Change Point Detection for Streaming High-Dimensional Time Series

Masoomeh Zameni[1(✉)], Zahra Ghafoori[1], Amin Sadri[2], Christopher Leckie[1], and Kotagiri Ramamohanarao[1]

[1] University of Melbourne, Melbourne, Australia
{masoomeh.zameni,zahra.ghafoori,caleckie,kotagiri}@unimelb.edu.au
[2] RMIT University, Melbourne, Australia
amin.sadri@rmit.edu.au

Abstract. An important task in analysing high-dimensional time series data generated from sensors in the *Internet of Things* (IoT) platform is to detect changes in the statistical properties of the time series. Accurate, efficient and near real-time detection of change points in such data is challenging due to the streaming nature of it and the presence of irrelevant time series dimensions. In this paper, we propose an unsupervised Information Gain and permutation test based change point detection method that does not require a user-defined threshold on change point scores and can accurately identify changes in a sequential setting only using a fixed short memory. Experimental results show that our efficient method improves the accuracy of change point detection compared to two benchmark methods.

Keywords: Change detection · Information Gain · Permutation test

1 Introduction

Change point detection is detecting time points when the statistical properties of a time series, such as the mean or variance, changes. This problem actively studied for a variety of applications such as human activity segmentation (Fig. 1). Existing change point detection methods are batch [1] or sequential [2]. Batch change point detection methods require the whole time series data to be stored for detecting change points. Sequential change point detection methods, however, detect change points as data points arrive. In real-world applications, data is usually streaming which necessitates sequential methods. Most change point detection approaches estimate the probability distribution of the data points in the windows before and after the time point t and then assess the divergence between these two distributions using a user-defined threshold. However, estimating the distribution and setting the user-defined threshold are highly domain dependant. Moreover, in high-dimensional time series only some dimensions may be responsible for a change, which degrades the performance of probability distribution-based techniques. In this paper, we propose a *Sequential Information*

G. Li et al. (Eds.): DASFAA 2019, LNCS 11448, pp. 515–519, 2019.
https://doi.org/10.1007/978-3-030-18590-9_78

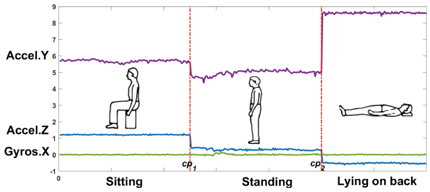


Fig. 1. Change point detection for activity segmentation.

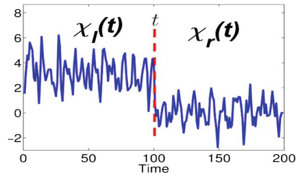


Fig. 2. A sample change point at time point t.

Gain Permutation-based (SequentialIGPerm) method that sequentially detects change points in high-dimensional time-series. Our distribution-free method uses a permutation test to validate candidate change points.

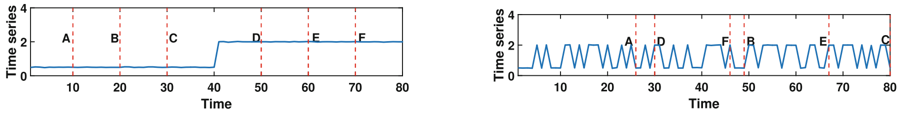
$X = \langle x(1), \cdots, x(t), \cdots \rangle$ denotes an *m-dimensional streaming time series*, where $x(t) \in \Re^m$ arrives at time t from m sensors. Let $\chi_l(t) := \{x(t-N+i)\}_{i=1}^{N}$ and $\chi_r(t) := \{x(t+i)\}_{i=1}^{N}$ be the windows that include N data points before and after time t, respectively (Fig. 2) that are generated according to some unknown probability density functions pdf_1 and pdf_2. Time t is a change point if $pdf_1 \neq pdf_2$. A change point detection method divides X into k non-overlapping *segments*. Each change point cp_j denotes the point between two consecutive segments. SequentialIGPerm aims to sequentially detect the change points at the arrival of a new sequence of the data.

2 Background

The *Information Gain* (IG) metric proposed in [3] used for calculating the change point scores in a batch setting. The increase in IG, the expected reduction in the entropy, for segmenting the high-dimensional time series X into k segments (using $k-1$ change points) is defined using the cost function $\mathcal{L}_k$ as $\mathcal{L}_k(X,T) = H(X) - \sum_{j=1}^{k} \frac{|\mathbf{s}_j|}{|X|} H(\mathbf{s}_j)$, where $T = \{cp_1, cp_2, \cdots, cp_{k-1}\}$ is the set of detected change points, $\mathbf{s}_j$ is the j^{th} segment, $|.|$ is the length operator, $H(X)$ is the entropy of the whole time series as a segment, and $H(\mathbf{s}_j)$ is the entropy of the segment $\mathbf{s}_j$ measured as $H(\mathbf{s}_j) = -\sum_{i=1}^{m} p_{ji} \log p_{ji}$, where m is the number of time series dimensions, and p_{ji} ($p_{ji} \leqslant 1$, $\sum_{i=1}^{m} p_{ji} = 1$) is the mass ratio of time series dimension c_i in segment $\mathbf{s}_j$ (the area under time series dimension c_i in segment $\mathbf{s}_j$ divided by the sum of the areas under all dimensions in segment $\mathbf{s}_j$) defined as $p_{ji} = \frac{\sum_{q \in \mathbf{s}_j} q(c_i)}{\sum_{d=1}^{m} \sum_{q \in \mathbf{s}_j} q(c_d)}$, where $q(c_i)$ denotes the value of the q^{th} point of the i^{th} time series dimension within the segment $\mathbf{s}_j$. The greatest reduction in the entropy of the signals is achieved when the time series is divided into the most coherent segments (the lowest variance segments) [3].

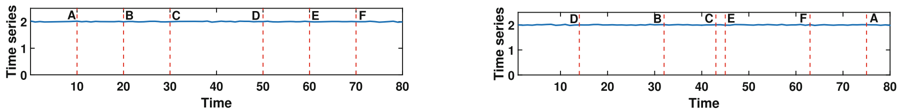
3 Proposed Method

For each data point $x(t)$ in a new data sequence, SequentialIGPerm computes its change point score (IG) using the data points in the two sub-windows $\chi_l(t)$



(a) The original window W (b) The permuted window W'

Fig. 3. Example of permutation test when a significant change point exists at $t = 41$. $\mathcal{L}(W,t) = 0.57$, whereas $\mathcal{L}(W',t)$ is reduced to 0.02 ($\mathcal{L}(W,t) \gg \mathcal{L}(W',t)$).



(a) The original window W (b) The permuted window W'

Fig. 4. Example of permutation test when no change point exists at $t = 41$. $\mathcal{L}(W,t) = 0.0087$, whereas $\mathcal{L}(W',t)$ is 0.0086 ($\mathcal{L}(W,t) \approx \mathcal{L}(W',t)$).

and $\chi_r(t)$. This indicates how much the IG of segmenting $W = \{\chi_l(t), \chi_r(t)\}$ is increased when $x(t)$ is considered as a change point. As adding each change point always increases IG (see [3] for the proof), we propose a method that tests if the increase of IG is likely to be due to statistical change in the data. We use a permutation test that generates a new window $W' = \{\chi'_l(t), \chi'_r(t)\} = \{x(\Pi(i)); x(i) \in W, i = \{1, .., |W|\}\}$, such that $\Pi(i = t) = t$ and $\Pi(i \neq t) \in \{1, 2, \cdots, |W|\} - \{t\}$ is the permutation function. W' is generated by randomly permuting data points around $x(t)$ in W. The data distribution is the same for both W and W' but the order of data points in W and W' is different. Then, we compute the increase in IG, i.e., $\mathcal{L}_1(W,t)$ and $\mathcal{L}_1(W',t)$, when W and W' are segmented into two segments considering time $t = cp_j$ as the change point.

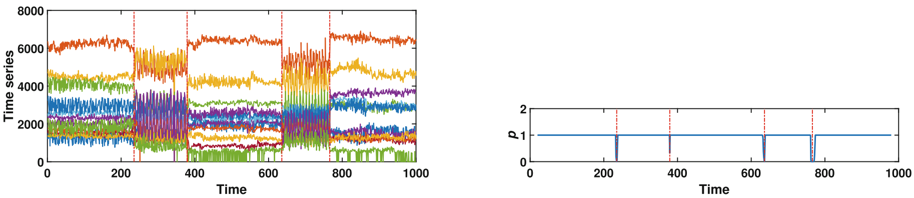
Theorem 1. *Define the alternative hypothesis $H_1 : \mathcal{L}(W,t) \gg \mathcal{L}(W',t)$ against the null hypothesis $H_0 : \mathcal{L}(W,t) \approx \mathcal{L}(W',t)$. If a change point occurs at time t, H_0 would be rejected.*

Proof. Consider two scenarios: (a) no change point occurs at time t, and (b) a change point occurs at time t. We prove that H_0 is rejected in scenario (b) but is not rejected in scenario (a). In scenario (a) $\chi_l(t) \in pdf_1$ and $\chi_r(t) \in pdf_2$ and $pdf_1 \approx pdf_2 \Rightarrow \mathcal{L}(W,t) \approx \epsilon$. In contrast, $\chi'_l(t) \in \{pdf_1, pdf_2\}$ and $\chi'_r(t) \in \{pdf_1, pdf_2\}$ and $pdf_1 \approx pdf_2 \Rightarrow E(\mathcal{L}(W',t)) \approx \epsilon$, where E is the expected value and ϵ represents a small value. Thus, $\mathcal{L}(W,t) \approx E(\mathcal{L}(W',t))$, H_0 is not rejected and t is rejected as a change point. In scenario (b), $\chi_l(t) \in pdf_1$ and $\chi_r(t) \in pdf_2$ and $pdf_1 \neq pdf_2 \Rightarrow \mathcal{L}(W,t) \gg \epsilon$. In contrast, $\chi'_l(t) \in \{pdf_1, pdf_2\}$ and $\chi'_r(t) \in \{pdf_1, pdf_2\}$ and $pdf_1 \approx pdf_2 \Rightarrow E(\mathcal{L}(W',t)) \approx \epsilon$. Thus, $\mathcal{L}(W,t) \gg E(\mathcal{L}(W',t))$, H_0 is rejected and t is accepted as a change point (Figs. 3, 4).

To test H_0, we generate a number of randomly permuted windows W'. Then, we check how often $(\mathcal{L}(W,t) - \mathcal{L}(W',t)) > \epsilon$. This probability of an increase in IG, i.e., $\mathcal{L}(W,t)$, by chance is the p-value of the data under the null hypothesis. The p-value is computed as $p = \frac{\sum_{g=1}^{N_p} I(\mathcal{L}(W'_g,t)-\mathcal{L}(W,t)<\alpha)}{N_p}$, where N_p is the number of permuted windows generated for each time stamp t, W'_g is the g^{th} permuted window and $\alpha \in [0,1]$ is a parameter set in an unsupervised manner using the training data. If p is low, e.g., below 0.05, then $\mathcal{L}(W,t)$ is significant change point is confirmed at time t (Fig. 5).

4 Experimental Results

We compare our method, SequentialIGPerm, with two sequential change point detection methods, namely ecp [4], and RuLSIF [2]. In SequentialIGPerm, $N = 20$, i.e., $|W| = 40$. The reason for choosing a small window size is to reduce the detection delay. For each W, $k_p = 20$ randomly permuted windows W' are generated to calculate the p-value. We observed a subtle change in the accuracy when we increased k_p from 20 to 2000. We set $k_p = 20$ in our experiments. In the p-value calculation, α is set to 0.3, 0.1, and 0.03, respectively for the synthetic, DSA, and RFID datasets based on our training data in an unsupervised manner. We assume that training data X_{train} including k change points is available to set our only parameter α. The value of k is estimated using the method proposed in [3]. The value of α is set based on the average of the minimum $\mathcal{L}$ for these k true change points and the maximum $\mathcal{L}$ for the non-change points in X_{train}. The time points with a p-value lower than 0.05 are considered as change points for the test



(a) The original RFID dataset (b) p-values

Fig. 5. p-values after applying the permutation test on the RFID dataset.

Table 1. AUC and run-time results

Data	Type	Length	m	#CP	AUC			Run-time (sec)		
					SIGP[a]	ecp [4]	RuLSIF [2]	SIGP	ecp	RuLSIF
GME25 [5]	Synthetic	26000	25	25	**1**	0.99	0.9	0.02	0.01	0.1
GME100 [5]	Synthetic	101000	100	100	**1**	0.99	0.9	0.09	0.01	0.13
Syn1000 [3]	Synthetic	10000	1000	200	**1**	0.55	0.88	1.3	0.01	1.79
DSA [6]	Real	14000	45	111	**0.96**	0.71	0.88	0.04	0.01	0.11
RFID [7]	Real	6812	12	44	**0.91**	0.69	0.85	0.01	0.01	0.1

[a] SequentialIGPerm.

data in all our experiments. The distance margin $\delta = 20$ is set for all datasets. The parameters in ecp are set as recommended in [4]. The results for RuLSIF are reported based on the best results obtained by setting parameter values recommended in [2]. Table 1 reports the *Area Under the ROC Curve* (AUC) and run-time of the methods. The results show that SequentialIGPerm detects the true change points more accurately with than the benchmarks with comparable run-time, which confirms the accuracy of our method is not degraded by irrelevant dimensions in high-dimensional time series. The time complexity of our method, ecp, and RuLSIF are $O(nmk_p)$, $O(n\,|W|^2)$, and $O(nm^2\,|W|^2\,L)$, respectively, where L is the number of retrospective subsequences that are considered for change point detection at each time t.

5 Conclusion

We proposed a novel distribution and user-defined threshold-free approach to detect change points in streaming high-dimensional time series data using an information gain metric and a permutation test. Our experimental results showed the advantages of our method compared to the examined benchmark techniques.

References

1. Vasko, K.T., et al.: Estimating the number of segments in time series data using permutation tests. In: ICDM, pp. 466–473 (2002)
2. Liu, S., et al.: Change-point detection in time-series data by relative density-ratio estimation. Neural Netw. **43**, 72–83 (2013)
3. Sadri, A., et al.: Information gain-based metric for recognizing transitions inhuman activities. In: PMC, vol. 38, pp. 92–109 (2017)
4. Matteson, D.S., et al.: A nonparametric approach for multiple change point analysis of multivariate data. J. Am. Stat. Assoc. **109**(505), 334–345 (2014)
5. Rajasegarar, S., et al.: Measures for clustering and anomaly detection in sets of higher dimensional ellipsoids. In: IJCNN, pp. 1–8 (2012)
6. Daily and sports activities data set (2014). https://archive.ics.uci.edu/ml/datasets/Daily+and+Sports+Activities
7. Yao, L., et al.: Unobtrusive posture recognition via online learning of multi-dimensional RFID received signal strength. In: ICPADS, pp. 116–123 (2015)