

Adaptive Novelty Detection with Generalized Extreme Value Distribution

Jan Vrba

University of Chemistry and Technology, Prague
 Faculty of Chemical Engineering, Department of Computing and Control Engineering
 Czech Republic
 Email: jan.vrba@vscht.cz

Abstract—This paper introduces the new adaptive novelty detection method. The proposed method is using generalized extreme value distribution to evaluate the absolute value of adaptive system weight increments in time. The detection of novelty is threshold-based and the threshold ζ corresponds to the value of joint probability density function. Performance of the proposed algorithm is shown on artificial data. For comparison also results of Learning Entropy algorithm are shown, as this algorithm also evaluates the increments of adaptive weights.

keywords—signal processing, adaptive systems, adaptive algorithms, novelty detection, generalized extreme value distribution

I. INTRODUCTION

Novelty detection topic is still important nowadays, as the learning systems are getting more and more popular. Many various approaches were developed during decades of research [1], [2]. The two main approaches were established. Namely, it is the statistical approach and the neural network approach. However, some novelty detection methods belong to both groups, thus alternative classification can be also found [3]. One of the novelty detection methods that can be considered as a neural network approach, is the Learning Entropy algorithm [4]. This algorithm is evaluating the adaptive systems adaptive weight increments instead of the prediction error and estimates Learning Entropy of each sample. The Learning Entropy reflects the effort of the adaptive system to learn the novelty in the data. Authors in their later works extend original algorithm with prediction error evaluation [5] or with the different technique of Learning Entropy estimation [6] and shows, that evaluation of weight increments can be a useful approach to novelty detection topic [7], [8].

In this paper, the new algorithm for novelty detection is introduced. This algorithm is based on extreme value theory and evaluation of adaptive weights increments. Performance of Learning Entropy algorithm is shown to compare.

II. ADAPTIVE SYSTEM SPECIFICATION AND NLMS ALGORITHM

In the experiments, the simple linear adaptive model is used. The output of this model at a discrete time

index k is given as

$$y(k) = h_1(k) \cdot x_1(k) + h_2(k) \cdot x_2(k), \quad (1)$$

which is equivalent with form

$$y(k) = \mathbf{h}^T(k) \cdot \mathbf{x}(k), \quad (2)$$

where $\mathbf{h}^T(k) = [h_1(k), h_2(k)] \in R^2$ is the vector of adaptive weights, $\mathbf{x}^T(k) = [x_1(k), x_2(k)] \in R^2$ is the input vector. The vector of adaptive weights is updated with every new sample obtained. So the update can be written as

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \Delta\mathbf{h}(k), \quad (3)$$

where $\Delta\mathbf{h}(k)$ is the vector of weights increments.

The normalized-least-mean-squares (NLMS) algorithm was used for the experiments. The update of this adaptive algorithm is given as

$$\Delta\mathbf{h}(k) = \frac{\mu \cdot \mathbf{h}(k) \cdot e(k)}{a + \mathbf{x}(k) \cdot \mathbf{x}^T(k)}, \quad (4)$$

where $a \in R$ is small positive constant used to avoid division by zero, $\mu \in R$ is the learning rate and $e \in R$ is the output error defined as

$$e(k) = d(k) - y(k), \quad (5)$$

where $d(k) \in R$ is the measured output and k is the discrete time index.

III. NOVELTY DETECTION ALGORITHMS

A. Review of Learning Entropy

According to [4], [7], the Learning Entropy (or the Approximate Individual Sample Learning Entropy) is given as follows

$$E_A(k) = \frac{1}{n \cdot n_\alpha} \sum_{j=1}^n \sum_{i=1}^{n_\alpha} f_{\alpha_i} (|\Delta h_i(k)| > \alpha_i \cdot \overline{|\Delta h_i(k)|}), \quad (6)$$

where $f_{\alpha_i}(TRUE) = 1$, $f_{\alpha_i}(FALSE) = 0$ and n is the number of adaptive weights, n_α is the number of detection sensitivities

$$\alpha = [\alpha_1, \dots, \alpha_{n_\alpha}]. \quad (7)$$

Term $\overline{|\Delta h_i(k)|}$ is the average value of the weight increments in some window of length M . The Learning

Entropy algorithm needs at least 2 parameters to run, namely one detection sensitivity and length of the window. Optimal performance of the algorithm is achieved by using a higher number of detection sensitivities as some certain values.

B. Proposed Algorithm for Novelty Detection

A normal distribution is usually used in many novelty detection algorithms, however normal distribution is not good in handling extremely rare events. In this case, the Generalized extreme value (GEV) distribution can be used [9], [10], [11].

In this subsection, the new novelty detection algorithm is proposed. Proposed algorithm is based on the assumption, that the absolute value of adaptive weight increment, in case of sample perturbation or step change in parameters of measured system, are drawn from the GEV distribution (as we are going to investigate the absolute value of adaptive weight increment on data with perturbation, the normal distribution is not good choice). The algorithm evaluates the value of GEV probability density function (pdf) for every new weight increment. After that, it compares the increment with the threshold to estimate if the new sample is in the novelty class and then updates the estimation of parameters of the GEV pdf. It is possible to estimate the parameters of the GEV pdf only based on the last M number of samples as same as use all samples available.

Assume the probability density function of i -th adaptive parameter increment at the time index k in form

$$f_i(|\Delta h_i(k)|) = \frac{1}{\sigma_i} t_i(|\Delta h_i(k)|) \xi_i^{+1} e^{-t_i(|\Delta h_i(k)|)}, \quad (8)$$

with

$$t_i(|\Delta h_i(k)|) = \begin{cases} \left(1 + \xi_i \left(\frac{|\Delta h_i(k)|}{\sigma_i}\right)\right)^{\frac{-1}{\xi_i}} & \xi_i \neq 0, \\ e^{-\left(\frac{|\Delta h_i(k)| - \mu_i}{\sigma_i}\right)} & \xi_i = 0, \end{cases} \quad (9)$$

where parameters $\xi_i \in R$ (shape), $\mu_i \in R$ (location) and $\sigma_i > 0$ (scale) are estimated numerically by maximum likelihood method with each newly obtained weight increment Δh_i .

Now it is possible to evaluate the value of pdf for each adaptive parameter and based on the threshold, it is possible to estimate, whether the obtained sample is novel or not in respect to adaptive parameters. Assuming, that each adaptive parameter change is independent of changes of the other adaptive parameters leads to joint probability distribution function given as follows

$$f(|\Delta \mathbf{h}(k)|) = \prod_{i=1}^n f_i(|\Delta h_i(k)|). \quad (10)$$

The threshold ζ can be set in a way that the newly obtained sample is classified as novel if the equation

$$f(|\Delta \mathbf{h}(k)|) \leq \zeta \quad (11)$$

holds. The proposed algorithm thus needs to set only one (ζ) or two (ζ, M) parameters to run. Pseudocode of the proposed algorithm follows.

Algorithm 1 GEV distribution based adaptive ND

- 1: initial estimation of the GEV parameters ξ, μ, σ
 - 2: set novelty detection threshold ζ
 - 3: **for** each new obtained $d(k)$ **do**
 - 4: update adaptive model to get $\Delta \mathbf{h}(k)$
 - 5: compute $f(|\Delta \mathbf{h}(k)|)$
 - 6: **if** $f(|\Delta \mathbf{h}(k)|) \leq \zeta$ **then**
 - 7: successful novelty detection in sample $d(k)$
 - 8: **end if**
 - 9: update parameters ξ, μ, σ
 - 10: **end for**
-

The initial estimation of the parameters ξ, μ, σ can be obtained by using some initial set of samples and the setting of the parameter ζ depends on data. The experimental results to show the performance of the proposed algorithm follows in the next section. There is also shown the result of Learning Entropy algorithm.

IV. EXPERIMENTAL ANALYSIS

In this section, the proposed algorithm is studied in two basic testing schemes. The first schema is a sample perturbation detection, the second experimental schema is the performance in parameters of measured system change. The proposed method is compared with the Learning Entropy algorithm, which is also using the change of adaptive parameters to detect novelty in data. The synthetic data are used in this study for both experiments. Threshold ζ is not set as the aim is to present the performance of the proposed algorithm. The initial estimation of parameters ξ, μ, σ is done on same experiments using a different seed for randomness generation.

A. Sample Perturbation Detection

The adaptive model (1) is used to track the desired discrete signal d given by (12). At every discrete time sample index k the measured inputs $x_1(k)$ and $x_2(k)$ are obtained together with the measured output $d(k)$. The relation between the input vector \mathbf{x} and the output scalar value $d(k)$ is

$$d(k) = x_1(k) + x_2(k) + v(k), \quad (12)$$

where v represents additive Gaussian white noise, so the $v \sim N(0, 0.1)$. The values of the system inputs are drawn from normal distribution $N(0, 1)$. At time $k = 10^4$ there is a perturbation lasting one sample

$$v(10000) \sim N(0.5, 0.1). \quad (13)$$

An experiment is conducted for 11×10^3 samples. Fig. 1 shows the peak of jointed pdf that corresponds with a perturbation in the data. Fig. 2 shows the adaptive system output. Notice, that there is no significant change in data at sample $k = 10000$. Histogram of absolute values of adaptive weight h_1 increments is in Fig. 3. Histogram with adaptive weight h_2 is similar so it is omitted. Fig 4 demonstrates the results of

Learning Entropy algorithm with sensitivity parameters $\alpha = [1, 5, 10]$ and length of window $M = 30$. These parameters were obtained experimentally to enable Learning Entropy algorithm to detect perturbation in signal $d(k)$ at time $k = 10^4$. The peak in LE represents maximal learning effort of an adaptive system and thus can be interpreted as successful novelty detection. Fig. 5 shows the prediction error of an adaptive system during an experiment.

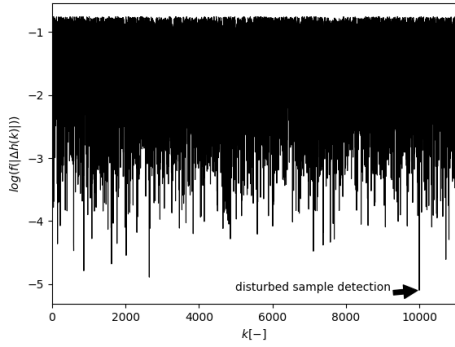


Figure 1. The logarithm of joint pdf $f(|\Delta h(k)|)$ during experiment A.

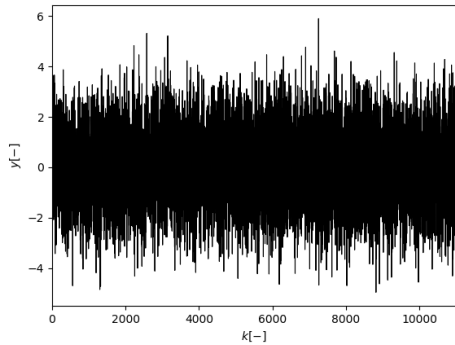


Figure 2. Output of the adaptive system $y(k)$ during experiment A.

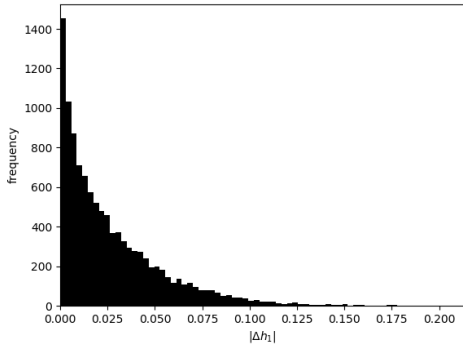


Figure 3. Histogram of absolute values of adaptive weight h_1 increments during experiment A.

B. Parameter Change Detection

The scheme of an experiment is the same as in the previous case, but the measured system is time-variant

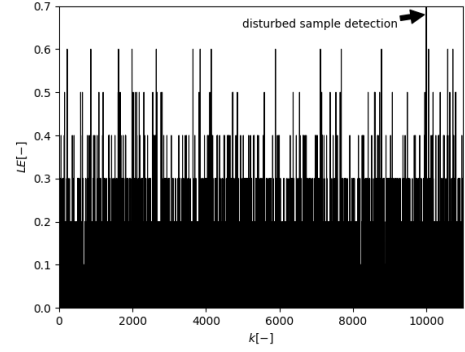


Figure 4. Learning Entropy algorithm during experiment A.

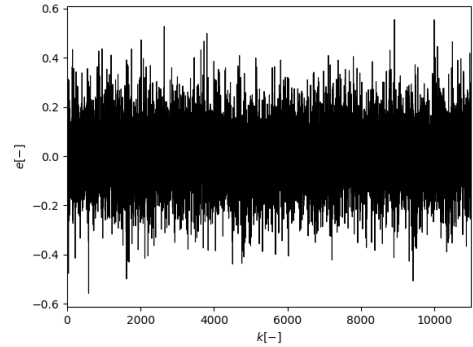


Figure 5. Prediction error e during experiment A.

and there is the step change in parameters at discrete time $k = 10^4$. The relation between inputs $x_1(k)$, $x_2(k)$ and output $d(k)$ is given as

$$d(k) = \begin{cases} x_1(k) + x_2(k) + v(k) & k \leq 9999, \\ 1.6x_1(k) + 0.4x_2(k) + v(k) & k \geq 10^4, \end{cases} \quad (14)$$

so the mean value of the desired signal $d(k)$ is not influenced by this change.

The experiment is conducted for 11×10^3 samples. Figure 6 shows the peak of joint pdf that corresponds with a perturbation in the data.

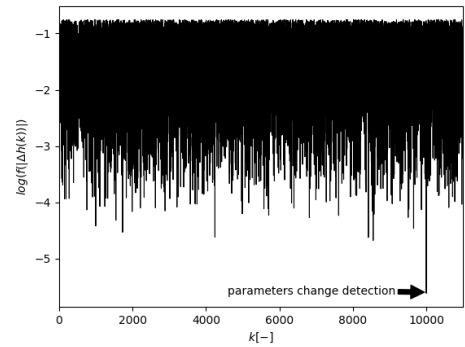


Figure 6. The logarithm of joint pdf $f(|\Delta h(k)|)$ during experiment B.

Fig. 7 shows the adaptive system output. Notice,

that there is no significant change in data at sample $k = 10000$. Histogram of absolute values of adaptive weight h_1 increments is in Fig. 8. Histogram with adaptive weight h_2 is similar so it is omitted. Fig. 9 demonstrates the results of Learning Entropy algorithm with sensitivity parameters $\alpha = [2, 3, 4, 7, 9]$ and length of window $M = 30$. These parameters were obtained experimentally to achieve at least partially successful detection. Fig. 10 shows the prediction error of an adaptive system during an experiment.

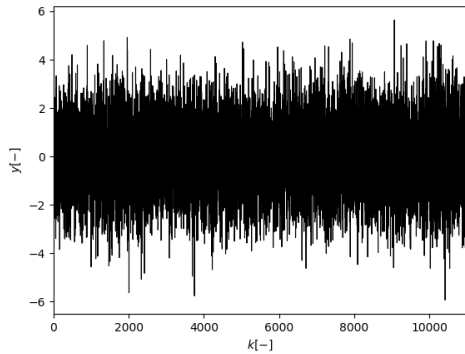


Figure 7. Output of the adaptive system $y(k)$ during experiment B .

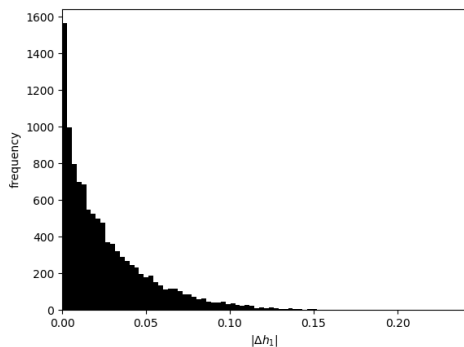


Figure 8. Histogram of absolute values of adaptive weight h_1 increments during experiment B .

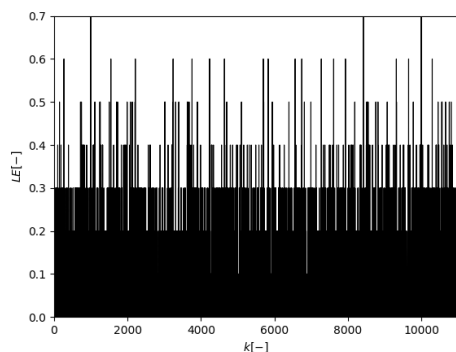


Figure 9. Learning Entropy algorithm during experiment B .

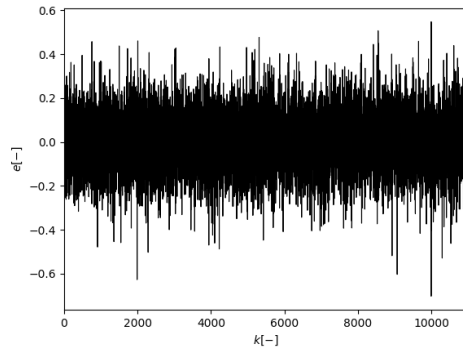


Figure 10. Prediction error e during experiment B .

V. CONCLUSION

This paper introduces the new novelty detection algorithm and its experimental study. The algorithm is evaluating the absolute value of adaptive system weight increments using generalized extreme value distribution. Two experiments with artificial data including novelty are presented. It is shown, that proposed algorithm is able to detect novelty in both experiments and comparison with Learning Entropy algorithm is presented.

ACKNOWLEDGMENT

The author would like to thank Matous Cejnek for developing PADASIP (Python Adaptive Signal Processing library). Financial support from specific university research (MSMT No 21-SVV/2018).

REFERENCES

- [1] M. Markou and S. Singh, "Novelty detection: a review part 1: statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481 – 2497, 2003.
- [2] M. Markou and S. Singh, "Novelty detection: A review - part 2: Neural network based approaches," *Signal Processing*, vol. 83, pp. 2499–2521, 2003.
- [3] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215 – 249, 2014.
- [4] I. Bukovsky, "Learning entropy: Multiscale measure for incremental learning," *Entropy*, vol. 15, no. 10, pp. 4159–4187, 2013.
- [5] M. Cejnek and I. Bukovsky, "Influence of type and level of noise on the performance of an adaptive novelty detector," in *2017 IEEE 16th International Conference on Cognitive Informatics and Cognitive Computing*, IEEE, jul 2017.
- [6] J. Vrba, "Vyuziti fuzzy systemu a algoritmu learning entropy pro detekci zmen stavu bioprocesu," *Automatizacia a riadenie v teorii a praxi ARTEP 2017 : 11. rocnik konferencie odbornikov z univerzit, vysokych skol a praxe.*, 2017.
- [7] I. Bukovsky, M. Cejnek, J. Vrba, and N. Homma, "Study of learning entropy for onset detection of epileptic seizures in EEG time series," in *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, jul 2016.
- [8] I. Bukovsky and C. Oswald, *Intelligent Systems in Cybernetics and Automation Theory: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol 2: ... in Intelligent Systems and Computing*. Springer, 2015.
- [9] C. Forbes, M. Evans, N. Hastings, and B. Peacock, *Statistical Distributions*. Wiley, 2010.
- [10] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer London, 2013.
- [11] H. joo Lee and S. J. Roberts, "On-line novelty detection using the kalman filter and extreme value theory," in *2008 19th International Conference on Pattern Recognition*, IEEE, dec 2008.