

Supporting the Page-Hinkley Test with Empirical Mode Decomposition for Change Detection

Raquel Sebastião (orcid.org/0000-0001-5717-1415)
and José Maria Fernandes (orcid.org/0000-0002-0681-9354)

Institute of Electronics and Informatics Engineering of Aveiro (IEETA)
Department of Electronics, Telecommunications and Informatics (DETI)
University of Aveiro
3810-193 Aveiro, Portugal
{raquel.sebastiao,jfernan}@ua.pt

Abstract. In the dynamic scenarios faced nowadays, when handling non stationary data streams it is of utmost importance to perform change detection tests. In this work, we propose the Intrinsic Page Hinkley Test (iPHT), which enhances the Page Hinkley Test (PHT) eliminating the user-defined parameter (the allowed magnitude of change of the data that are not considered real distribution change of the data stream) by using the second order intrinsic mode function (IMF) which is a data dependent value reflecting the intrinsic data variation. In such way, the PHT change detection method is expected to be more robust and require less tunes. Furthermore, we extend the proposed iPHT to a blockwise approach. Computing the IMF over sliding windows, which is shown to be more responsive to changes and suitable for online settings. The iPHT is evaluated using artificial and real data, outperforming the PHT.

Keywords: Data streams; Page Hinkley Test; Empirical Mode Decomposition; Personalized Approach; Change Detection; Sliding Windows;

1 Introduction

Nowadays information is gathered as a continuous flow of data streams. A data stream is a sequence of information in the form of transient data that arrives continuously (possibly at varying times) and is potentially infinite. Data streams may not be strictly static as underlying properties evolve over time presenting non-stationary distributions and demanding the detection of changes. The survey [2] discusses several detection approaches, such as, ADaptive WINDowing, Drift Detection Method, Early Drift Detection Method.

The PHT [6] is a sequential analysis technique typically used for monitoring change detection in the average of a Gaussian signal. However, it is relatively robust in the face of non normal distributions and has been recently applied for concept drift detection in data streams [3]. In this work we propose the iPHT: a change detection approach enhancing the PHT through the elimination of the

user-defined parameter (the allowed magnitude of change) by using the second order IMF.

The IMF are the result of the Empirical Mode Decomposition (EMD) of a signal, which is an adaptive and efficient decomposition method that handles data on time domain [4]. Relying on time scale characteristics of the data, the EMD can successfully handle non stationary and nonlinear data [4]. Moreover, unlikely other signal processing decomposition methods, EMD makes no assumptions on the incoming data. This means that iPHT can be applied without any prior hypothesis on the data as both PHT and EMD can be applied over an arbitrary data stream without any constraints. Furthermore, we extend the iPHT to a blockwise approach. Computing the IMF over sliding windows advances a contribution to an online strategy.

2 Proposed Strategy

The Page-Hinkley Test requires an input parameter and an input threshold, which are defined by the user according to the data properties and application purposes. By supporting the PHT with the EMD we can replace the input parameter, which controls the allowed magnitude of change, by an IMF.

2.1 Page Hinkley Test

The two-sided PHT [6, 3] detects both increases and decreases in the mean of a sequence. For testing online changes it runs two tests in parallel, considering a cumulative variable defined as the accumulated difference between the observed values and their mean until the current moment.

2.2 Empirical Mode Decomposition

The EMD is being successfully used in real-time applications, such as biomedical problems [7] and image processing [5], among others. There are several free implementations of EMD available: the one used in this work was developed by Alan Tan ¹.

The EMD consists of successively decomposing the original data $x(t)$ into IMF $c_i(t), i = 1, \dots, n$ and into the monotonic residual $r(t)$. Once the first IMF is removed from the original data, the procedure is successively applied to the residual. This process will decompose the original data into the highest frequency component (c_1) to the lowest frequency component (c_n), until the residual $r(t)$ is a monotonic function from which no more IMF can be extracted:

$$x(t) = \sum_{i=1}^n c_i(t) + r(t), \text{ where } n \text{ is the number of IMF.}$$

¹ available at <http://www.mathworks.com/matlabcentral/fileexchange/19681-hilbert-huang-transform> (accessed in March 17th 2016)

2.3 Intrinsic Page Hinkley Test

The lower IMF orders are related with higher frequency of the data and higher IMF orders describe the data baseline wander. We are interested in using the lower IMF orders to establish the magnitude of the changes that are allowed. In this context the PHT is supported through the EMD, by replacing the δ parameter with an IMF. Hence we are losing the user-defined parameter of the PHT by using a component that will depend on the data itself. The iPHT tests are the following:

For increase cases: $iU_0 = 0$ $iU_T = (iU_{T-1} + x_T - \bar{x}_T - IMF_T)$ $m_T = \min(iU_t, t = 1 \dots T)$ $iPH_U = iU_T - m_T$ If $iPH_U > \lambda$: change detected	For decrease cases: $iL_0 = 0$ $iL_T = (iL_{T-1} + x_T - \bar{x}_T + IMF_T)$ $M_T = \max(iL_t, t = 1 \dots T)$ $iPH_L = M_T - iL_T$ If $iPH_L > \lambda$: change detected
---	---

(\bar{x}_T is the mean until the current sample.)

The threshold λ depends on the admissible false alarm rate. Increasing λ entails fewer false alarms, but might miss or delay changes. The order of the IMF used to replace δ depends on the relation between the several IMF orders and the original data. The information shared decrease with the IMF order: the first IMF shares more information with the original data. As δ corresponds to the magnitude of the changes that are allowed, the original data and the IMF order must be positive related. On the other side, they must not share too much information. Therefore, the IMF order will be studied, in the section 3, evaluating the covariance and the Spearman correlation between the data and several IMF orders.

The EMD requires the knowledge of the entire data stream to compute the IMF. To overcome this drawback, we propose a blockwise approach for the extraction of the IMF: the IMF are computed only with the data that is inside a sliding window. The main difficulty is how to select the appropriate length of the window, establishing a trade-off between good stability for the computation of the IMF and good adaptability to evolving scenarios.

3 Results and Discussion

3.1 Artificial Data Design

The artificial data sets were generated according to normal distributions, varying only the mean parameter and setting the standard deviation parameter equal to 1. Each data stream consists of 2 parts, each of with size $N = 5000$. Different changes were simulated by varying among 3 levels of magnitude and 3 rates (or speed) of change, obtaining a total of 9 types of changes. For high, medium and low magnitude levels, μ changed from 0 to 5, 3 and 2, respectively. The rates were defined assuming that the samples from the first part are from the old distribution and the $5000 - ChangeLength$ last samples are from the new

distribution. The rates of change were defined as high, medium and low, for a *ChangeLength* of 1, $0.25 * 5000$ and $0.5 * 5000$, respectively.

For each type of changes, 10 data streams were generated with different seeds. Two data sets, with similar characteristics, were generated, one to train and adjust the parameters and another to test and evaluate the proposed iPHT.

Choosing the Order of the IMF. Figure 1 shows the covariance between the original data and the correspondent five first IMF orders (average with error bars for 10 runs on data generated with different seeds). The Spearman correlation presented a similar behavior, and for both measures, the higher values are obtained for the IMF_1 . When comparing the original data with the IMF_1 and with the remain IMFs, it can be observed an abrupt decay and a repeated pattern along the different changes. Considering these results on the shared information, IMF_2 was chosen to replace the user-defined parameter δ of the PHT.

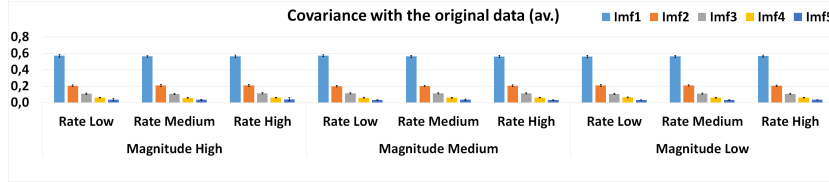


Fig. 1. Covariance between the data and the five first IMF orders

Sensibility of the λ Threshold. The λ change detection threshold was adjusted using the training set and evaluating the total number of false alarms (FA), the total number of missed detections (MD) and the detection delay time (DDT). The results are shown in Table 1. Independently of the λ , the proposed iPHT detects all the changes. This is relevant as the MD are a major concern: missing a change, besides inducing a bias on the model, would prevent the system to perform the appropriate action. The number of FA (desirable to be null) can be controlled by the change detection threshold. It can be observed the absence of FA for λ greater than 300. Considering that the DDT was quite similar for $\lambda = 400$ and for $\lambda = 500$, and following a conservative approach, we decided on establishing the $\lambda = 500$. As expected, the DDT when using $\lambda = 500$ is greater for changes with low rate and magnitude.

Comparison with the PHT. This experiment assesses the advantage of iPHT over PHT. For PHT, the λ was set to 500 and δ varied from 0.0001 to 1. Independently of the δ , both approaches detected all the changes without FA. The DDT of the iPHT was 722 ± 71 (average and standard deviation of 10 runs for the 9 types of changes). For the PHT, as the δ varied from 0.0001 to 1, the DDT

Table 1. DDT when varying the λ : average and standard deviation of 10 runs. In parenthesis is the number of runs, if any, where the iPHT presents FA or MD.

		λ			
Magnitude	Rate	100	300	500	700
High	Low	341 \pm 174 (14;0)	628 \pm 118 (0;0)	901 \pm 75(0;0)	1112 \pm 55 (0;0)
	Medium	225 \pm 66 (18;0)	456 \pm 98 (0;0)	647 \pm 65 (0;0)	789 \pm 53 (0;0)
	High	15 \pm 4 (19;0)	45 \pm 13 (1;0)	80 \pm 19 (0;0)	120 \pm 19 (0;0)
Medium	Low	476 \pm 139 (16;0)	860 \pm 154 (0;0)	1201 \pm 127 (0;0)	1592 \pm 113 (0;0)
	Medium	399 \pm 94 (15;0)	641 \pm 79 (0;0)	869 \pm 54 (0;0)	1056 \pm 47 (0;0)
	High	23 \pm 8 (18;0)	60 \pm 28 (0;0)	127 \pm 30 (0;0)	196 \pm 29 (0;0)
Low	Low	611 \pm 176 (20;0)	1131 \pm 99 (0;0)	1532 \pm 92 (0;0)	1849 \pm 71 (0;0)
	Medium	398 \pm 145 (19;0)	735 \pm 139 (0;0)	1027 \pm 105 (0;0)	1263 \pm 87 (0;0)
	High	36 \pm 17 (15;0)	108 \pm 30 (0;0)	210 \pm 31 (0;0)	314 \pm 30 (0;0)

varied from 723 ± 69 to 1818 ± 31 (average and standard deviation of 10 runs for the 9 types of changes). These results show that the iPHT and PHT presented similar performance, but the iPHT has the advantage of adjusting only the change detection threshold.

Extension to Online Processing. This experiment evaluates the relation between DDT and the window length (from 100 to 1000) used to compute the IMF of the iPHT. A large window, while providing more data to compute more accurate IMF, will contribute to higher delays, since the testing for changes is performed within large time steps. A smaller window will, at a first glance, allow to detect changes earlier. However, it compromises the stability of the IMF, which may preclude the detection of changes.

The DDT is shown in Figure 2. This experiment stands out that choosing correctly the length of the sliding window, computing the IMF in blockwise does not impair the detection delay time. Indeed, using a sliding window of length 100 the DDT is similar when computing the IMF with the entire data stream. It also can be observed that, an increase of the length of the sliding window leads to higher DDT. One result that must be underlined is that using the entire data stream or using sliding windows, the iPHT does not presented FA nor MD. It can also be observed that the advantage of using a smaller sliding window is strengthened when detecting abrupt changes, which is explained by the ability of a window with small length to better catching abrupt than smooth changes.

Robustness to Detect Changes in the Presence of Noise. In this experiment noisy data was generated by adding different percentages of pink noise to the test data set. Figure 3 shows the obtained results by varying the amount of noise from 0% to 50%. The DDT was obtained when performing the iPHT over a sliding window of length 100. The iPHT presents a regular performance along the different amounts of noise, which are similar to the DDT in the absence of noise. This experiment sustains the argument that the iPHT, when performed in blockwise with a sliding window of length 100, is robust against noise while effectively detects changes in the data. It should be noted that, even in the presence of great amount of noise, the proposed iPHT did not miss any change.

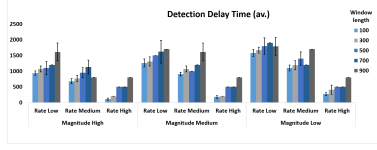


Fig. 2. DDT for different lengths of the sliding window (average and standard deviation of 10 runs of the training set)

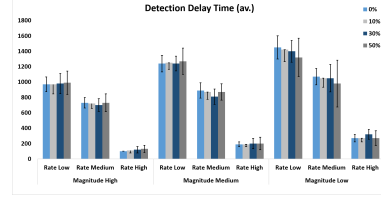


Fig. 3. DDT (average of 10 runs) of the iPHT with different amounts of noise and computed over a sliding window of length 100 (blockwise)

Although, with an amount of 50% of noise, it presents a FA for changes with high magnitude and low rate and with medium magnitude and rate.

3.2 Industrial Data

This industrial data set was obtained within the scope of the work presented in [1], with the objective of designing different machine learning classification methods for predicting surface roughness in high-speed machining. Data was obtained by performing tests in a Kondia HS1000 machining center equipped with a Siemens 840D open-architecture CNC. The tests were done with different cutting parameters, using sensors for registry vibration and cutting forces. For change detection purposes, the measurements of the cutting speed on X axes from 7 tests were joined sequentially in order to have only one data set with 6 changes with different magnitudes and sudden and low rates, as shown in Figure 4. For PHT, δ was set to 1 and the iPHT used the IMF_2 . The λ was set to 2000 (for both tests). Although presenting 8 FA against 7 of the PHT, the iPHT detects all the 6 changes with mean DDT of 3512 against only 5 detected changes with mean DDT of 8458 of the PHT.

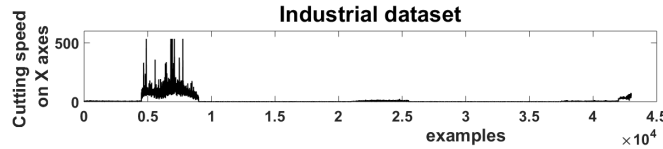


Fig. 4. The cutting speed on X axes from 7 tests sequentially joined

4 Conclusions

This paper proposes the iPHT, a test for change detection through the enhancement of the PHT with the IMF, eliminating the user-defined parameter which controls the allowed magnitude of change.

The results over the designed experiments have shown that the iPHT is a feasible approach to detect changes under different evolving scenarios, even in the presence of noise. The obtained results also sustain its application over a sliding window, which is an advantage that leads future research on this approach for online settings. While presenting similar performance to the PHT (regarding DDT, FA and MD), the iPHT has the advantage of adjusting only the change detection threshold.

With these properties, the iPHT has potential for detecting changes in several interesting scenarios, like internet of things and sensor networks context, for human and environmental monitoring. Our next step is to apply the iPHT in a physiological monitoring setting over heart rate. In such cases, a change can be seen as a biomarker for emotion changes and changes in fatigue or stressful conditions. Furthermore, iPHT must be evaluated and compared to other methods in terms of computational time and memory usage.

Acknowledgments. This work was supported by the Portuguese Science Foundation (FCT) through national funds, and co-funded by the FEDER, within the PT2020 Partnership Agreement and COMPETE2020 under projects IEETA (UID/CEC/00127/2013) and VR2market (funded by the CMU Portugal program, CMUP-ERI/FIA/0031/2013). Raquel Sebastião acknowledges her Post-Doc grant (BPD/UI62/6777/2015).

References

1. Correa, M., Bielza, C., Pamies-Teixeira, J.: Comparison of bayesian networks and artificial neural networks for quality detection in a machining process. *Expert Syst. Appl.* **36**(3), 7270–7279 (2009). DOI 10.1016/j.eswa.2008.09.024
2. Gama, J., et al.: A survey on concept drift adaptation. *ACM Computing Surveys* **46**(4) (2014). DOI 10.1145/2523813
3. Gama, J., Sebastião, R., Rodrigues, P.: On evaluating stream learning algorithms. *Machine Learning* **90**(3), 317–346 (2013). DOI 10.1007/s10994-012-5320-9
4. Huang, N.E., et al.: The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **454**(1971), 903–995 (1998). DOI 10.1098/rspa.1998.0193
5. Linderherd, A.: Variable sampling of the empirical mode decomposition of two-dimensional signals. *International Journal of Wavelets, Multiresolution and Information Processing* **03**(03), 435–452 (2005). DOI 10.1142/S0219691305000932
6. Page, E.S.: Continuous inspection schemes. *Biometrika* **41**(1-2), 100–115 (1954). DOI 10.1093/biomet/41.1-2.100
7. Santillan-Guzman, A., Fischer, M., Heute, U., Schmidt, G.: Real-time empirical mode decomposition for eeg signal enhancement. In: *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pp. 1–5 (2013)