Volume XXII 2019
ISSUE no.1
MBNA Publishing House Constanta 2019

Scientific Bulletin of Naval Academy

SBNA PAPER • OPEN ACCESS

Hyperparameters optimisation for time varying signals

To cite this article: M. Rogobete and C. Racuciu, Scientific Bulletin of Naval Academy, Vol. XXII 2019, pg. 196-199.

Available online at www.anmb.ro

ISSN: 2392-8956; ISSN-L: 1454-864X

doi: 10.21279/1454-864X-19-I1-027
SBNA© 2019. This work is licensed under the CC BY-NC-SA 4.0 License

Hyperparameters optimisation for time varying signals

Marius Rogobete[1], Ciprian Racuciu[1]

[1] "Titu Maiorescu" University, Bucharest, Romania
E-mail: marius.rogobete@gmx.de

Abstract. For the most machine learning methods, for cyclo-stationary or even stochastic signals, the performance depends critically on hyperparameters. Moreover, the tuning of more hyperparameters based on the feedback of the performance model will leak an increasingly significant amount of information about the validation set into the model. Therefore, we propose in this research two classes of hyperparameters, a general class that makes the characterization of general signal curve and the second, a specific class that define special parameters connected to the phenomena type (e.g. sensor type).

1. Introduction

The time-dependent signals cannot be modeled using invariant time patterns. But the analysis and modeling of those processes should consider that systems which vary over time can be approximated by invariant linear time patterns over short intervals. For long periods, processes should be analyzed by time-dependent modeling.

A general feature of time-sensitive signals is that they contain non-stationary transient events. The characteristics of such non-stationary processes use time-varying parametric models. There are two main classes of methods for solving these models.

The first class uses recursive estimation of variable coefficients according to time, and the second is to constrain the evolution of the coefficient of linear or nonlinear combinations of basic functions with appropriate properties, called deterministic and stochastic regressive approaches.

2. Model Overview

There are many models able to be applied for this purpose. For time-varying processes an autoregressive model is defined by:

$$f(t) = \sum_{k=1}^{R} \mu_k (t) f(t-k) + \sum_{l=1}^{S} \rho_l(t) v(t-l) + err(t) \quad (1)$$

where

$\mu_k(t)$ *and* $\rho_l(t)$ are *ARX(R,S)* time-varying parameters

$v(t-l)$ is the input signal

R and S are the model orders (user choice values)

err(t) is the prediction error function

ARX function estimates parameters of ARX model using least squares using specific model structure:

$$f(t) + \mu_1 f(t-1) + \cdots + \mu_k f(t-k) = \rho_1(t) v(t-n_s) + \quad \ldots \quad + \rho_l(t) v(t-l-n_s+1) + err(t) \quad (2)$$

where

- n_s is number of input samples that occur before the input affects the output, also called the dead time in the system
- $f(t-1) \dots f(t-k)$ are previous outputs on which the current output depends
- $v(t-n_s) + \dots + \rho_l(t)v(t-l-n_s+1)$ previous and delayed inputs on which the current output depends

The assumption is the maximum value orders are time invariant.
Therefore, the method transforms time varying parameters in multi-wavelet basis function $\zeta_n(t)$, where *n=1,2, … ,L*.

After several substitutions results:

$$f(t) = \sum_{k=1}^{R} \sum_{n=1}^{L} \lambda_{k,n} f_n(t-k) + \sum_{l=1}^{S} \sum_{n=1}^{L} \sigma_{l,n} v_n(t-l) + err(t) \quad (3)$$

The model approximates the time varying parameters which could be assimilate as hyper-parameters after several adjustments. The approximation of $\mu_k(t)$ *and* $\rho_l(t)$ supposes to estimate same function values, as the numerical method request.

2.1. Function based on Multi-Wavelet

According to the wavelet theory, a scalar function that is scalar integrable $f \epsilon L^2(R)$ can be approximated with multiresolution wavelet decomposition:

$$f(x) = \sum_k \lambda_{l_0,k} \emptyset_{l_0,k}(x) + \sum_{j>j_0} \sum_k C\Psi_{j,k}(x) \quad (4)$$

where

$$\Psi_{l,k}(x) = 2^{\frac{l}{2}} \Psi(2^l x - k) \quad (5)$$

$$\Phi_{l,k}(x) = 2^{\frac{l}{2}} \Phi(2^l x - k) \quad (6)$$

are the wavelet family, shifted and dilated from mother wavelet Ψ and scaling function $\emptyset$.
The wavelet decomposition coefficients are $\lambda_{l_0,k}$ and $\sigma_{l,n}$.
A square integrable function *y(x)*, as multiresolution analysis theory demonstrates, can be approximated for a scale level resolution enough large as:

$$\Phi_{l,k}(x) = 2^{\frac{l}{2}} \Phi(2^l x - k) \quad (7)$$

The coefficients $\mu_k(t)$*and* $\rho_l(t)$ of (1) are approximated using functions of the family:

$$\mu_k(t) = \sum_{i \epsilon \Gamma_s} c_{k,i}^{(s)} \Phi_{k,i}^{(s)} \left(\frac{t}{N}\right) + \sum_{i \epsilon \Gamma_u} c_{k,i}^{(u)} \Phi_{k,i}^{(u)} \left(\frac{t}{N}\right) + \sum_{i \epsilon \Gamma_h} c_{k,i}^{(h)} \Phi_{k,i}^{(h)} \left(\frac{t}{N}\right)$$

$$\rho_l(t) = \sum_{i \epsilon \Gamma_s} d_{l,i}^{(s)} \Phi_i^{(s)} \left(\frac{t}{N}\right) + \sum_{i \epsilon \Gamma_u} d_{l,i}^{(u)} \Phi_i^{(u)} \left(\frac{t}{N}\right) + \sum_{i \epsilon \Gamma_h} d_{l,i}^{(h)} \Phi_i^{(h)} \left(\frac{t}{N}\right) \quad (8)$$

where :

$1 \leq s \leq u \leq h \leq 4, \quad t = 1,2, \dots, N$;
N is the observations number of the signal.

The experimental results for s=2, u=3 and h=4 recover in a right manner the time varying coefficients.

However, the first and second order B-splines which computes $\Phi_i^{(n)}$ are working well [11] as is evident from the time-dependent coefficients expression.

3. Time-varying Signal model

The sequence for processing SHA-512 method, suitable for a low-level implementation:
Assuming the time-varying model as:

$$f(t) = \mu_1(t) f(t-1) + \mu_2(t) f(t-2) + err(t) \tag{9}$$

the definitions for time-varying parameters are:

$$\mu_1(t) = 2\cos(2\pi y(t))$$
$$\mu_2(t) = -1, \quad t = 1, \dots, 1000 \tag{10}$$

where *y(t)* are values defined by the user.

4. Block Least Mean Square

The Least Mean Square algorithms, e.g. conventional and normalized , LMS are effectively solving the dynamic regression problem.
For example, the adaptive algorithm of BLMS:

$$err(mL+i) = s(mL+i) - \hat{z}^H (m)\, v\,(mL+i) \tag{11}$$

$$\hat{z}(m+1) = \hat{z}(m) + \delta \sum_{l=0}^{L-1} err^*(mL+j)\, v(mL+j) \tag{12}$$

where

$0 \le i \le L-1, \quad m = 1,2, \dots$
L is the length of the block;
δ is the step size parameter;
M is the number of taps;
err(t) is the error at time t.

The finite vector $\boldsymbol{V}(t)$:

$$\boldsymbol{V}(t) = \begin{bmatrix} v(t) \\ \dots \\ v(y-M+1) \end{bmatrix} \tag{13}$$

and

$$\widehat{\boldsymbol{Z}}(m) = \begin{bmatrix} \hat{z}_0(m) \\ \dots \\ \hat{z}_{M-1}(m) \end{bmatrix} \tag{14}$$

Finally, the complex BLPS algorithm is implementing using the FFT algorithm using method described in [11].

5. Conclusions

The ARX model presented in this work gives a good estimation for time-varying parameters. As they can be used subsequently to improve the estimation of the model parameters and are tuned for a better prediction of signal model then these parameters satisfy the hyperparameters characteristics (defined in Applied Predictive Modelling [1])

For the future work, should process both the standard deviations and the mean absolute error for different the parameter estimated. The statistically will confirm or not the better performance of the multi-wavelet basis function method. The classical normalized least mean square (NLMS) method could be also used for simulation and compared with the method based on multi-wavelet basis functions.

References

[1] M. Kuhn, K. Johnson, Applied Predictive Modeling , 2nd 2018 Edition, Springer, March 30, 2018), ISBN-10: 1461468485, ISBN-13: 978-1461468486

[2] Y. L. Singh and P. Gupta (2008), ECG to Individual Identification, Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference

[3] S. Safie, J. Soraghan and L. Petropoulakis (2011), ECG biometric authentication using Pulse Active Width (PAW), Biometric Measurements and Systems for Security and Medical Applications (BIOMS), 2011 IEEE Workshop on, pp. 1 - 6.

[4] S. K. Jagtap (2012), The impact of digital filtering to ECG analysis: Butterworth filter application, Communication, Information & Computing Technology (ICCICT), 2012 International Conference

[5] S.A., Billings, H.L., Wei, and M/A.Balikhin, , 'Generalized multiscale radial basis function networks', Neural Networks, 2007, 1081–1094

[6] Goodwin, G. C. and Sin K. S., Adaptive Filtering, Prediction, and Control, Englewood Cliffs, NJ: Prentice-Hall, 1984.

[7] Haykin, S., Adaptive Filter Theory, 4th ed., Pearson Education, 2002.

[8] Kokangul, A. (2008), 'A combination of deterministic and stochastic approaches to optimize bed capacity in a hospital unit', Computer Methods and Programs in Biomedicine, 90, 56-65.

[9] Mallat, S. G. (1989), 'A theory for multiresolution signal decomposition: the wavelet representation', IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 674-693.

[10] Nagumo, J. and Noda, A. (1976), 'A learning method for system identification', IEEE Trans. Automat. Cont., 12, 282-287.

[11] J. C. Lee, C. K. Un, "Block Realization of Multirate Adaptive Digital Filters", IEEE Trans. Acoust. Speech Signal Process., vol. 34, pp. 105-117, Feb. 1986.