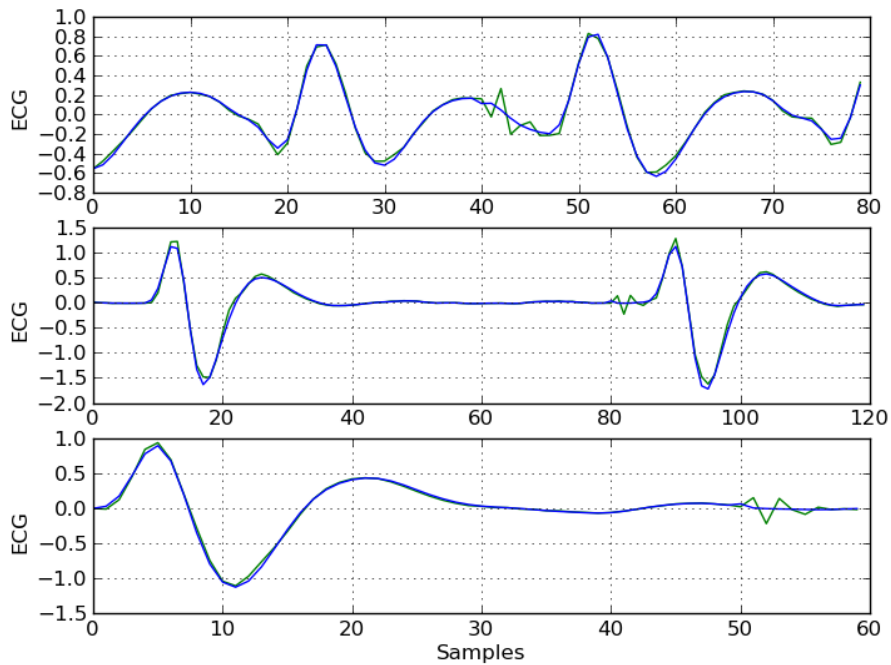


Figure 6. Details of prediction in areas of introduced perturbations in real measured ECG (blue – real signal, green – output of neural unit)

3. DISCUSSION

As presented in previous chapters, the potentials of this method for novelty detection in biomedical signals, is quite promising. For demonstration, we chose a predictive model, which was relatively simple by structure. Here, such model was not capable to properly predict the measured ECG, yet the introduced novelty detection was able to distinguish the difference between prediction error caused by the model insufficiency and prediction error caused by novelty in the samples. This method also seems to be highly useful on artificial signals containing significant noise, where the model is unable to learn such noise, but can distinguish the noise from new phenomena in the measured data. Also on artificial data without noise we can certainly note, that novelty detection distinguishes the prediction error from the novelty in data, also with even a significantly small, signal perturbation.

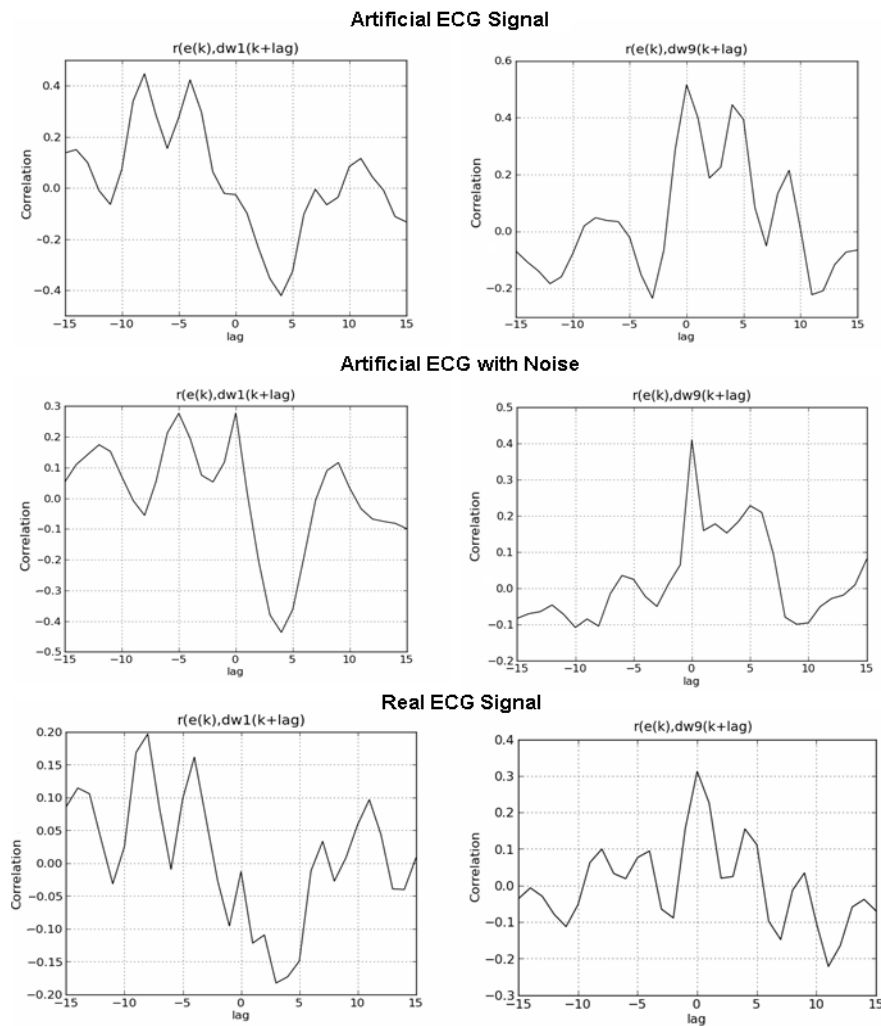


Figure 7. Correlation coefficients between adaptive weight increments and error, for Δw_1 and Δw_9 in the artificial ECG signal without noise (upper pair), artificial ECG signal with noise (middle pair) and real ECG signal (lower pair)

A further remark may be seen in Figure 7. Where, the correlation coefficients between the adaptive weight increments and error, over various samples of lag are analysed. For

demonstrative purposes, Figure 7 only includes an analysis over the Δw_1 and Δw_9 adaptive weight increments with the error. Here, we may note that in all cases of the corresponding correlation coefficients, statistically insignificant correlation between the adaptive weight increments and error is seen and the same was observed for all other weight increments. This is justified by the notion that the prediction error of the model carries the information about the size of inaccuracy, while the adaptive weight increments carry information about how much the model tried to adapt on new data. Thus, even when the model error may be high, this may not necessarily be related with the adaptive weight increments (for more on this principle, please refer to the works of [20]–[22]). In case of a high error value, caused by common phenomena (what is out of the models capability to learn), novelty detection doesn't mark those particular samples, because the model has already recognized its inability to learn, which also further justifies the functionality of this method for novelty detection.

Another remark is in regards to the small time consumption in detection of this newly introduced method. This is due to the nature of the model being merely multiplication of values, which was calculated during the prediction of the data. Novelty detection doesn't slow down the process of prediction. This fact is a big advantage especially in processing the signal in real time, what could be a crucial feature in the case of ECG or other biomedical signals. Furthermore, as we can see on the results obtained by application of novelty detection on real ECG (Fig. 5), this technique of detection of unexpected data highlights the onset of arrhythmia from the other data. According to this method it is possible to see the onset of arrhythmia before it appears in the measured signal or prediction error. The first few samples show that there is inconsistency in the data, which appears approximately 1000 samples before the start of the arrhythmia signal. This model could be used for the detection of arrhythmia after proper optimizing and, it is an objective for further research.

4. CONCLUSION

Unusual behaviour of adaptable parameters and actual prediction error together, establishes novel learning-system-based measure of novelty that is introduced in this paper. The method presented in this paper is different from other methods and from the methods in [20] & [21], because this method is using actual error of the predictive model, together with cognitive information in changes of adaptive weights of the model. This is an interesting approach, because the prediction accuracy and behaviour of a learning model are not necessarily correlated with each other [21]; however, both values together provide us with important information about consistency at every sample with temporary system dynamics. These facts are the reason why this and similar models of novelty detection are important for future research and development.

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues from *Tohoku University*, namely from the *Yoshizawa-Sugita Lab* (formerly *Yoshizawa-Homma Lab*) and from the *Department of Radiological Imaging and Informatics*, for their vital, continuous cooperation and support. The authors would also like to acknowledge the SGS12/177/OHK2/3T/12 grant, for support of this work.

REFERENCES

- [1] Markou, M.; Singh, S. Novelty detection: a review—part 1: statistical approaches, *Signal Processing*, Volume 83, Issue 12, 2003, pp. 2481-2497.
- [2] Markou, M.; Singh, S. Novelty detection: a review—part 2: neural network based approaches, *Signal Processing*, Volume 83, Issue 12, 2003, pp. 2499-2521.

- [3] Pincus, S.M. Approximate Entropy as a Measure of System Complexity. *Proc Natl Acad Sci, USA* 88: 1991, pp. 2297–2301.
- [4] Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circul. Physiol.*, 2000, 278, H2039–H2049.
- [5] Costa, M.; Goldberger, A.L.; Peng, C.K. Multi-scale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* 2002, 89, 68102.
- [6] Costa, M.; Goldberger, A.L.; Peng, C.K. Multi-scale entropy analysis of biological signals. *Phys. Rev. E Stat. Nonlin Soft Matter Phys.* 2005, 71, 021906.
- [7] Yin, L.; Zhou, L. Function Based Fault Detection for Uncertain Multivariate Nonlinear Non-Gaussian Stochastic Systems Using Entropy Optimization Principle, *Entropy* 2013, 15, pp.32-52.
- [8] Vorburger, P.; Bernstein, A.: Entropy-based Concept Shift Detection.,*Proc. 6th Int. Conf. Data Min.*, 2006, pp.1113–1118.
- [9] Willsky, A. A survey of design methods for failure detection in dynamic systems. *Automatica*, 1976, vol. 12, pp.601–611.
- [10] Gertler, J. Survey of model-based failure detection and isolation in complex plants. *IEEE Contr. Syst. Mag.* 1988, vol. 8, pp.3–11.
- [11] Isermann, R. Process fault detection based on modeling and estimation methods: A survey. *Automatica*, 1984, vol. 20, pp. 387–404.
- [12] Frank, P.M. Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—A survey and some new results. *Automatica*, 1990 vol. 26, pp.459–474.
- [13] Widmer, G.; Kubat, M. Learning in the Presence of Concept Drift and Hidden Contexts, *Machine Learning*, April M. M. 1996, Volume 23, Issue 1, pp.69-101.
- [14] Polycarpou, M.M.; Trunov, A.B. Learning Approach to Nonlinear Fault Diagnosis: Detectability Analysis. *IEEE Transactions on Automatic Control*, 2000, vol. 45, no. 4.
- [15] Polycarpou, M.M.; Helmicki, A.J. Automated fault detection and accommodation: A learning systems approach. *IEEE Trans. Syst., Man Cybern.*, 1995, vol. 25, pp.1447–1458.
- [16] Demetriou, M.A.; Polycarpou, M.M.: Incipient Fault Diagnosis of Dynamical Systems Using Online Approximators. *IEEE Transactions on Automatic Control*, 1998, Vol. 43, No. 11.
- [17] Trunov, A.B.; Polycarpou, M.M. Automated fault diagnosis in nonlinear multivariable systems using a learning methodology. *IEEE Transactions on Neural Networks*. 2000, vol. 11, no. 1, pp.91–101.
- [18] Alippi, C.; Roveri, M. Just-in-Time Adaptive Classifiers—Part I: Detecting Nonstationary Changes. *IEEE Transactions on Neural Networks*, 2008, Vol. 19, No. 7.
- [19] Grossberg, S.; Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world. *Neural Networks*, 2013, 37, pp.1–47.
- [20] Bukovsky, I.; Bila, J. Adaptive Evaluation of Complex Dynamic Systems using Low-Dimensional Neural Architectures. *Advances in Cognitive Informatics and Cognitive Computing, Series: Studies in Computational Intelligence*, Vol. 323/2010, eds. D. Zhang, Y. Wang, W. Kinsner, Springer-Verlag Berlin Heidelberg, 2010, pp.33-57.
- [21] Bukovsky, I.; Kinsner, W.; Bila, J. Multi-scale Analysis Approach for Novelty Detection in Adaptation Plot. 3rd Sensor Signal Processing for Defence 2012 (SSPD 2012), Imperial College London, UK.
- [22] Bukovsky, I.: “Learning Entropy: Multiscale Measure for Incremental Learning”, manuscript ID entropy-39174, *Entropy*, special issue on Dynamical Systems, ISSN 1099–4300, (accepted paper 22/09/2013).
- [23] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [24] van Rossum, G., Drake, F. L. (eds): *Python Reference Manual*, PythonLabs, Virginia, USA, 2001. Available at <http://www.python.org>
- [25] Ascher, D., Dubios, P., F., Hinsen, K., Hugunin, J., Oliphant, T.: *Numerical Python*, Lawrence Livermore National Laboratory, Livermore, California, USA, 2001.
- [26] Yoshizawa-Sugita Lab and Homma Lab, Tohoku University, Japan.

Short Biography of Authors

Matouš Cejnek received his Bachelor's Degree at Czech Technical University in Prague in 2012. Currently he is a Master's student, with expected PhD studies to follow in 2014 at CTU. His research focuses on non-conventional neural networks for adaptive systems and novelty detection in time series and biomedical applications. Matous's work has been awarded in local and international student competitions in 2013.

Peter Mark Beneš received his Bachelor's degree with honours at Czech Technical University in Prague (CTU) in 2012. Currently, he is a Master's student, with expected PhD studies to follow at CTU in 2014. His research focuses on non-conventional neural networks for adaptive identification and control of industrial systems including hoist mechanisms and skew control of rail-based mechanisms. Peter's work has been awarded in local and international student competitions also with an industrial BOSCH award in 2013.

Ivo Bukovsky graduated from Czech Technical University in Prague where he received his Ph.D. in the field of Control and System Engineering in 2007 and became associate professor since 2013. His research interests include higher-order neural networks, adaptive evaluation of time series and systems, multi-scale-analysis approaches, control and biomedical applications. He was a visiting researcher at the University of Saskatchewan (2003), at the University of Manitoba in Canada (2010), and he was a visiting professor at Tohoku University in 2011.