Time Series Segmentation through Automatic Feature Learning

ABSTRACT

Traditional changepoint detection methods look for boundaries that are defined as abrupt variations in the generative parameters of a data sequence. However, we observe that *breakpoints* – human-specified boundaries – occur on more subtle boundaries that are non-trivial to detect with these statistical methods. In this work, we propose a new semi-supervised approach, based on deep learning, that outperforms existing techniques and learns the more subtle *breakpoint* boundaries with a high accuracy. Through extensive experiments on various real-world data sets – including human-activity sensing data, speech signals, and electroencephalogram (EEG) activity traces – we demonstrate the effectiveness of our algorithm for practical applications. Furthermore, we show that our approach achieves significantly better performance than previous methods.

CCS CONCEPTS

• **Information systems → Information systems applications**; • **Computing methodologies → Machine learning approaches**; **Machine learning algorithms**;

KEYWORDS

Deep Learning, Automatic Segmentation, Changepoint detection

ACM Reference Format:

. 2018. Time Series Segmentation through Automatic Feature Learning. In *Proceedings of ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN 2018).* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Changepoints are abrupt changes in the trends of a data sequence. Bayesian techniques [1–3, 6, 7] discover these by looking for changes in the parameters of the distribution that generates the sequence. Considering the generality of this problem, many techniques exist in the literature [1–3, 6, 7]. These techniques attempt to capture the generative process through a pre-determined model and aim to look for changes in the parameters of the generative process. For learning expert-specified boundaries, these models often fail – since such changes are not easily captured by a pre-specified model of the generative process and changepoints do not typically fall along parameter-shift boundaries.

The changepoints specified by experts often arise when the state transition is a function of latent temporal properties of an underlying process that are difficult to capture in a pre-specified model. These rules are encoded as latent features in these traces and practically impossible to detect with the existing generative-model based changepoint detection algorithms. To differentiate from these statistically-detectable changepoints, we hereafter refer to the human-specified changepoints as *breakpoints*. Furthermore, we propose a novel algorithm that uses deep learning techniques to detect *breakpoints* without any prior assumptions about the generative process. Our method automatically learns the features that are most useful to represent the input data and thus can discover hidden structure in real-world time series data. Note that our approach has broad applicability for general changepoint detection even outside of its application to breakpoint detection as considered in this paper.

Figure 1 shows a comparison of our approach to a Bayesian changepoint detection technique from the literature [1]. Our technique requires careful parameter tuning however, in this work we discovered a simple, reliable tuning heuristic that *works for all the data sets we tested*. We discuss this technique in later sections.

2 BREAKPOINT DETECTION

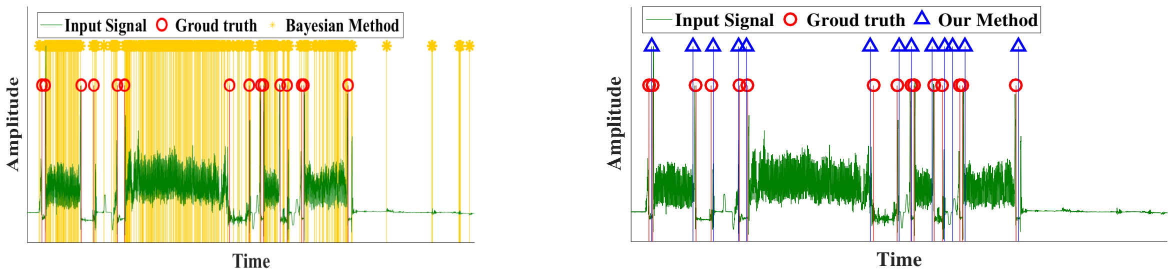
After we extract the representative features of the input data through the deep learning technique described above, we calculate the distance between two features corresponding to consecutive time windows. For the t-th timestamp, the distance between the consecutive features f_t and f_{t-1} can be computed as

$$Dist_t = \frac{||f_t - f_{t-1}||_2}{\sqrt{||f_t||_2 \times ||f_{t-1}||_2}} \tag{1}$$

IPSN 2018, April 2018, PORTO, PORTUGAL

ACM ISBN 123-4567-24-567/08/06…$15.00

https://doi.org/10.475/123_4



(a) (b)

Figure 1: The performance of breakpoint detection under different methods using a smartphone sensor data set for activity recognition [4]. The green line represents the original signal and the red circle line is the ground truth of breakpoints. The yellow star lines in (a) represents the detected breakpoints by using existing Bayesian method. The blue triangle lines in (b) represent the detected breakpoints by using our method. We can see that our method significantly outperforms the previous approaches in finding breakpoints for real-world applications.

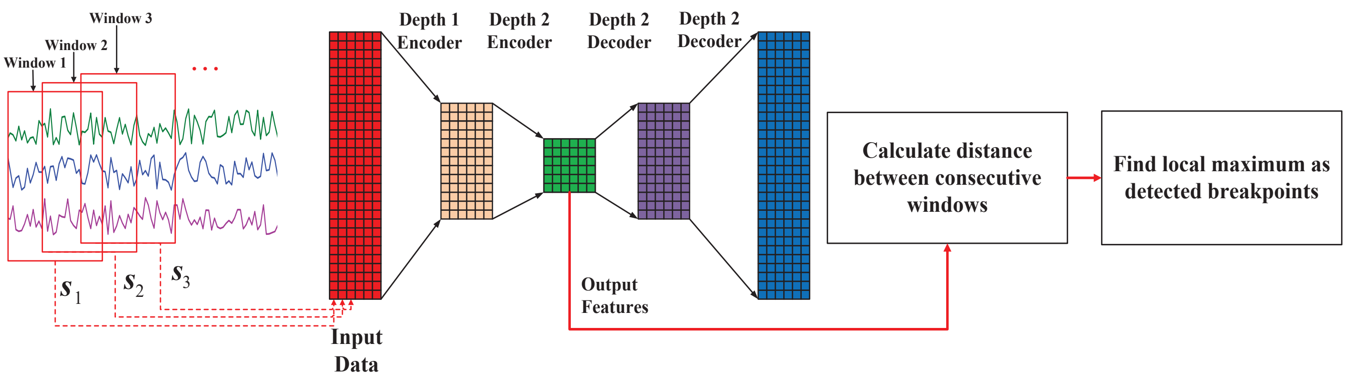


Figure 2: Pipeline for our breakpoint detection system. We first segment the input data into a series of windows and then apply autoencoder models in deep learning to extract representative features for the input data. These extracted features can then be utilized to calculate the distance between consecutive windows and the timestamps corresponding to local-maximal distance can be detected as breakpoints.

where the numerator is the Euclidean distance [5] between features corresponding to consecutive time windows, and the denominator serves as a normalization term.

Based on the computed distance of $\{Dist_t\}_{t=1}^{T/N_w}$ in Eq. 1, we construct a distance curve and select all the peaks (local-maximal) in the curve as breakpoints detected by our approach (see details in Figure 2).

It is worth noting that our approach can be broadly applied for general changepoint detection even outside of its application to breakpoint detection as considered in this paper. Compared with the state-of-the-art methods [1–3, 6, 7], our approach 1) has no assumption on the generative models for the input data and 2) the features are automatically extracted from the input data making them representative for analysis.

REFERENCES

[1] Ryan Prescott Adams and David JC MacKay. 2007. Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742* (2007).

[2] Jushan Bai. 1997. Estimation of a change point in multiple regression models. *Review of Economics and Statistics* 79, 4 (1997), 551–563.

[3] Daniel Barry and John A Hartigan. 1993. A Bayesian analysis for change point problems. *J. Amer. Statist. Assoc.* 88, 421 (1993), 309–319.

[4] CrowdSignals. [n. d.]. CrowdSignals.io. http://crowdsignals.io/. ([n. d.]).

[5] Michel Marie Deza and Elena Deza. 2009. Encyclopedia of distances. In *Encyclopedia of Distances*. Springer, 1–583.

[6] Chandra Erdman and John W Emerson. 2008. A fast Bayesian change point analysis for the segmentation of microarray data. *Bioinformatics* 24, 19 (2008), 2143–2148.

[7] Bonnie K Ray and Ruey S Tsay. 2002. Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis* 23, 6 (2002), 687–705.