Reevaluating the Change Point Detection Problem with Segment-based Bayesian Online Detection

Erick Draayer
New Mexico State University
edraayer@nmsu.edu

Huiping Cao
New Mexico State University
hcao@nmsu.edu

Yifan Hao
New Mexico State University
yifan@nmsu.edu

ABSTRACT

Change point detection is widely used for finding transitions between states of data generation within a time series. Methods for change point detection currently assume this transition is instantaneous and therefore focus on finding a single point of data to classify as a change point. However, this assumption is flawed because many time series actually display short periods of transitions between different states of data generation. Previous work has shown Bayesian Online Change Point Detection (BOCPD) to be the most effective method for change point detection on a wide range of different time series. This paper explores adapting the change point detection algorithms to detect abrupt changes over short periods of time. We design a segment-based mechanism to examine a window of data points within a time series, rather than a single data point, to determine if the window captures abrupt change. We test our segment-based Bayesian change detection algorithm on 36 different time series and compare it to the original BOCPD algorithm. Our results show that, for some of these 36 time series, the segment-based approach for detecting abrupt changes can much more accurately identify change points based on standard metrics.

CCS CONCEPTS

• **Mathematics of computing → Time series analysis**.

KEYWORDS

change point detection; time series analysis; Bayesian

ACM Reference Format:

Erick Draayer, Huiping Cao, and Yifan Hao. 2021. Reevaluating the Change Point Detection Problem with Segment-based Bayesian Online Detection. In *Proceedings of the 30th ACM Int'l Conf. on Information and Knowledge Management (CIKM '21), November 1–5, 2021, Virtual Event, Australia.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459637.3482167

1 INTRODUCTION

A common problem in time series analysis is identifying when a time series changes between states of data generation. This type of analysis is known as change point detection (CPD). CPD searches for abrupt changes within a time series. These abrupt changes are assumed to occur at a single point, known as the change point, which marks the transition from one state to another state. For example, finding abrupt changes in the state of a hospital patient's blood pressure is crucial for their well-being. CPD analysis is applied to a wide range of fields. such as human motion analysis [3], climate change detection [7], speech recognition [10], and finance [13].

Current CPD algorithms assume an instantaneous change between states of data generation. Therefore these algorithms are designed to find a single point that describes when this change occurs. However, while a change within a time series may be abrupt, it may actually happen over a relatively short segment of time. CPD algorithms designed to detect change as a single point may produce false positives or false negatives. Figure 1 shows an example of a time series that changes states of data generation. This change occurs over the segment between the red dotted lines. An algorithm designed to detect change as a single point might classify all, some, or none of the points between the red dotted lines as change points depending on the algorithm. In this situation, it is better to evaluate all the data points between the red dotted lines as a segment to determine if an abrupt change occurred. The segment of abrupt change described here is not like segments found by segmentation algorithms. A segment of abrupt change is a single occurrence not part of repeating patterns found by segmentation methods.

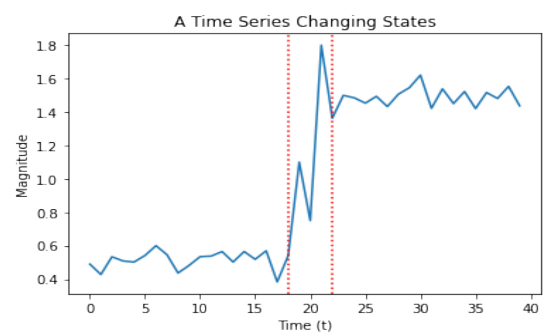

Figure 1: Example of simple time series displaying a change in state of data generation over a brief period of time.

In this paper we explore the idea of finding segments of abrupt changes within a time series by implementing a segment-based mechanism for the widely used CPD algorithm, Bayesian Online Change Point Detection (BOCPD) [1]. BOCPD has recently been shown to outperform several other CPD algorithms, old and new, across many different time series [14], and has been extensively studied across many survey papers [2, 4, 5, 8, 12, 14, 15]. BOCPD also has the advantage of being adaptable for both offline and online CPD problem settings. Therefore we focus our efforts on creating a segment-based mechanism for this method.

CIKM '21, November 1–5, 2021, Virtual Event, Australia.

ACM ISBN 978-1-4503-8446-9/21/11...$15.00
https://doi.org/10.1145/3459637.3482167

Our proposed method, Segment-based Bayesian Online Detection (SB-BOCPD), analyzes a small window of the most recent data from a stream of data instead of a single point. This window is compared to a set of probability density functions estimated from prior data within the stream to determine if the window captures an abrupt change in the time series. We test SB-BOCPD on a data set consisting of 31 real-world time series and 5 artificially generated time series. This data set was recently established from a survey paper that compiled a collection of time series from several domains and employed multiple human annotators to establish labels for change points [14]. Van den Burg and Williams [14] also adapted two standard metrics, covering and F_1-score, to handle annotations from multiple sources for evaluating CPD performance. Therefore we use covering and F_1-score to compare SB-BOCPD to BOCPD. These metrics encompass two prominent viewpoints that CPD should be evaluated as either a clustering or classification problem, respectively.

The contributions of this paper are summarized as follows:

- Our work is the first to to explore detecting abrupt changes in a time series as a short segment rather than a single point.
- We create Segment-Based Bayesian Online Detection (SB-BOCPD), a change detection algorithm based on Bayesian change point detection but adapted to include a segment-based mechanism to detect abrupt changes within a small window of time.
- We validate our proposed method on a total of 36 time-series data using two standard metrics for CPD evaluation.

2 METHODOLOGY

Let $S = \{x_1, x_2, ..., x_i, ...\}$ be a univariate time series data stream where x_i is the most recent observation at time stamp i. A segment of S from time stamp a to time stamp b where $a < b$ is denoted as $x_{a:b}$. Given a univariate time series data stream, S, our method returns an ordered set of non-overlapping change segments denoted as $C = \{c_{a_0:b_0}, c_{a_1:b_1}, ..., c_{a_n:b_n}\}$

2.1 Segment-based Bayesian Online Detection

Our proposed method is built upon Bayesian Online Change Point Detection (BOCPD) [1]. Like BOCPD, our method estimates a probability distribution over a run length using observations up to some recent data point within S. A recursive message-passing algorithm is used for calculating the joint distribution over the current run lengths. The main difference between our method and BOCPD is that our method creates a window, W_l, of the most recent data from the time series data stream. This window introduces a new parameter, l, which describes the length of the window.

$$W_l = \{x_{i-l}, x_{i-l+1}, ..., x_i\} \tag{1}$$

W_l is the set of the l most recent observations from the time series data stream. This window is evaluated to determine if it contains an abrupt change.

SB-BOCPD first initializes the parameters that describe the selected probability distribution. This set of parameters is denoted as $v_i^{(r)}$ and $X_i^{(r)}$. These parameters are inherited by the choice of probability distribution and usually refer to size, shape, and other statistical properties of the distribution.

After a new observation is read, W_l is updated and the predictive probability is calculated. Traditional BOCPD calculates a predictive probability for the most recent data point, x_t, as follows:

$$\pi_i^{(r)} = P(x_i | v_i^{(r)}, X_i^{(r)}) \tag{2}$$

The variable r is the current run length and x_t is the newest observation from the data stream. The predictive distribution for run time, r, is described by the set of parameters $v_i^{(r)}$ and $X_i^{(r)}$. In order to evaluate if W_l describes an abrupt change, we calculate the average of $(x_{i-l}, x_{i-1+1}, ..., x_i)$ in W_l, denoted as w_{avg}. We substitute x_i with w_{avg} in Eq. (2). We evaluated multiple methods for analyzing W_l to determine the presence of abrupt change. These methods included Kolmogrov-Smirnov test, energy statistics [11], distance metrics, and comparing probability density function values calculated from W_l versus those calculated from past observations outside W_l. However, none of these approaches performed as well as w_{avg}.

We change which observations are used to estimate the probability distributions over the run lengths. BOCPD normally estimates these distributions using all observations up to the most recent one, denoted as $x_{1:i}$. Our method estimates the probability distributions using all observations up to the beginning of the window, denoted as $x_{1:i-l}$. Therefore, our prediction function changes to Eq. (3).

$$\pi_{i-l}^{(r)} = P(w_{avg} | v_{i-l}^{(r)}, X_{i-l}^{(r)}) \tag{3}$$

From the predictive probability, we calculate the growth and change probability as defined in Eq. (4) and Eq. (5) respectively. H is a hazard function defined by the user that describes when a time series is expected to switch states.

$$P(r_i = r_{i-1} + 1 | x_{1:i-l}) = P(r_{i-1}, x_{1:i-l}) \pi_i^{(r)} (1 - H(r_{i-1})) \tag{4}$$

$$P(r_i = 0 | x_{1:i-l}) = \sum_{r_{i-1}} P(r_{i-1}, x_{1:i-l}) \pi_i^{(r)} (H(r_{i-1})) \tag{5}$$

We then update the run length posterior distribution, $P(r_i | x_{1:i-l})$ by calculating and dividing by the evidence defined as $P(x_{1:i-l} = \sum_{r_i} P(r_i, x_{1:i-l})$. Lastly, the parameters for the distribution, $X_i^{(r)}$ and $v_i^{(r)}$, are also updated before predicting if W_l is a segment of abrupt change.

To determine if the current window, W_l, is a segment of abrupt change, the maximum posterior at the current run length is calculated. If the change probability, $P(r_i = 0 | x_{1:i-l})$, is the maximum posterior probability then W_l is classified as a change segment. If change segments overlap, then the union of the segments is taken. This gives us our final output of non-overlapping change segments, $C = \{c_{a_0:b_0}, c_{a_1:b_1}, ..., c_{a_n:b_n}\}$. The code for our algorithm is available online (https://github.com/ecdraayer/SB-BOCPD)

2.2 Evaluation of Method

Our proposed method, SB-BOCPD, finds a set of non-overlapping change segments, whereas BOCPD and other CPD algorithms return a set of change points. In order to directly compare the performance of these two algorithms, we represent our change segments as a single point so that we can use standard metrics for evaluating CPD algorithms. We represent the segments as a single point by using the median of the change segment, denoted as $c_{mid_i} = c_{(a_i+b_i)/2}$.

We want to avoid creating another metric for evaluating change segment detection so that our method can be directly compared with CPD methods. Also, datasets labeled with change segments instead of change points would be needed before developing metrics tailored for evaluating change segment detection methods. Many metrics already exist for CPD, which makes evaluating the performance of CPD algorithms difficult across multiple papers that each use their own. Therefore, we focus on adapting our output for metrics that are well established and widely used for CPD.

Even though our method introduces a new parameter, l, which describes the window length. For this parameter, offline analysis or prior knowledge about the time series should help find the appropriate value of l. For example, a sensor recording human movement at a frequency of 500 Hertz can be expected to experience abrupt change within a single second and therefore l should be set to 500. Having a window also means that our proposed method needs to read $l + 1$ observations from the stream before the algorithm can start detecting abrupt changes. However, this disadvantage is negligible since the length of W_l should be very small (generally $\ll 1\%$) relative to the length of the time series data stream. Note that if $l = 1$, then SB-BOCPD becomes BOCPD.

3 EXPERIMENTS

Datasets. Our method is tested on 31 real-world time series and 5 synthetic time series established in a recent benchmark study [14]. The 5 synthetic time series are quality_control_1-5. The rest of the time series were collected from various online sources. Each time series is standardized. Time series were selected based on displaying abrupt changes and interesting behavior of seasonality, and outliers. This benchmark study also created annotations for each time series by employing several data scientists to mark points where they perceived abrupt changes. Each time series has five sets of annotations created by five different data scientists. All time series tested are univariate, but our proposed method can also perform on a multivariate time series stream. In a multivariate setting, change can be detected by running independent methods for each variable of the time series data stream, turning the problem into multiple univariate CPD problems. Therefore we test our proposed method only on 1-dimensional time series streams.

Baselines. We compare our proposed method, Segment-based Bayesian Online Detection (SB-BOCPD) to Bayesian Online Change Point Detection (BOCPD). We compare our method solely against BOCPD for two reasons: (i) BOCPD is a popular CPD algorithm used in offline and online settings and recently been shown to outperform many other CPD algorithms in a recent benchmark study [14] and (ii) we are researching the idea of detecting change within a short segment versus a single point.

Evaluation Metrics. We use F_1-score [9] and covering [6] as evaluation metrics for our algorithm. These are two standard metrics and were adapted to handle evaluating on time series data with multiple annotations in a recent benchmark study [14]. Both metrics range from scores 0 to 1.0. The ideal score is 1.0 and the worst possible score is 0 for both metrics. These metrics measure performance based on one of two major views that CPD is a classification or a clustering problem.

Table 1: F_1-scores and Covering for BOCPD and SB-BOCPD

	F_1-Score		Covering Score	
Dataset	BOCPD	SB-BOCPD	BOCPD	SB-BOCPD
bank	1.000	1.000	1.000	1.000
bitcoin	0.733	0.615	0.822	0.782
brent_spot	0.609	0.597	0.667	0.652
businv	0.588	**0.776**	0.693	**0.696**
centralia	1.000	0.909	0.753	0.675
children_per_woman	0.712	**0.810**	0.801	**0.861**
co2_canada	0.924	0.679	0.773	0.631
construction	0.709	**0.750**	0.585	**0.679**
debt_ireland	1.000	0.958	0.688	**0.814**
gdp_argentina	0.947	0.889	0.737	0.737
gdp_croatia	1.000	1.000	0.708	**0.833**
gdp_iran	0.862	**1.000**	0.583	**0.814**
gdp_japan	1.000	1.000	0.802	0.802
global_co2	0.889	**0.929**	0.758	0.758
homeruns	0.829	**0.879**	0.694	0.694
jfk_passengers	0.776	**0.966**	0.837	**0.873**
lga_passengers	0.704	0.683	0.547	**0.596**
measles	0.947	0.947	0.951	0.951
nile	1.000	1.000	0.888	0.857
ozone	0.857	**1.000**	0.602	**0.852**
quality_control_1	1.000	1.000	0.996	0.989
quality_control_2	1.000	1.000	0.927	0.917
quality_control_3	1.000	1.000	0.997	0.997
quality_control_4	0.787	0.780	0.673	0.673
quality_control_5	1.000	1.000	1.000	1.000
rail_lines	0.966	0.966	0.768	**0.872**
ratner_stock	0.868	**0.889**	0.906	0.874
robocalls	0.966	**1.000**	0.808	0.752
scanline_126007	0.921	0.748	0.631	**0.654**
scanline_42049	0.962	0.799	0.892	0.778
seatbelts	0.683	**0.824**	0.800	0.688
uk_coal_employ	0.868	**0.966**	0.920	**0.928**
unemployment_n1	0.876	0.843	0.669	0.628
usd_isk	1.000	0.814	0.737	**0.863**
us_population	0.785	0.615	0.853	0.787
well_log	0.832	0.495	0.793	0.679
Average	0.878	0.865	0.785	**0.795**

$$F_1 = \frac{2PR}{P + R} \tag{6}$$

$$P = \frac{|TP(Y^*, C)|}{|C|}, \; R = \frac{1}{K}\sum_{k=1}^{K}\frac{|TP(Y_k, C)|}{|Y_k|} \tag{7}$$

As a classification problem, each observation is classified as either a "change-point" or "not change-point". In this scenario, the F_1-score can be used to evaluate the performance of the CPD algorithm. Eq. (6) defines the F_1-score where P is precision and R is recall. Precision is defined in Eq. (7) where C is the set of change points found by an algorithm and Y^* is the combined set of all annotations established for the time series. $TP(Y^*, C)$ is the set of true positives for Y^* and C. The recall, R, is defined by Eq. (7) and is the average of the recalls calculated using each set of change points established by annotators. A true positive, TP, is determined based on margin of error, M. If the difference between the ground truth change point and detected change point is $< M$ then it is counted as a true positive. We use margin of error $M = 5$ in all experiments. When viewed as a clustering problem, CPD algorithms partition the time series according to the location of the detected change points. Partitions are created at each change point. In this scenario, covering can be used as a metric to evaluate performance based on similarity between partitions from the CPD algorithm output and the ground truth.

$$C(B', B) = \frac{1}{T} \sum_{A \in B} |A| max_{A' \in B'} J(A, A') \quad (8)$$

Eq. (8) defines covering where B' and B are two sets of partitions of a time series. $J(A, A')$ is the intersection over union of the two sets A and A', also known as the Jaccard Index. T is the total length of the time series dataset. Since our dataset has multiple annotators, we compute the average covering across all annotators for the final measure of performance.

Parameter Choice. BOCPD and SB-BOCPD both use the Gaussian distribution with the negative inverse gamma prior. No truncation of the run length is applied. The initial mean of the distribution is set to 0. We perform a grid search for the parameters α_0, β_0, κ_0, and λ_{gap} used by both BOCPD and SB-BOCPD. The parameters α_0, β_0, and κ_0 describe the Gaussian distribution with negative inverse gamma prior. The parameter λ_{gap} is a constant associated with the hazard function of BOCPD and SG-BOCPD. The hazard function is defined as $H(t) = 1/\lambda_{gap}$. We also perform a grid search for the parameter, l that describes the window length in SB-BOCPD. The parameters are varied as follows:

- α_0 = (0.01,1,100)
- β_0 = (0.01,1,100)
- κ_0 = (0.01,1,100)
- λ_{gap} = (50,100,200)
- l = (3,5,8,13,21,34)

Performance Comparison. Table 1 shows covering and F_1-scores across the 36 different time series. Scores highlighted in bold indicate SB-BOCPD outperformed BOCPD for that time series.

Table 1 shows SB-BOCPD slightly outperforms BOCPD on average in terms of covering. Closer inspection shows large differences between scores for most of the datasets. For the datasets gdp_iran and ozone we can see a major performance increase of over 40% in favor of SB-BOCPD. Other datasets such as construction, gdp_ireland, gdp_croatia, usd_isk, and rail_lines show other instances of substantially better performance for SB-BOCPD. However, BOCPD substantially outperforms SB-BOCPD on datasets such as centralia, co2_canada, scanline_42049, us_population, and well_log.

Table 1 also shows BOCPD slightly outperforms SB-BOCPD on average in terms of F_1-score. We can see a similar pattern of many datasets having substantial differences between F_1-scores. The datasets businv, children_per_woman, jfk_passengers, ozone, and seatbelts show substantially better performance for SB-BOCPD. The datasets bitcoin, co2_cananda, scanline_126007, scanline_42049, us_population, and well_log show BOCPD substantially outperforming SB-BOCPD.

Table 1 shows a similar pattern between covering and F_1-scores of certain datasets performing much better with one method over the other. Datasets such as ozone and construction seem to be better suited for our proposed method, SB-BOCPD. Other datasets such as scanline_42049 and well_log remain better suited for BOCPD.

Inspection of individual datasets where SB-BOCPD performed well tended to have abrupt changes over a segment. Annotations of change points from the multiple annotators of these datasets were typically in disagreement about the exact point, but concentrated in the same area of the time series. Datasets where BOCPD performed well tended to have abrupt changes marked by a single point. Therefore we attribute the major differences in performance between BOCPD and SB-BOCPD to the behavior of abrupt changes in the datasets.

Computational Cost. SB-BOCPD maintains the linear space- and time-complexity, $O(n)$, per time stamp of BOCPD. The additional calculation of w_{avg} per time stamp is $O(m)$ where $m << n$ and the only additional memory is for the m observations of W_l. Truncation of low probability estimations in the tail of the distribution can further reduce running time.

4 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed the idea of detecting abrupt changes in time series by examining short segments of time. Traditionally, change detection algorithms are built upon the assumption that changes in the state of data generation of a time series happen instantaneously. We argue that this assumption is flawed, and that it may be better to detect abrupt changes within a short segment of time for some time series. We demonstrate this by adapting Bayesian Online Change Point Detection (BOCPD), a popular CPD algorithm, to use a segment-based mechanism for detecting change within a short window. We call our method Segment-Based Bayesian Online Detection (SB-BOCPD). Our method is compared to BOCPD using two standard metrics for CPD across 36 different time series datasets. Our results show major differences in performance between these two methods, suggesting that improvements to change detection can be made by detecting abrupt change as a segment.

For future research we plan to adapt other CPD algorithms with a segment-based mechanism. There is also a lot of potential to develop more sophisticated segment-based mechanisms. For example, our segment-based mechanism for BOCPD relies on a fixed window length. However, abrupt changes within a time series may happen over varying lengths. Developing a segment-based mechanism with a self-adjustable window length may provide better results. Time series segmentation, which traditionally partitions a time series using a set of single points called cut points, may also benefit from a similar segment based approach. Instead of searching for cut points, a segment based method would search for cut segments.

ACKNOWLEDGMENTS

This work is partially supported by NSF #1757207, #1914635, and GAANN DOE grant number P200A180005.

REFERENCES

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

[2] Samaneh Aminikhanghahi and Diane J Cook. 2017. A survey of methods for time series change point detection. *Knowledge and information systems* 51, 2 (2017), 339–367.

[3] Samaneh Aminikhanghahi and Diane J Cook. 2017. Using change point detection to automate daily activity segmentation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 262–267.

[4] Michele Basseville. 1988. Detecting changes in signals and systems—a survey. *Automatica* 24, 3 (1988), 309–326.

[5] Sooyoung Cheon and Jaehee Kim. 2010. Multiple change-point detection of multivariate mean vectors with the Bayesian approach. *Computational Statistics & Data Analysis* 54, 2 (2010), 406–415.

[6] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. 2010. The pascal visual object classes (voc) challenge. *International journal of computer vision* 88, 2 (2010), 303–338.

[7] Zhenmei Ma, Shaozhong Kang, Lu Zhang, Ling Tong, and Xiaoling Su. 2008. Analysis of impacts of climate variability and human activity on streamflow for a river basin in arid region of northwest China. *Journal of hydrology* 352, 3-4 (2008), 239–249.

[8] Aleksey S Polunchenko and Alexander G Tartakovsky. 2012. State-of-the-art in sequential change-point detection. *Methodology and computing in applied probability* 14, 3 (2012), 649–684.

[9] C. J. Van Rijsbergen. 1979. Information Retrieval. *Butterworths* (1979).

[10] David Rybach, Christian Gollan, Ralf Schluter, and Hermann Ney. 2009. Audio segmentation for speech recognition using segment features. In *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 4197–4200.

[11] Gábor J Székely. 2003. E-statistics: The energy of statistical samples. *Bowling Green State University, Department of Mathematics and Statistics Technical Report* 3, 05 (2003), 1–18.

[12] Alexander G Tartakovsky and George V Moustakides. 2010. State-of-the-art in Bayesian changepoint detection. *Sequential Analysis* 29, 2 (2010), 125–145.

[13] Sven Thies and Peter Molnár. 2018. Bayesian change point analysis of Bitcoin returns. *Finance Research Letters* 27 (2018), 223–227.

[14] Gerrit JJ van den Burg and Christopher KI Williams. 2020. An evaluation of change point detection algorithms. *arXiv preprint arXiv:2003.06222* (2020).

[15] S Zacks. 1983. Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation. In *Recent advances in statistics*. Elsevier, 245–269.