ROC Analysis of Extreme Seeking Entropy for Trend Change Detection

Jan Vrba, Jan Mareš
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 is dedicated to the evaluation of the ROC curve of recently introduced Extreme Seeking Entropy algorithm. The ROC curve is evaluated for a trend change in the signal that contains additive Gaussian noise. The resulting ROC curve of the Extreme Seeking Entropy algorithm is compared with other adaptive novelty detection methods, namely Learning Entropy and Error and Learning Based Novelty Detection as those algorithms are also evaluating the adaptive weights increments. The ROC curves are evaluated for multiple noise variances and area under those ROC curves is estimated.**

***keywords*—signal processing, adaptive systems, adaptive algorithms, novelty detection, trend change detection**

I. Introduction

Novelty detection (ND) is still an important topic and adaptive methods are getting importance as more and more data are processed nowadays. In decades many different approaches have been developed [1], [2], [3]. However, every ND method is performing well for specific data sets, so there is no universal method at all and thus the new ND methods are still needed. Two main approaches can be recognized in the ND topic, statistical (i.e. [4]) and neural network approach. From this point of view, the Learning Entropy (LE) [5] and Error and Learning Based Novelty Detection (ELBND) [6] can be classified as neural network approaches. The recently introduced Extreme Seeking Entropy algorithm (ESE) [7] can be classified as a combination of both.

In this, article we are going to estimate a receiver operating characteristic (ROC) curves [8] of the LE, ELBND and ESE algorithms. The ROC curve is a useful tool to visualize the ability of binary classification systems to classify data as the classification threshold is changing as well as to compare the performance of classification systems [9]. The scenario of trend change detection, where data contains additive gaussian noise, is evaluated. The trend change detection is a problem that is related i.e. to fault diagnosis [10]. The ESE, LE and ELBND methods have been selected as all of them are evaluating the adaptive weight increments that are obtained via sample-by-sample learning algorithms and all of those methods are utilizing simple adaptive systems to detect novelty in the data.

II. Adaptive System and Learning Algorithm Specification

In the experiments, the simple higher order neural unit (HONU) adaptive model is used. The output of this model at a discrete time index k is given as

$$y(k) = w_1(k)\cdot x_1(k) + w_2(k)\cdot x_2(k) + \\ + w_3(k)\cdot x_1(k)\cdot x_2(k) \quad (1)$$

which is equivalent with form

$$y(k) = \boldsymbol{w}^T(k)\cdot \boldsymbol{x}(k) \quad (2)$$

where

$$\boldsymbol{w}^T(k) = [w_1(k), w_2(k), w_3(k)] \in R^3$$

is the vector of adaptive weights and

$$\boldsymbol{x}^T(k) = [x_1(k), x_2(k), x_1(k)\cdot x_2(k)] \in R^3$$

is the input vector. The vector of adaptive weights is updated with every new sample obtained, formally the update of the adaptive weight vector is given as

$$\boldsymbol{w}(k+1) = \boldsymbol{w}(k) + \Delta\boldsymbol{w}(k), \quad (3)$$

where $\Delta\boldsymbol{w}(k)$ is the vector of weights increments. The initial setting of every adaptive weight is drawn from uniform distribution $U(-1,1)$.

The generalized gradient descent (GNGD) algorithm was used for the experiments. The update of this adaptive algorithm is given as

$$\Delta\boldsymbol{w}(k) = \eta(k)e(k)\mathbf{x}(k) \quad (4)$$

where $e(k)$ is prediction error given as

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

and

$$\eta(k) = \frac{\mu}{\|\mathbf{x}(k)\|_2^2 + \epsilon(k)}$$

$$\epsilon(k) = \epsilon(k-1) - \rho\mu\frac{e(k)e(k-1)\mathbf{x}^T(k)\mathbf{x}(k-1)}{(\|\mathbf{x}(k-1)\|_2^2 + \epsilon(k-1))^2}$$

where $d(k)$ is the output of the system, $\mu \in R$ is the learning rate, $\epsilon \in R$ is a compensation term, and ρ is the step size adaptation parameter, which should be chosen as $0 \leq \rho \leq 1$ to preserve the stability of the GNGD. The learning rate μ was set as $\mu = 0.5$ for all experiments.

ISBN 978-80-261-0722-4,

III. Novelty Detection Algorithms

A. Extreme Seeking Entropy Algorithm

The Extreme Seeking Entropy (ESE) algorithm [7] is utilizing the last n_s adaptive weights increments to detect novelty in the data. With every new adaptive weight increment obtained the peak-over-threshold (POT) method is proceed and if the increment is bigger than threshold ζ_i given by the POT method, the parameters of Generalized Pareto distribution are estimated and the probability of the weight increment is evaluated. Then ESE is calculated according to following formula (6).

$$ESE(|\Delta\mathbf{w}(k)|) = -\log\prod_{i=1}^{n}(1 - f_{cdf_i}(|\Delta w_i(k)|)) \tag{6}$$

where

$$f_{cdf_i}(|\Delta w_i(k)|) = \\ = \begin{cases} 0, |\Delta w_i(k)| < \zeta_i \\ F_{(\xi_i,\mu_i,\sigma_i)}(|\Delta w_i(k)|), |\Delta w_i(k)| \geq \zeta_i. \end{cases}$$

and $F_{(\xi_i,\mu_i,\sigma_i)}$ is cumulative distribution function of the generalized Pareto distribution with location parameter μ_i, scale parameter σ_i and shape parameter ξ_i and n is the number of adaptive weights.

Algorithm 1 Extreme Seeking Entropy Algorithm

```
1: set n_s and choose POT method
2: initial estimation of the GPDs parameters ξ_i, μ_i,
   σ_i for each adaptable parameter
3: for each new d(k) do
4:    update adaptive model to get Δw(k)
5:    proceed POT method
6:    if |Δw_i|(k) > ζ_i then
7:       update parameters of GPDs ξ_i, μ_i, σ_i
8:    end if
9:    compute ESE according to (6)
10: end for
```

The POT method was chosen according to [11] as 10%, so the parameters of GPD are estimated from the 10% of the highest weight increments.

B. Error and Learning Based Novelty Detection Algorithm

Error and Learning Based Novelty Detection algorithm (ELBND) utilize the prediction error and adaptive weight increments [6]. The novelty in the obtained sample is given as follows.

$$ELBND(k) = \sum_{i=1}^{n}|e(k) \cdot \Delta w_i(k)| \tag{7}$$

Note that there are multiple approaches to estimate ELBND [6], for the experiments we used the sum version.

C. Learning Entropy Algorithm

Recently the new version of the Learning Entropy algorithm has been published [5]. The direct version avoids the need for multiple scales [12] and estimates the learning entropy of the obtained sample at discrete time index k as

$$LE(k) = \sum_{i=1}^{n} z(|\Delta w_i(k)|) \tag{8}$$

with special $z - score$ that is given as

$$z(|\Delta w_i(k)|) = \frac{|\Delta w_i(k)| - \overline{|\Delta w_i^M(k-1)|}}{\sigma(|\Delta w_i^M(k-1)|)} \tag{9}$$

where $\overline{|\Delta w_i^M(k-1)|}$ is the mean of the last M adaptive ith weight increments and $\sigma(|\Delta w_i^M(k-1)|)$ is their standard deviation.

IV. Evaluation of the ROC curve

In this section, the ESE algorithm is compared with the Learning Entropy algorithm and Error and Learning Based Novelty Detection algorithm. The comparison is evaluated in trend change detection case with the synthetic dataset. The structure of the experiment was chosen as it is in [7] to extend already published results.

A. Description of the Experiment

The output of the system d at the discrete time index k is given by the equation

$$\begin{gathered} d(k) = x_1(k) + x_2(x) + 0.01 \cdot k + v(k) \\ 0 \leq k < 200 \end{gathered} \tag{10}$$

where $v(k)$ is additive Gaussian noise with zero mean and standard deviation σ_n and $x_1(k)$, $x_2(k)$ are inputs with values drawn from the uniform distribution $U(-1,1)$. At time $k = 200$ there is a step-change in the signal and the equation (10) changes to

$$\begin{gathered} d(k) = x_1(k) + x_2(x) + (0.01 + \delta) \cdot k + v(k) \\ 200 \leq k \leq 399 \end{gathered} \tag{11}$$

where the value of parameter δ is drawn from the uniform distribution $U(-0.02, 0.02)$ and is constant for the $200 \leq k \leq 399$. The initial estimation of GPD parameters and LE values was done using 1200 samples that were obtained from the system given by equation (10). Note that for the ESE algorithm the value of was set to $n_s = 1200$.

B. Receiver Operating Characteristic

To construct the receiver operating characteristic (ROC) the dataset with the same number of positive and negative cases is desirable. Each performed experiment consists of 400 samples and totally 10 000 experiments were performed. To obtain a suitable dataset for ROC every experiment was re-sampled as follows.

$$\begin{aligned} ND_r(i) = max\{&ND(i \cdot 10), ND(i \cdot 10 + 1), \\ &ND(i \cdot 10 + 2), \ldots, ND(i \cdot 10 + 9)\}, \\ &i = 0, 1, \ldots, 39 \end{aligned} \tag{12}$$

where $ND(k)$ represents the novelty measure in data at the discrete time index k (value of ESE, LE, $ELBND$). Every re-sampled data sequence was divided into two sets. Set P contain one positive sample $P = \{ND(20)\}$. Set N contain 39 negative samples, $N = \{ND(0), \ldots ND(19), ND(21), \ldots ND(39)\}$. For each experiment totally two samples that represents sets P and N were used to evaluate ROC, note that the sample from set N was drawn randomly. Totally there were 10 000 experiments performed for multiple values of noise variance σ_n. The values of noise variance were selected as follows $\sigma_n = \{0.1, 0.2, 0.5, 1.0, 2.0, 2.5\}$ and signal-to-noise ratio (SNR) is evaluated based from all experiments for each noise variance as

$$SNR = 10 \log_{10} \frac{\sigma_s}{\sigma_n} \tag{13}$$

where σ_s is the standard deviation of the output of the system. The True positive rate (TPR) correspond to hit rate and it is a ratio of true positive cases (TP) for a given threshold to the total number of positive cases.

$$TPR = \frac{TP}{10000} \tag{14}$$

False positive rate (FPR) is given as ratio of false positive cases (FP) for given threshold to total number of negative cases.

$$FPR = \frac{FP}{10000} \tag{15}$$

C. Results

The resulting ROC curves for various SNR are depicted in following Figures 1-6. The blue solid line depicts the results of the Extreme Seeking Entropy algorithm, the green dotted line Learning Entropy algorithm, the red dashed line Error and Learning Based Novelty Detection algorithm and black black dash-dotted line random classifier. For each ROC curve the

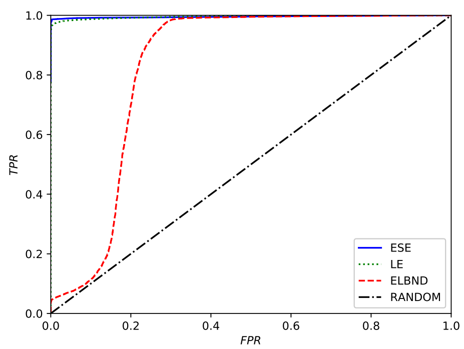

Figure 1. ROC curve for trend change detection of the signal containing additive Gaussian noise with variance $\sigma = 0.1$ and corresponding $SNR = 35.8\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

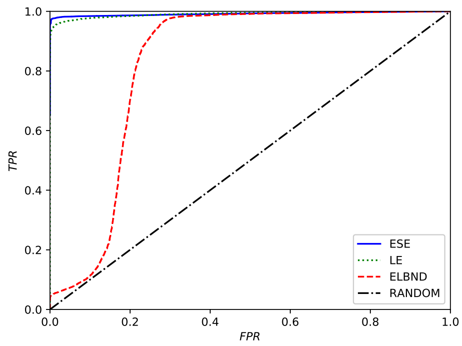

Figure 2. ROC curve for trend change detection of the signal containing additive Gaussian noise with variance $\sigma = 0.2$ and corresponding $SNR = 30.0\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

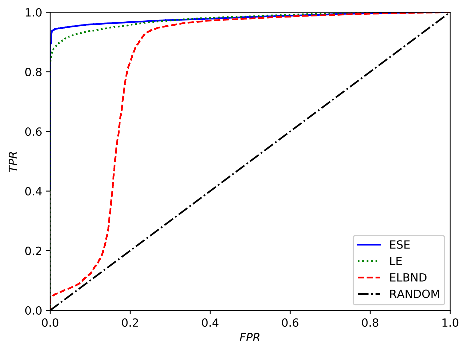

Figure 3. ROC curve for trend change detection of the signal containing additive Gaussian noise with variance $\sigma = 0.5$ and corresponding $SNR = 21.7\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

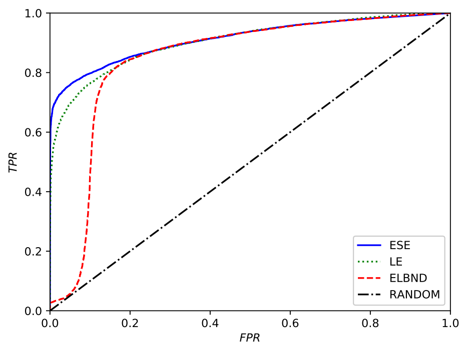

Figure 4. ROC curve for trend change detection of the signal containing additive Gaussian noise with variance $\sigma = 1.0$ and corresponding $SNR = 16.2\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

area under the curve $(AUROC)$ was estimated using

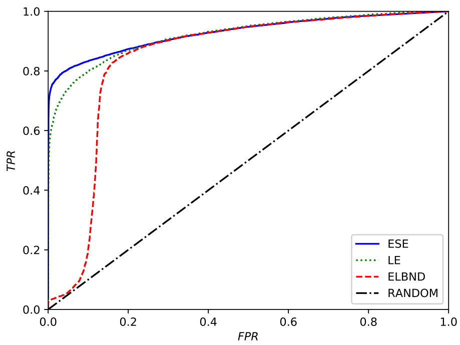

Figure 5. ROC curve for trend change detection of signal containing additive Gaussian noise with variance $\sigma = 2.0$ and corresponding $SNR = 10.88\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

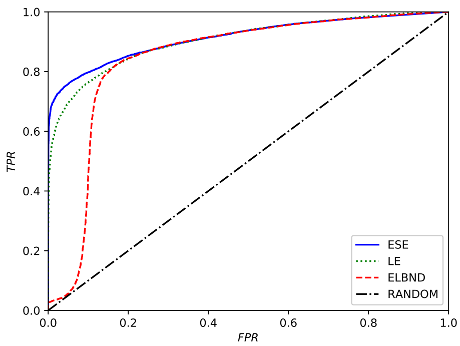

Figure 6. ROC curve for trend change detection of the signal containing additive gaussian noise with variance $\sigma = 2.5$ and corresponding $SNR = 9.2\ dB$. The acronyme ESE is the Extreme Seeking Entropy, LE is the Learning Entropy, ELBND is the Error and Learning Based Novelty Detection method and RANDOM is the random classifier.

trapezoidal rule:

$$AUROC = \int_0^1 TRP(FPR) \approx$$
$$\approx \sum_{j=1}^{19999} \frac{TPR(FPR(j)) + TPR(FPR(j+1))}{2} \cdot$$
$$\cdot (FPR(j+1) - FPR(j)) \quad (16)$$

and resulting $AUROC$s are in following table I. For all six experiments, the detection rate was evaluated. The successful detection is considered if the maximum in the ND score corresponds with the positive sample. The results of detection rates are depicted in the following table II.

TABLE I
$AUROC$ FOR TREND CHANGE DETECTION

		AUROC		
σ_n	*SNR [dB]*	*ESE*	*LE*	*ELBND*
0.1	35.8	0.9954	**0.9952**	0.8234
0.2	30.0	**0.9920**	0.9912	0.8299
0.5	21.7	**0.9816**	0.9777	0.8288
1.0	16.2	**0.9576**	0.9496	0.8263
2.0	10.8	**0.9286**	0.9214	0.8397
2.5	9.2	**0.9134**	0.9056	0.8446

TABLE II
DETECTION RATES FOR TREND CHANGE DETECTION

		Detection rate		
σ_n	*SNR [dB]*	*ESE*	*LE*	*ELBND*
0.1	35.8	98.88	**98.92**	60.00
0.2	30.0	**98.14**	98.03	59.61
0.5	21.7	**95.18**	95.08	59.65
1.0	16.2	**90.42**	89.96	57.67
2.0	10.8	**81.27**	78.51	57.69
2.5	9.2	**75.86**	71.56	57.16

V. CONCLUSION

This paper is presenting the ROC curves for trend change detection problem with artificial data. The ROC curves and area under ROC curve are evaluated for various values of SNR as the signal contains additive Gaussian noise. The comparison of Extreme seeking entropy algorithm, Learning entropy algorithm and Error and Learning Based Novelty detection algorithm is shown. From the obtained results it seems that the ESE and the LE algorithms slightly outperform the ELBND algorithm in the presented experiments. The drawback of using ESE is its higher computational complexity.

ACKNOWLEDGMENT

The authors would like to thank Matous Cejnek for developing PADASIP (Python Adaptive Signal Processing library). This work was supported from the grant of Specific university research – grant No. A1_FCHI_2020_002.

REFERENCES

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

[2] M. Markou and S. Singh, "Novelty detection: a review—part 1: statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481 – 2497, 2003.

[3] A. Zimek and P. Filzmoser, "There and back again: Outlier detection between statistical reasoning and data mining algorithms," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, Aug. 2018.

[4] M. Desforges, P. Jacob, and J. Cooper, "Applications of probability density estimation to the detection of abnormal conditions in engineering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 212, no. 8, pp. 687–703, 1998.

[5] I. Bukovsky, W. Kinsner, and N. Homma, "Learning entropy as a learning-based information concept," *Entropy*, vol. 21, p. 166, feb 2019.

[6] M. Cejnek and I. Bukovsky, "Concept drift robust adaptive novelty detection for data streams," *Neurocomputing*, vol. 309, pp. 46–53, 2018.

[7] J. Vrba and J. Mareš, "Introduction to extreme seeking entropy," *Entropy*, vol. 22, p. 93, Jan. 2020.

[8] J. P. Egan, *Signal detection theory and ROC-analysis*. Academic press, 1975.

[9] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[10] M. R. Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Engineering Applications of Artificial Intelligence*, vol. 20, pp. 133–146, Mar. 2007.

[11] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: A critique," *The Annals of Statistics*, vol. 11, pp. 1019–1031, Dec. 1983.

[12] I. Bukovsky, "Learning entropy: Multiscale measure for incremental learning," *Entropy*, vol. 15, pp. 4159–4187, Sept. 2013.