IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Robust Approach to Estimation of the Intensity of Noisy Signal with Additive Uncorrelated Impulse Interference

Andriy Lozynskyy
Karpenko Phisico-Mechanical Institute
NAS of Ukraine
Lviv, Ukraine
lozynsky@ah.ipm.lviv.ua

Igor Romanyshyn
Karpenko Phisico-Mechanical Institute
NAS of Ukraine
Lviv, Ukraine
romanyshyn@ipm.lviv.ua

Bohdan Rusyn
Karpenko Phisico-Mechanical Institute
NAS of Ukraine
Lviv, Ukraine
rusyn@ipm.lviv.ua

Volodymyr Minialo
Karpenko Phisico-Mechanical Institute
NAS of Ukraine
Lviv, Ukraine
minialo@ipm.lviv.ua

***Abstract* – A robust approach to estimation the intensity of a noisy signal with additive uncorrelated impulse interference is proposed. An occurrence of the additive uncorrelated impulse interference leads to increasing of the observed signal dispersion within some sections with impulse interference. Robustness of the intensity estimation is achieved by decreasing the influence of sections with impulse interference. A number of nonlinear filtering methods basing on lower envelope detection are developed: two-parameter recursive filter, dilation filter, clipping derivative filter and filters based on order statistics. Proposed approach was approbated by a numerical simulation. Numerical simulation is validated the efficiency of the proposed approach for estimation the intensity of a noisy signal with additive uncorrelated impulse interference at dynamic data mining and data stream mining.**

Keywords – noisy signal, additive uncorrelated impulse interference, random signal parameters estimation, robust method, nonlinear filtering

I. Introduction

Estimation of quasistationary noisy signal intensity with additive noise is the important problem in dynamic data mining and data stream mining.

If a noisy signal is represented as a set of random variables then a problem of estimation of noisy signal parameters is reduced to a problem of estimation of these random variables parameters [1-15].

A basis for estimation of random variables parameters is probability density function. Mainly, at regular random signal analysis, one is estimated a mean value (amplitude characteristic) and dispersion (energy characteristic).

At analysis of one-dimensional random variables that conforms to a first two moments: mean value and dispersion (location and scale). First two moments are fully characterize a normal distribution. In classical statistics for their estimation are used the arithmetic mean and standard deviation.

The arithmetic mean and standard deviation are very responsive to outliers and, therefore, are unreliable estimations in this case [1, 6]. The most responsive to outliers is standard deviation. As classical example is the Tukey method, which illustrates the influence of outliers on robustness of standard deviation estimation [2]. It is shown, only two “bad” measurements out of thousand can substantially worse the estimation of standard deviation versus mean absolute deviation [1, 2, 8]. Mean absolute deviation is a robust parameter for estimation of random variable scale [1]. At the same time, the impulse interference is affecting on it also, as on the arithmetic mean.

In presence of outliers, the median is the most insensitive value of location estimation versus mean value [8, 15]. It should be noticed that Kolmogorov paper is one of the first describing the application of median for estimation of random variable location [18].

For estimation a random variables location at outliers occurrence on the base of order statistics, α - trimmed values and α - Winsorized are used [13].

α - trimmed value is obtained, when in variational series α percent of series length are rejected from the beginning and from the end; the arithmetic mean of rested series, in this case, acts as location parameter.

α - Winsorized value is obtained, when in variational series, α percent of series length from the beginning of series are substituted by the least value of rested series, and α percent (also) - from the end of series by the largest value. The arithmetic mean of Winsorized series, in this case, acts as location parameter.

Using methods of ordinal statistics to time series filtering is performing with “window” filtering. In other words, the interval (“window”) is chosen for data filtering and this interval is moving sequentially, sample by sample.

On the other hand, a problem of estimation the intensity of a noisy signal with additive noise is reduced to signal filtering in frequency and time domain [15, 22].

However, modern methods don’t provide a good estimation of location and scale with impulse interference.

978-1-5386-2874-4/18/$31.00 ©2018 IEEE

II. Filtering Methods on the Basis of Lower Envelope Detection

A. Statement of problem

A problem of estimation the intensity of a noisy signal is reduced to estimation the dispersion of random signal with interference.

Dispersion of random signal with additive uncorrelated noise has two components: a desired dispersion of random signal and noise dispersion. Impulse interference leads to increasing a dispersion of the registered signal on interferenced interval. Estimation of random signal dispersion is reduced to estimation of random variable location and is achieved by decreasing influence of these intervals.

Let's have a sequence of noisy samples s_{n_i} of signal intensity:

$$s_{n_i} = s_i + \xi_i\,,\ \forall i = \overline{1, N}\,, \tag{1}$$

where $s_i \geq 0$ - utility noisy signal, $\xi_i \geq 0$ - random interference with random occurrence time and random duration.

The problem of estimation the mean value (location) of a random variable is set:

$$\bar{s} = \frac{1}{N}\sum_{i=1}^{N} s_i \tag{2}$$

on basis of noisy samples (1).

It's clear that averaging of s_{n_i}, leads to:

$$\bar{s}_{n_i} = \bar{s} + \bar{\xi} = \frac{1}{N}\sum_{i=1}^{N}\left(s_i + \xi_i\right), \tag{3}$$

that can't be a good estimator.

In case of normal distribution of random variables s_i, ξ_i, distribution of sum (1), usually is labeled as distributions with «heavy-tailed» [1, 16]. Numerous experiments have shown that application of standard methods (on basis of mean value calculation (3), median filtering, calculation of trimmed- average, Winsorization, etc.) is not a good estimator of location (2).

A new approach is proposed for estimation the intensity of a noisy signal with uncorrelated noise that is based on s_{n_i} lower envelope signal detection. It allows to decrease, substantially, the influence of intervals with impulse interference at estimation of random variable location (2). According to the proposed approach, a number of methods were elaborated: two-parameter recursive filter, dilation filter, clipping derivative filter as well as a series of modified methods, based on order statistics (median filtering, trimmed-average, Winsorization).

B. Two-parameter recursive filter

A method that based on nonlinear filtering for estimation of mean value (2) of noisy realization (1) is proposed.

These filters are based on exponential smoothing [20-22] and operation as follows:

$$\begin{aligned} & y_1 = s_1 + \xi_1, \\ & y_i = \begin{cases} k_1\left(s_i + \xi_i\right) + \left(1 - k_1\right)y_{i-1},\ s_i + \xi_i > y_{i-1} \\ k_2\left(s_i + \xi_i\right) + \left(1 - k_2\right)y_{i-1},\ s_i + \xi_i \leq y_{i-1} \end{cases} \\ & i = 2,3,\ldots N \end{aligned} \tag{4}$$

Here $k_1 \sim 0$ - coefficient, near-zero, $k_2 \sim 1$ - coefficient, near-one. Such selection of coefficients means that if current value of the registered signal exceeds a previous filtered value (usually interferenced value) then we receive a value that is near to previous (first row of y_i in (4)). In case of noise absence, filtering is performed according to lower envelope. Such filtering method at which i - value of filtering signal is expressed in terms of $i-1$ (with weight coefficient) leads to displacement of sequence to the right on one step. For elimination of such effect is used an inversely filtration - from N - sample to first:

$$\begin{aligned} & \tilde{y}_N = y_N, \\ & \tilde{y}_i = \begin{cases} k_1 y_i + \left(1 - k_1\right)\tilde{y}_{i+1},\ y_i > \tilde{y}_{i+1} \\ k_2 y_i + \left(1 - k_2\right)\tilde{y}_{i+1},\ y_i \leq \tilde{y}_{i+1} \end{cases} \\ & i = N-1, N-2, \ldots, 1 \end{aligned} \tag{5}$$

Such two-stage filtering allows us to eliminate the effect of sequence displacement that appears during one-stage filtering procedure.

It should be marked that single executing of the proposed filtering procedures, may not be enough to filter the data. Therefore, the filtration continues until the desired result is achieved.

C. Dilation method

Dilation is a filtering method, simpler versus two-parameter and it consists in the following:

$$\begin{aligned} & y_1 = s_1 + \xi_1; id = 0; \\ & y_i = \begin{cases} s_i + \xi_i; id = 0; s_i + \xi_i \leq y_{i-1} \\ y_{i-1}; id = id + 1; s_i + \xi_i > y_{i-1} \end{cases} \\ & i = 2,3,\ldots, N \end{aligned} \tag{6}$$

As one can see, dilation method is based on replacement of the samples in increasing interval to previous value. The maximal size of this interval is given by the parameter $id_{\max}$, which is determined on the estimation of the duration of the interference. The parameter id changes to $id_{\max}$, then reset, filter output is assigned to an current value and the filtration cycle (6) is repeated. Dilation method, as well as two-parameter method, leads to displacement of sequence to the right. Next stage is used for elimination of this effect:

$$\tilde{y}_N = y_N; id = 0;$$

$$\tilde{y}_i = \begin{cases} y_i; id = 0; y_i \le \tilde{y}_{i+1} \\ \tilde{y}_{i+1}; id = id + 1; y_i > \tilde{y}_{i+1} \end{cases} \quad (7)$$

$$i = N-1, N-2, \ldots, 1$$

D. Method of clipping derivative

The present method is based on comparison of current sample with previous one and when difference between them exceed a defined value, current value is substituted by this value.

$$y_1 = s_1 + \xi_1;$$

$$y_i = \begin{cases} s_i + \xi_i; (s_i + \xi_i) - (s_{i-1} + \xi_{i-1}) \le y_{i-1} p; \\ s_i + \xi_i + p; (s_i + \xi_i) - (s_{i-1} + \xi_{i-1}) > y_{i-1} p \end{cases} \quad (8)$$

$$i = 2, 3, \ldots, N$$

The parameter p is chosen on the assumption of allowable increasing of signal intensity (1). Trimming of difference is performed adaptively - proportionally to current filter output.

As in the previous methods, reiteration from end to start sequence may be used for elimination the effect of sequence displacement.

E. Methods on basis of order statistics

Order statistics methods are based on forming of variational series during "window" filtering and selection the value near the beginning of variational series as location parameter.

III. Numerical Simulation

A noisy signal with interference was simulated as sum of two random variables with a distribution close to the normal and with given mean values. Probability of interference appearance was 0.03, and probability of disappearance – 0.1.

Figure 1 presents an example of simulated signal and results of filtering by two-parameter recursive filter with $k_1 = 0.05$, $k_2 = 0.5$ parameters.

Figure 2 shows a comparison results of two median filters (standard and modified), each of which is based on 17 samples. An input signal was simulated similarly to the previous case, with the probability of interference appearance rising from experiment to experiment from 0 to 0.5 (dotted line), and probability of "disappearance" equal to 0.5.

Here, a horizontal (black) line corresponds to mean value (etalon) that doesn't change from the experiment to experiment and was equal to 2. A green line - to the result of modified median filtering with the fourth value of each "window" was taken into account. And yellow line - to the result of standard median filtering at which the ninth (central) value of each "window" was taken into account.

As one can see, based on the simulation signal, modified filtering is the better way to estimate of the location (1).

Filtering results (location estimation and its standard deviation) versus level of the interference is presented in Table I. It is shown the results of standard and modified median filtering (SMF and MMF, accordingly) in the following format: (location ± standard deviation)/etalon for different interference levels.

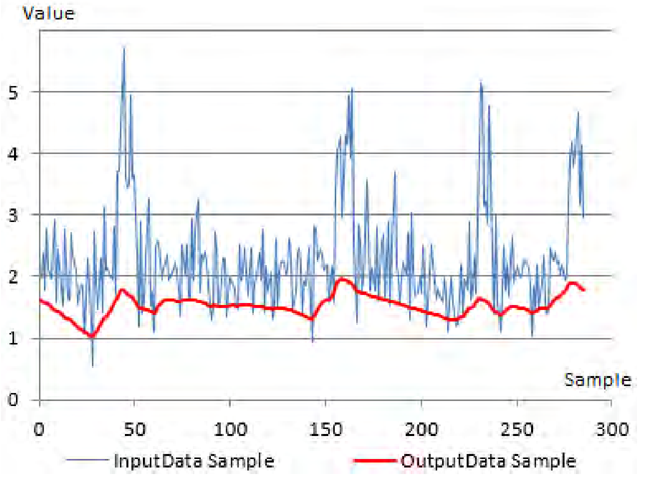

Fig. 1. Simulated signal with interference and filtering result at lower envelope by two-parameter recursive filter

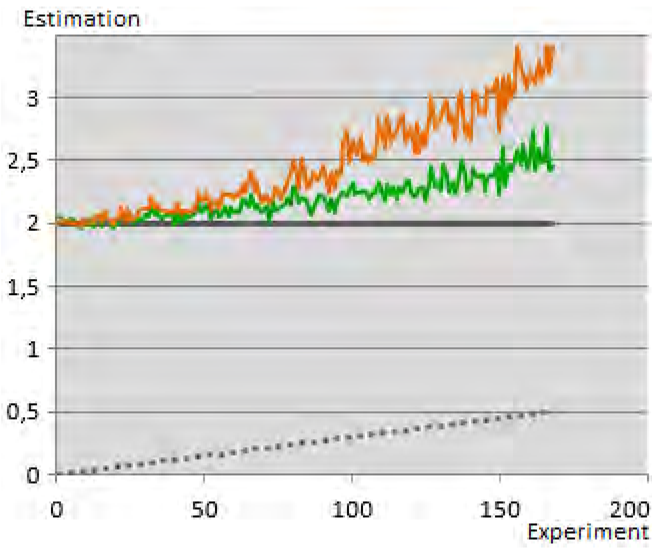

Fig. 2. Comparison of standard and modified median filters

TABLE I. DEPENDENCE OF THE MEDIAN FILTERING RESULTS FOR SOME LEVELS OF INTERFERENCE

Interference level	SMF	MMF
0	1.00±0.01	1.00±0.02
1	1.24±0.03	1.20±0.03
2	1.46±0.06	1.25±0.05
3	1.67±0.10	1.28±0.06
4	1.87±0.11	1.32±0.07

As one can see, without interference a standard median filtering method gives a less deviation for location estimation. At increasing level of interference, a modified median filtering method, based on the selection of less element, prevails over standard.

Table II shows the results of filtering (location and deviation) using different methods, developed by authors. Input signal was simulated similarly to the previous case, and probability of the interference appearance and "disappearance" was equal to 0.5 both.

The base of order statistics consists of 17 samples. For the standard median filtering, as usual, a central (ninth) value was selected. For the modified median filtering – the fourth value, trimm–mean was performed basing on first 1-3 values, Winsorization – on the base of 2-4 values. Derivative clipping parameter was 0.07 and dilation parameter value was 11.

TABLE II. FILTERING METHODS COMPARISON

Method	Location/etalon	Deviation/etalon
Median filter	1.688	0.094
Recursive filter	1.331	0.043
Dilation	1.210	0.059
Derivation crop	1.185	0.054
Trimm-filter	1.212	0.045
Winsorization	1.262	0.055
Modified median filtering	1.293	0.059

As one can see, a standard median filter is substantially yield to developed filtering methods, based on lower envelope detection.

IV. CONCLUSIONS

A new approach to estimation of the noisy signal intensity with additive uncorrelated impulse interference in the field of dynamic data mining and data stream mining, based on filtering using lower envelope detection, is proposed. For this, a number of nonlinear filtering methods are developed, namely: two-parameter recursive filter, dilation filter, clipping derivative filter and filters based on order statistics.

Results of numerical experiments are indicated about insufficient applicability of standard filtering methods of noisy signals of considered class and appreciable benefit (on accuracy) of developed methods, based on the proposed approach, versus standard.

REFERENCES

[1] S. A. Ayvazyan, I. S. Enyukov, L. D. Mehsalkin, Applied statistics: Rudiments of simulation and data preprocessing. M.:Finances and Statistics, 1983.

[2] P. J. Huber, Robust statistics. M.: Mir, 1984. (In Russian)

[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, Robust statistics. M.: Mir, 1989. (In Russian)

[4] J. W. Tukey, A survey of sampling from contaminated distributions. In: Contributions to Prob. and Statist. (Ed. Olkin I. et al.). Stanford Univ. Press. 1960, pp. 448–485.

[5] A. W. F. Edwards, "Three Early Papers on Efficient Parametric Estimation," Statistical Science, vol. 12, no. 1, pp. 35-47, 1997.

[6] G. Shevlyakov, and P. Smirnov, "Robust Estimation of the Correlation Coefficient: an Attempt of Survey," Austr. J. of Statistics, vol. 40, no.1&2, pp. 147-156, 2011.

[7] P. O. Smirnov, Robust methods and algorithms of estimation the correlation data characteristics on the basis of new high-performance and rapid robust scale estimations. (Candidate dissertation). St. Petersburg, 2013.

[8] C. Croux, and C. Dehon, Robust estimation of location and scale. Encyclopedia of Environmetrics, A.-H. El-Shaarawi and W. Piegorsch (eds). John Wiley & Sons Ltd: Chichester, UK, Retrieved from 2013. https://feb.kuleuven.be/public/u0017833/PDF-FILES/Croux_Dehon5.pdf.

[9] G. E. P. Box, "Non-Normality and Tests on Variance" Biometrika, vol. 40, pp. 318–335, 1953.

[10] P. J. Bickel, and E. L. Lehmann, "Descriptive Statistics for nonparametric models. I.," Introduction. The Annals of Statistics, vol. 3, no. 5, pp. 1038-1044, 1975.

[11] P. J. Bickel, and E. L. Lehmann, "Descriptive Statistics for nonparametric models. II.," Location. The Annals of Statistics, vol. 3, no. 5, pp. 1045-1069, 1975.

[12] P. J. Bickel, and E. L. Lehmann, "Descriptive Statistics for nonparametric models. III.," Dispersion. The Annals of Statistics, vol. 4, no. 6, pp. 1139-1158, 1976.

[13] D. E. Tyler, A short course on robust statistics. Retrieved from http://www.rci.rutgers.edu/~dtyler/ShortCourse.pdf.

[14] P. J. Rousseeuw, and C. Croux, "Alternatives to the Median Absolute Deviation," Journal of the American Statistical Association, vol. 88, 424, pp. 1273-1283, 1993.

[15] Christophe Leys, Christophe Ley UGent, Olivier Klein, Philippe Bernard and Laurent Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," Journal of Experimental Social Psychology, vol. 49(4), pp.764-766, 2013. Retrieved from http://dx.doi.org/10.1016/j.jesp.2013.03.013.

[16] A. Chakrabarty, "Large Deviations for Truncated heavy-tailed random variables: a boundary case," Indian J. Pure Appl. Math., vol.48 (4), pp. 671-703, 2017.

[17] R. A. Fisher, "On the Mathematical Foundations of Theoretical Statistics," Phil. Trans. R. Soc. Lond. A., vol. 222, pp. 309-368, 1992. doi: 10.1098/rsta.1922.0009.

[18] A. N. Kolmogorov, "The method of the median in the theory of errors," Mathematical collect., vol.38, no. 3-4, pp. 47-50, 1931.

[19] A. A. Lyubushin, Analysis of data from geophysical and environmental monitoring systems. M.: Nauka, 2007.

[20] E. S. Gardner, Exponential smoothing: the state of the art. Part II. Houston, 2005.

[21] Yu. S. Dodonov, and Yu. A. Dodonova, "Stable measures of central tendency: weighing as probable alternative of data truncation at the response time analysis," Psychological researches, vol. 5(19), pp. 1–14, 2011. Retrieved from http://psystudy.ru.

[22] Predicting time series using exponential smoothing. Retrieved from https://www.mql5.com/ru/articles/346.