APPLICATION OF JENSEN-SHANNON DIVERGENCE IN SMART GRIDS

Istvan Pinter[1*], Lorant Kovacs[1], Andras Olah[1,2], Rajmund Drenyovszki[1,3], David Tisza[1,2] and Kalman Tornai[1,2]

[1]Kecskemet College, Faculty of Mechanical Engineering and Automation,
Izsaki ut 10, H-6000 Kecskemet, HUNGARY
[2]Pazmany Peter Catholic University, Faculty of Information Technology
Prater utca 50/a, H-1083 Budapest, HUNGARY
[3]University of Pannonia, Department of Computer Science and Systems Technology,
Egyetem u. 10, H-8200 Veszprem, HUNGARY
** Corresponding author e-mail: pinter.istvan@gamf.kefo.hu*

Abstract
A lot of problem in smart grids, such as automatic detection of appliance categories, customer categories or anomaly detection are rooting in segmentation of non-stationary time series into stationary subsequences. This problem is often referred to as Change Point Detection (CPT). In this paper an information theoretic measure: the Jensen-Shannon Divergence will be shown as promising candidate tool for solving CPT tasks in smart grids.

Keywords:
Smart grid, Change Point Detection, Jensen-Shannon Divergence

1. Introduction

Smart grid is new vision of future evolution of electricity systems. In contrast to the present grid, smart grid is adaptively managing the balance (supply and demand) of the system, and can handle the challenge of incorporation of renewable energy resources. These features are based on a two-way communication system and a SCADA (Supervisory Control and Data Acquisition) system from the top to the bottom.

The new system components imply new applications that require new signal processing methods as well. One of the relevant problems is the consumption management of an electricity customer that requires handling of traditional appliances as well. The smart meter has to have information about the type of the device connected to the grid even in the case of being not able to communicate its statistical parameters. Furthermore, the smart meter has to recognize the change of the appliance to another one. As a result the task of categorization is twofold: at first stationary segments in the time series of consumption data should be identified, and then a category detection algorithm should be performed. This is quite the same if the task is to categorize the customer in order to predict its behavior. Another task when stationary segments have to be detected is outlier detection, i.e. finding not reliable data in a huge dataset.

From a mathematical point of view the above mentioned problems have a common component: non-stationary time series has to be segmented into stationary ones, or by other words, change points in the signal should be identified. In this paper an information theoretic measure, the Jensen-Shannon Divergence (JSD) will be used for Change Point Detection (CPT) in a smart grid environment. It will be shown that the method is capable of detecting seasonality and category change in the time series.

The paper will be organized as follows: In Section 2. the JSD and its properties will be introduced. In Section 3. the generalized JSD and JSD-contour will be explained as a tool for CPT. In Section 4. numerical results will be presented, finally in Chapter 5. conclusions will be drawn.

2. The Jensen-Shannon Divergence

The JSD is an information theoretical measure to determine the distance of two discrete probability distributions [1,2,]. Let's denote $P = \{p_1, p_2, \dots, p_i, \dots p_K\}$ and $Q = \{q_1, q_2, \dots q_i, \dots q_K\}$ two discrete random distributions where p_i and q_i indicate the probability of the occurance of ith symbol in P and Q, respectively. In order to introduce the definition of the $JSD(P,Q)$, one needs the average distribution

$$A = \frac{P+Q}{2} = \{a_1, a_2, \dots, a_\mathrm{i}, \dots a_K\}, \quad (1)$$

where $a_i = \frac{p_i+q_i}{2}$. The JSD can be expressed by both the Kullback-Leibler Divergence and by the Shannon-entropy. Let's denote $KLD(P,Q) = \sum_{k=1}^{K} p_k log \frac{p_k}{q_k}$ the Kullback-Leibler Divergence [1], by which the JSD can be expressed as

$$JSD(P,Q) = \frac{D_{KL}(P,A) + D_{KL}(Q,A)}{2} \quad (2)$$

Note that $KLD(P,Q) \neq KLD(Q,P)$, but $JSD(P,Q) = JSD(Q,P)$, and $JSD(P,Q) \geq 0$ furthermore $\sqrt{JSD(P,Q)}$ satisfies $\sqrt{JSD(P,Q)} \leq \sqrt{JSD(P,R)} +$

$\sqrt{JSD(R,Q)}$, hence it can be used as measure of distance.
The JSD can be expressed by the terms of the Shannon-entropy as well, gives as

$$JSD(P,Q) = H(A) - \frac{H(P)+H(Q)}{2} \qquad (3)$$

where $H(P) = -\sum_{k=1}^{K} p_k log p_k$. In the following section expression (3) is more convenient for our purposes, hence it will be used instead of (2).

3. The generalized JSD and the JSD-contour for CPT detection

The average distribution A in (1) can be seen as an averaged sum of the distributions P and Q, where the weights are 0.5. Let's introduce the weights $\omega_A, \omega_B > 0$, so that $\omega_A + \omega_B = 1$, than the average of the distribution is $A = \omega_P P + \omega_Q Q$ and using this new definition of the average distribution we arrive at the generalized JSD as

$$JSD(P,Q) = H(\omega_P P + \omega_Q Q) - [\omega_P H(P) + \omega_Q H(Q)] \qquad (4)$$

The purpose of this generalization is reducing the negative effect of unreliable estimates of the true statistical values by empirical ones using short segments.
The generalized JSD can be used for the segmentation of symbol-sequences introducing the idea of JSD-contour. Let us assume a symbol sequence $S = [s_1, s_2, ..., s_N] = [S_1, S_2]$ where $S_1 = [s_1, s_2, ..., s_{N_1}]$ and $S_2 = [s_1, s_2, ..., s_{N_2}]$ and the symbols take their values from a finite alphabet $s_i \in \{\alpha_1, ..., \alpha_k\}$. S_1 and S_2 will be referred to as left and right-series, respectively. Note that S_1 and S_2 are the real subsequences of the whole sequence S; our aim is to find the change point N_1.
Let us denote an arbitrary segmentation of the whole sequence S by $S_l(n)$ and $S_r(n)$ where n stands for the segmentation point, and $n = \{1, 2, ..., N-1\}$. The relative frequencies of the symbols in a subsequence can be described by a K-element vector $P(n) = [p_1, p_2, ..., p_K]$ where $p_i = \frac{n_i}{n}$ and n_i is the number of occurrences of symbol α_i, and $Q(n) = [q_1, q_2, ..., q_K]$. The JSD-contour is a sequence of the generalized JSDs for all the possible segmentations $n = \{1, 2, ..., N-1\}$.

$$JSDC(S) = [JSD(1), JSD(2), ..., JSD(N-1)]$$

where

$$JSD(n) = H(\omega_P P(n) + \omega_Q Q(n)) - [\omega_P H(P(n)) + \omega_Q H(Q(n))]$$

In [3] it has been proven, that the sharpest peak in the JSD-contour can be observed, if the weighting factors are set as

$$\omega_P = n/N \text{ and } \omega_Q = \frac{N-n}{N}, \qquad (5)$$

respectively. This choice is reasonable to suppress the uncertainty of the relative frequencies in the case of short subsequences. However there is further reason: In the case of (5) the distribution of the JSD(n) (since it is a random variable because of the usage of relative frequencies) can be calculated, and can be used to evaluate the distance of the distribution of two subsequences.
In Figure 1. results of a numerical example can be seen. A sequence of length of 2500 has been generated by concatenating of two subsequences: the left sequence consists of 500 while the right sequence of 2000 symbols. The distributions of S_1 and S_2 were $P = [0.3, 0.7]$ and $Q = [0.5, 0.5]$, respectively (i.e. binary example). From these distributions 1000 sequences were generated. For each the JSD-contour has been calculated which are depicted by blue curves. The mean of the JSD-contours is indicated by red.

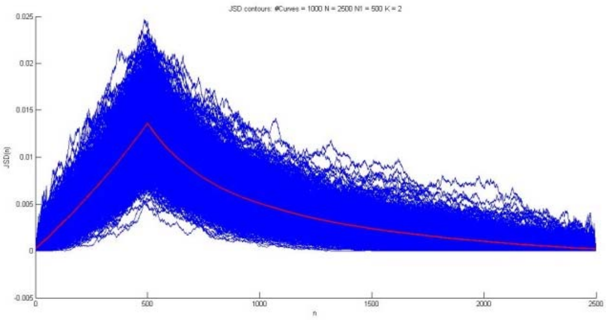

Figure 1. JSD-contour of 1000 randomly generated binary sequences (blue) and the mean of the contours (red)

In Figure 2. the distribution of the detected CPT by the JSD-contour can be seen for the same example. Note, that the true CPT of $n = 500$ is estimated with the largest probability and the variance is relatively low.

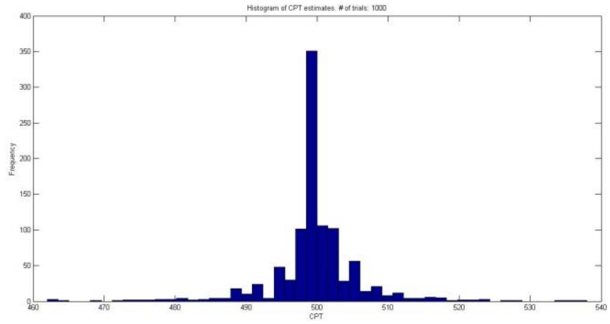

Figure 2. The histogram of the index of the maximum of the JSD-contour

4. Numerical results

In this section we test the performance of the JSD-contour based segmentation technique on real consumption data of households. In the numerical experiments the following assumptions has been made

- measurements are available from Smart Meters
- the non-stationary data series are composed from stationary sub-sequences
- the numerical measurement values can be mapped to a usable alphabet.

The original measurements were sampled by 12-bit AD-converter, resulting in an alphabet of $\mathbf{2^{12}} = \mathbf{4096}$ elements. We recognized that the JSD-contour based CPT detection does not perform satisfactory in the case of large alphabets. Therefore re-quantization of the measured sequence has been performed using a histogram-based method with histogram bin-width $\boldsymbol{h}$ and for number of bins $\boldsymbol{K}$ as

$$h = \frac{max - min}{\sqrt[3]{N}}$$

$$K = \left\lceil \sqrt[3]{N} \right\rceil$$

As a source of real data we used the UCI Machine Learning Repository [5]. In the database there are seven time-series available from December 2006 to November 2010. The sampling interval is 1 minute, the registered measures are global active power, global reactive power, voltage, intensity, sub-metering 1 (kitchen), sub-metering 2 (laundry), sub-metering 3 (water heater, air-conditioner).

As a test sequence for detecting the CPT we used artificially concatenated segments. Two scenarios has been tested:

- seasonal change (using winter-summer data of global active power);
- load profile change (using the data of sub-meters).

In Figure 3. and 4. The concatenated sequence can be seen with seasonal change point. The vertical magenta bar indicates the (artificial) change point. The JSD-contour is the red curve. In both cases the maximum of the JSD-contour seems to be a good estimation of the change points. The relative errors of the true change points and the approximated one are below 0.005. The relative error of the change point detection is measured by

$$e = \frac{|n_0 - \hat{n}_0|}{\hat{n}_0},$$

where n_0 is the artificial CPT, and $\hat{n}_0$ is the estimate of it by the maximum of the JSD-contour.

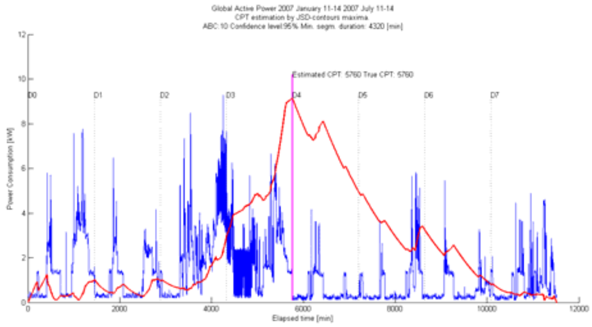

Figure 3. Concatenated seasonal data: left sequence January 11-14 2007, right sequence July 11-14 2007 (global active power)

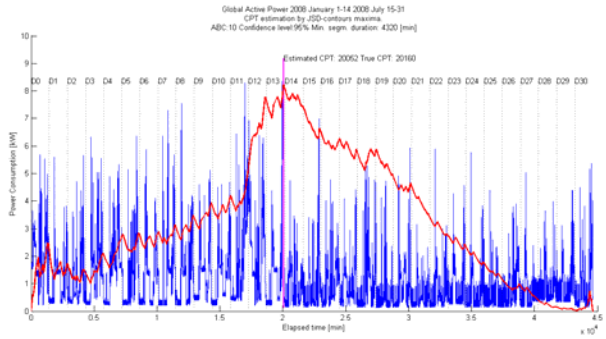

Figure 4. Concatenated seasonal data: left sequence January 1-14 2008, right sequence July 15-31 2008 (global active power)

In Figure 5.-7. The concatenated sequence can be seen with the change point coming from different sub-metering devices (but from the same time interval). In these cases the relative errors are below 0.03.

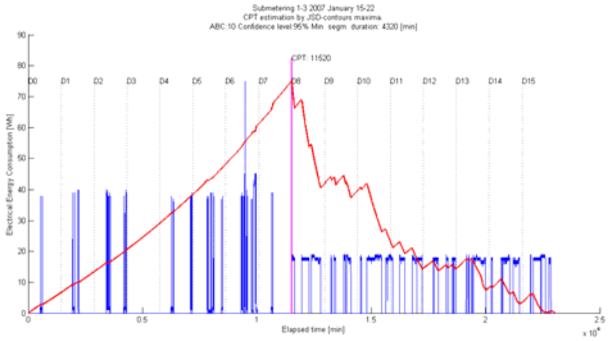

Figure 5. Concatenated home appliance data: left sequence: January 15-22 2007 ,sub-metering 1, right sequence: same days, sub-metering 3 (energy consumption)

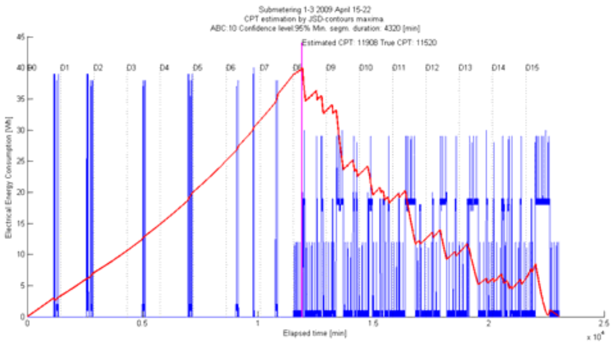

Figure 6. Concatenated home appliance data: left sequence: April 15-22 2009, sub-metering 1, right sequence: same days, sub-metering 3 (energy consumption)

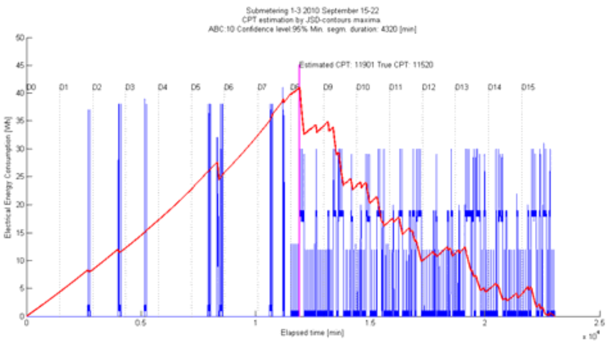

Figure 7. Concatenated home appliance data: left sequence: September 15-22 2010, sub-metering 1, right sequence: same days, sub-metering 3 (energy consumption)

5. Conclusions

This paper introduces the Jensen-Shannon Divergence Contour, which is capable to perform the segmentation of non-stationary sequences by an off-line manner. Extensive numerical analysis of real consumption data for the case of seasonal and category of appliance change point has been done. The results prove that the method can solve CPT problems in a smart grid environment.

6. Acknowledgement

This research and publication have been supported by the European Union and Hungary and co-financed by the European Social Fund through the project TÁMOP-4.2.2.C-11/1/KONV-2012-0004: National Research Center for the Development and Market Introduction of Advanced Information and Communication Technologies. This source of support is gratefully acknowledged.

7. References

[1] I. Csiszár, J. Körner: „Information Theory - Coding Theorems for Discrete Memoryless Systems“, Akadémiai Kiadó, Budapest, 1981.
[2] I. Deák: „Random Number Generators and Simulation“ Akadémiai Kiadó, Budapest, 1990.
[3] I. Grosse, P. Bernaola-Galván, P. Carpena, R. Román-Holdán, J. Oliver and H. E. Stanley: „Analysis of symbolic sequences using the Jensen-Shannon divergence“, Physical Review E, Volume 65, 2002, 041905 pp. 65-81.
[4] Lucien Birge and Yves Rozenholc: „How many bins should be put in a regular histogram', ESAIM: Probability and Statistics February 2006, Vol. 10. pp. 24-45.
[5] Bache, K., Lichman, M.: „UCI Machine Learning Repository“, Irvine, University of California, School of Information and Computer Science, [http:/archive.ics.uci.edu/ml], 2013