Physica A 471 (2017) 499–511

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Structural break detection method based on the Adaptive Regression Splines technique

Daniel Kucharczyk[a], Agnieszka Wyłomańska[b,*], Radosław Zimroz[b]

[a] Faculty of Pure and Applied Mathematics, Wroclaw University of Science and Technology, Janiszewskiego 14a, 50-370 Wrocław, Poland

[b] KGHM CUPRUM Ltd., CBR, Sikorskiego 2-8, 53-659 Wrocław, Poland

HIGHLIGHTS

- The new technique of structural break point detection is proposed.
- We propose the statistical test for testing two-regimes behaviour.
- The technique is based on the Adaptive Regression Splines method.
- We check the proposed methodology to simulated and real technical data.

ARTICLE INFO

Article history:
Received 14 May 2015
Received in revised form 16 December 2015
Available online 27 December 2016

Keywords:
Segmentation
Median
Statistical test
Structural break point detection

ABSTRACT

For many real data, long term observation consists of different processes that coexist or occur one after the other. Those processes very often exhibit different statistical properties and thus before the further analysis the observed data should be segmented. This problem one can find in different applications and therefore new segmentation techniques have been appeared in the literature during last years. In this paper we propose a new method of time series segmentation, i.e. extraction from the analysed vector of observations homogeneous parts with similar behaviour. This method is based on the absolute deviation about the median of the signal and is an extension of the previously proposed techniques also based on the simple statistics. In this paper we introduce the method of structural break point detection which is based on the Adaptive Regression Splines technique, one of the form of regression analysis. Moreover we propose also the statistical test which allows testing hypothesis of behaviour related to different regimes. First, the methodology we apply to the simulated signals with different distributions in order to show the effectiveness of the new technique. Next, in the application part we analyse the real data set that represents the vibration signal from a heavy duty crusher used in a mineral processing plant.

1. Introduction

Physical variables observed/measured by advanced data acquisition systems in real world applications are often processed, modelled, and analysed in order to extract information about the process. In many cases, long term observation

* Corresponding author.
E-mail addresses: daniel.kucharczyk@pwr.edu.pl (D. Kucharczyk), awylomanska@cuprum.wroc.pl (A. Wyłomańska), rzimroz@cuprum.wroc.pl (R. Zimroz).

http://dx.doi.org/10.1016/j.physa.2016.12.011
0378-4371/

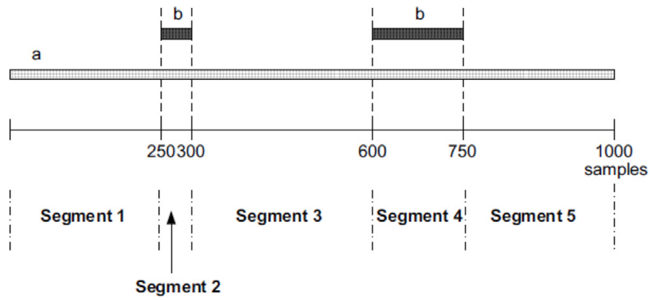

Fig. 1. The model of the signal that is a mixture of different segments related to different processes.
Source: [1].

(long might mean very different period depends on application) in fact consists of different processes that coexist or occur one after the other. There are many examples that during single observation Process A reveals very different properties than Process B and they cannot be processed/analysed with the same tools (simple example is switching from stationary to non-stationary signal). Less radical case is that two segments might have the same type of model, but with different order and values of parameters. It leads to conclusion that very first step in experimental data analysis is signal segmentation, i.e. dividing the raw observation into smaller pieces (segments) with homogeneous properties. Note that Process A or Process B might be single process or mixture of processes.

From signal processing perspective, one of the most fundamental reason for segmentation is finding locally stationary segments in non-stationary signals. In such a context the segmentation is a problem of identifying the time instants at which the statistics of the observed signal changes. Signal segmentation is also generalized and considered in the context of signal detection and localization by means of recognition and identification of the time of appearance of event that differs from "normal" signal, [1]. It could be generalized to a model proposed in [1] presented in Fig. 1.

The basis (criterion) for signal segmentation might be different in various domains, contexts and applications. Change of process from Process A to Process B might be related to appearance of seismic event, machine regime switching, fault in the system, financial crash on stock market, natural radiation, wind behaviour, etc. A key issue is to find proper description of processes in order to highlight the difference between them. In some cases, the difference between processes is visually identifiable, but often advanced processing is required to find breakpoint. Signal segmentation might be done in different manners. Some of the methods are based on the statistical properties of raw time series [2,3] or its modelling and analysis of model residuals [1,4,5]. The second group of segmentation methods takes under consideration not the raw signal but its transformation to other domain, like spectral [6], cepstral [7], time–frequency [8–10], etc. In the literature one can also find segmentation techniques based on the raw time series decomposition such as empirical mode or wavelets decomposition [11–13].

Signal segmentation has been applied in many areas. It is especially crucial in condition monitoring (to isolate shocks related to damage) [9,13], machine performance analysis (to find when machine operates under overloading, idle mode etc.) [14], experimental physics [2,15,16], biomedical signals (like ECG signals) [7,17–22], speech analysis (automatic speech recognition and understanding) [23–25], econometrics [26,27] and seismic signal segmentation [8,28–31]. The other areas where the segmentation problem appears one can find in [32–34].

In this paper we propose a novel segmentation technique. This method is an extension of the algorithm proposed in [2], where the main statistics used to segmentation was based on the empirical second moment of raw data. An approach proposed here is a step forward in this field. The introduced method is based on the behaviour of the average absolute deviation about the empirical median of given time series. For the data where some statistical properties (expressed mostly by the scale parameter) change over time, the simple statistics applied here changes its behaviour which leads to structural break detection. The proposed method is much more sensitive for such cases where the change point is not visually identifiable in the contrast to the method presented in [2]. Moreover it can be used to different raw signals without assumption of their distributions which is the main advantage. On the one hand we analyse visually the mentioned statistics but on the other hand the strict method of change point detection is introduced. The method uses the Adaptive Regression Splines technique which is a form of regression analysis. It is a non-parametric regression technique and can be seen as an extension of linear systems that automatically models non-linearities and interactions between variables [35]. Additionally, we propose a statistical test which allows for testing if given data satisfy the behaviour presented in Fig. 1. Because in the analysis we consider the case where only one change point exists then the time series $\{Z_i\}$ which is tested here, similar as in [2], can be written in the following mathematical form

$$X_i = \begin{cases} T & \text{for} \quad i \le l, \\ Y & \text{for} \quad i > l, \end{cases} \tag{1}$$

where l is a structural break point and T and Y are independent random variables with cumulative distribution functions $G(\cdot)$ and $H(\cdot)$, respectively.