EGLD system for noise identification in predictors ensemble context

Ryszard Szupiluk
Warsaw School of Economics
Al. Niepodległosci 162, 02-554 Warsaw, Poland
rszupi@wp.pl

Tomasz Ząbkowski
Warsaw University of Life Sciences (SGGW)
Faculty of Applied Informatics and Mathematics
Nowoursynowska 159, 02-787 Warsaw, Poland
tomasz_zabkowski@sggw.pl

***Abstract* — In this article we develop a noise identification method for data without temporal structure. The problem of noise detection will be presented in the framework of ensemble methods via blind signals separation.**

Keywords— independent components analysis, short term energy forecasting, neural networks, Extended Generalized Lambda Distribution

I. INTRODUCTION

The noise signals are typically modelled in stochastic processes approach as a sequence of independent identically distributed random variables. Such apparatus has deep and wide theoretical and practical meaning, but also it has some fundamental limitations. Especially, it is addressed for data with temporal structure [1]. In many problems, observation order is not important, or the analytical method simply neglects this properties. One of such examples is noise filtration from prediction results via independent component analysis methods, what can be treated as a kind of ensemble method [1].

In this filtration/ensemble approach, the prediction results from different models are treated as a multidimensional variable containing hidden constructive and destructive components. These latent components are identified using the independent component analysis (ICA) methods [3,4]. The key issue in this method is the correct classification and distinction between destructive and constructive components [5]. For this task we present novel concept based on Extended Generalized Lambda Distribution (EGLD) [6].

II. PREDICTION RESULTS IMPROVEMENT

We assume, that after the learning process, each prediction result x_i includes two types of latent components: constructive $\hat{\mathbf{s}}_j$, associated with the target, and destructive $\mathbf{s}_i$, associated with the inaccurate learning data, individual properties of models, missing data, not precise parameter estimation, distribution assumptions etc [2]. The relation between observed prediction $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T$ results and latent components $\mathbf{S} = [\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \ldots \hat{\mathbf{s}}_k, \mathbf{s}_{k+1}, \mathbf{s}_n]^T$ can be expressed as linear transformation can be assumed as

$$\mathbf{X} = \mathbf{AS}\,, \tag{1}$$

where matrix $\mathbf{A} \in R^{n \times n}$ represents the mixing system.

Our aim is to find the latent components and reject the destructive ones (replace them with zero). Next we mix the constructive components back to obtain improved prediction results as

$$\hat{\mathbf{X}} = \mathbf{A}\hat{\mathbf{S}} = \mathbf{A}[\hat{\mathbf{s}}_1, \hat{\mathbf{s}}_2, \ldots, \hat{\mathbf{s}}_k, \mathbf{0}_{k+1}, .., \mathbf{0}_n]^T\,. \tag{2}$$

Under some conditions the estimation **A** and **S** can be perform by ICA method where the Natural Gradient algorithm is one of popular methods [3]. For $\mathbf{A} \approx \mathbf{W}^{-1}$ it can be described by iterative rule as

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)\left[\mathbf{I} - E\left\{\mathbf{f}(\mathbf{S})\mathbf{S}^T\right\}\right]\mathbf{W}(k)\,, \tag{3}$$

where is a vector of nonlinearities with optimal form of $f_i(\mathbf{s}_i) = d\log(p_i(\mathbf{s}_i))/d\mathbf{s}_i$, $E\{.\}$ means expected value operator, $p_i(\mathbf{s}_i)$ is pdf of $\mathbf{s}_i$. The main problem after components **S** estimation is to identify destructive ones.

III. STATISTICAL ANALYSIS OF DESTRUCTIVE COMPONENTS

The component classification problem can be treated as the problem of finding similarity (or divergence) between the signals. Due to assumptions related to ICA application and its properties associated with higher order statistic we focus on signal characteristics in terms of kurtosis and skewness, what provide us to Extended Generalized Lambda Distribution (EGLD) model. EGLD system consists of two distributions: Generalized Lambda Distribution (GLD) and Generalized Beta Distribution (GBD) [6,7]. GLD is a distribution which doesn't have a probability density function or distribution function that can be specified explicit, but its distribution is defined by the inverse of the $F(y)$ distribution [6]:

$$F^{-1}(y) = \lambda_1 + \frac{y^{\lambda_3} - (1-y)^{\lambda_4}}{\lambda_2}\,, \tag{4}$$

where $0 \le y \le 1$ and $\lambda_1, \lambda_2, \lambda_3.\lambda_4$ are parameters of distribution, which exists if $\lambda_2\left(\lambda_3 y^{\lambda_3 - 1} + \lambda_4 (1-y)^{\lambda_4 - 1}\right)^{-1} \ge 0$.

The distribution parameters can be determined with the use of L-moments [8]. The other distribution, which is GBD, is determined by the probability density function [6]:

$$p(y) = c\beta_2^{-(\beta_3+\beta_4+1)}(y-\beta_1)^{\beta_3}(\beta_1+\beta_2-y)^{\beta_4}, \quad (5)$$

which exists in an interval of $[\beta_1, \beta_1+\beta_2]$ and c is a constant. The parameters of nonlinearity can be estimated with method of moments. From practical point of view the important advantage of the EGLD system is fact that most of distributions, for which kurtosis and skewness exist, can be expressed in coherent way by four parameters. For signal x_i we obtain parameters vector $\mathbf{B}(i) = [\beta_1(i), \beta_2(i), \beta_3(i), \beta_4(i)]$ for GBD or $\mathbf{\Lambda}(i) = [\lambda_1(i), \lambda_2(i), \lambda_3(i), \lambda_4(i)]$ for GLD. The signal comparison via EGLD system can be made in two ways. The first method aim to compare directly parameters of EGLD system what can measured by chosen p-norm

$$L_{ij} = \|\mathbf{\Lambda}(i) - \mathbf{\Lambda}(j)\|_p . \quad (6)$$

In Fig. 1, for p=2, we can observe, relation between signal to noise ratio (SNR) and L measure for mixture deterministic sinusoidal signal and Gaussian noise. The monotonically characteristic of this relation indicate for adequateness in signal randomness measure.

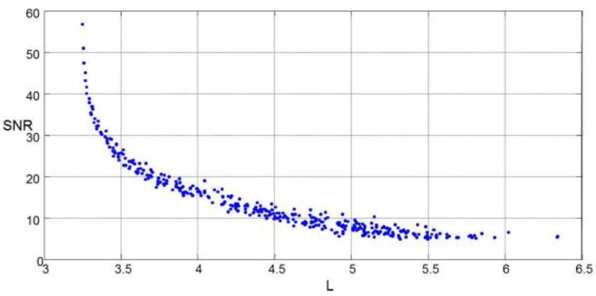

FIGURE. 1. THE RELATION BETWEEN L MEASURE (HORIZONTAL AXIS) AND SNR (VERTICAL AXIS) FOR NOISY SINUSOID.

The second method for signals comparison, via EGLD system, assume generation of simulated data from EGLD with particular lambda parameters, and next measuring distance between them by divergence function, for example, for Kullback-Leibler divergence we obtain similarity factor

$$D = D_{KL}(\mathbf{x}_i \| \mathbf{x}_i) = \sum_{k=1}^{N} \mathbf{x}_i(k) \ln \frac{\mathbf{x}_i(k)}{\mathbf{x}_j(k)}, \quad (7)$$

where $\mathbf{x}_i \propto \text{EGLD}(\mathbf{\Lambda}(i))$, $\mathbf{x}_j \propto \text{EGLD}(\mathbf{\Lambda}(j))$.

The practical application is presented in next paragraph.

IV. NUMERICAL EXPERIMENT

The validation test of the proposed concept with noise detection was performed on the problem of load prediction in the Polish power system using hourly data from 1988 till 1998. We trained six MLP neural networks with one hidden layer (with 12, 18, 24, 27, 30, 33 neurons respectively). After models decomposition we obtained six components, see Fig. 2, for which we calculated similarity factors D (7). Tab. 1 presents the results of primary models and after ICA filtration using mean absolute percentage error (MAPE). The best improvement was obtained after elimination of component s6 which can be interpreted as noise. Tab. 2 shows similarity between ICA components and the target measured with D factor (7). It should be noted that presented method can be applied to compare particular ICA components what is impossible to perform using correlation approach due to mutual statistical independence after ICA.

TABLE I. MODEL ERRORS MEASURED WITH MAPE.

Model	M1	M2	M3	M4	M5	M6
primary models	2.39	2.36	2.37	2.40	2.40	2.36
after ICA filtration	2.38	2.24	2.39	2.41	2.40	2.42

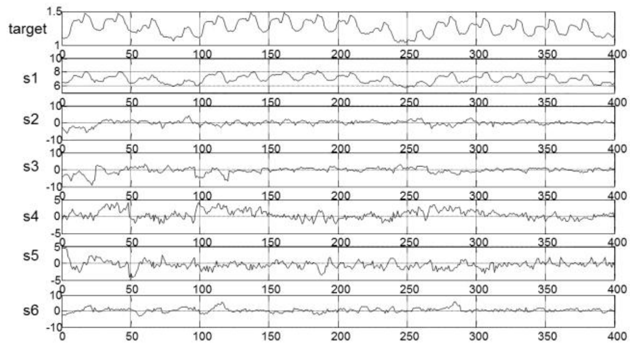

FIGURE 2. TARGET AND ICA COMPONENTS IN LOAD PREDICTION.

TABLE II. DIVERGENCE BETWEEN ICA COMPONENTS AND TARGET MEASURED WITH D.

s1	s2	s3	s4	s5	s6
0.1160	0.1519	0.1934	0.1959	0.1895	0.2075

V. CONCLUSIONS

We propose novel signals similarity measure based on the EGLD system. It can be applied both, for data with or without time structure, as well as for data which are mutually uncorrelated. The method is effective and it can be an alternative to correlation approach, especially for noise identification problems. The practical example confirmed validity of our approach.

ACKNOWLEDGMENT

The work was funded by the National Science Center based on decision number DEC-2011/03/B/HS4/05092.

REFERENCES

[1] S. V. Vaseghi, *Advanced signal processing and digital noise reduction*, Chichester, John Wiley & Sons, Stuttgart, 1997.

[2] R. Szupiluk, P. Wojewnik and T. Zabkowski, “Model Improvement by the Statistical Decomposition”. Lecture Notes in Computer Science, vol. 3070, 2004, pp. 1199-1204.

[3] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing. John Wiley, Chichester, 2002.

[4] A. Hyvarinen, J. Karhunen and E. Oja, Independent Component Analysis. John Wiley, New York, 2001.

[5] R. Szupiluk, P. Wojewnik and T. Zabkowski, “Noise detection for ensemble methods”. *Lecture Notes in Artificial Intelligence*, vol. 6113, 2010, pp. 471-478.

[6] Z.A. Karian, A. Zaven, and E.J. Dudewicz, Fitting statistical distributions: the Generalized Lambda Distribution and Generalized Bootstrap methods, Chapman & Hall, 2000.

[7] Z.A. Karian, E.J. Dudewicza and P. McDonald, “The extended generalized lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the Final Word on Moment Fits”. Communications in Statistics - Simulation and Computation, vol. 25, 1996, pp. 611–642.

[8] J. Karvanen, J. Eriksson and V. Koivunen, “Adaptive Score Functions for Maximum Likelihood ICA”. VLSI Signal Processing, vol. 32, 2002, pp. 83-92.