Causal Discovery-Driven Change Point Detection in Time Series

Shanyun Gao[1], Raghavendra Addanki[2], Tong Yu[2], Ryan A. Rossi[2], Murat Kocaoglu[1]

[1] Purdue University, West Lafayette, USA

[2] Adobe Research, San Jose, USA

Motivations

Traditionally, change detection focuses on shifts in **the joint distribution** of all variables—capturing system-wide changes but missing **local changes** in specific variables. Such localized changes often carry more real-world significance.

Moreover, many existing methods rely on the assumption of independently and identically distributed (IID) samples, which may not hold in practice.

Our Goal

*Can we detect changes targeting the underlying **causal mechanisms** of a time series—rather than joint or marginal distributional shifts—with theoretical guarantees and without assuming IID samples?*

Key Idea

Phase One (causal discovery) focuses on uncovering the underlying causal structure of the entire non-stationary time series.

Phase Two (change point detection) centers on change point detection within each time series segment, using the causal structures estimated in Phase One to guide the segmentation.

Dynamic α-relative density-ratio:

$$r_{\alpha_i}(x|\text{spa}(x)) = \frac{p}{(1-\alpha_i)p + \alpha_i p'} := \frac{p}{q_{\alpha_i}}$$

Dynamic α-relative Pearson Divergence and its Estimator:

$$PE_{\alpha_i}[p, p'] := \frac{1}{2}\mathbb{E}_{x\sim q_{\alpha_i}}[(r_{\alpha_i}(x|\text{spa}(\text{x})) - 1)^2]$$

where PE_{α_i} is estimated from a method called RuLSIF [3,4] with kernel functions.

Theorem. *Let $\{\widehat{PE}_{\alpha_i}\}_i$ be the estimated PE series for one time series segment $X^j(\Lambda) \subsetneq \mathbf{X}^j \subsetneq V$ and T_c^j denote the true change point in this time series segment. $\widehat{T}_c^j$ denotes the estimator of T_c^j obtained by:*

$$\widehat{T}_c^j = \arg_i \max\{\widehat{PE}_{\alpha_i}\}_i$$

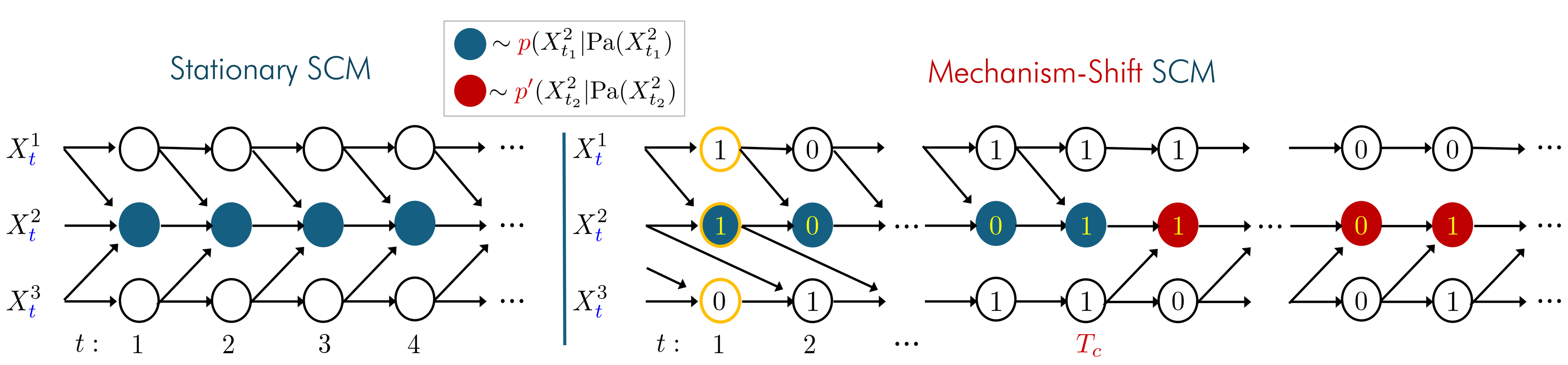
Under certain assumptions, we have that given large enough n_w,

$$\Pr\left(\left|\widehat{T}_c^j - T_c^j\right| < 2n_w\right) > \left(1 - \frac{a_w - 2}{b_{st}\log T_{sub}} - \frac{a_w}{T_{sub}}\right)\left(1 - \frac{1}{n_w}\right)$$

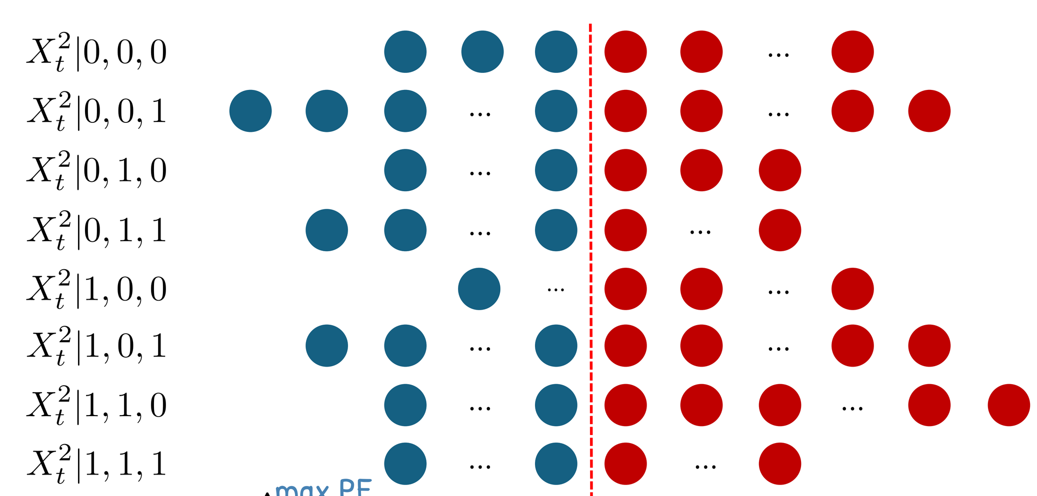
where $b_{st} = \left\lfloor \frac{\log T_{sub}}{n_{st}} \right\rfloor$, $a_w = \left\lceil \frac{T_{sub}}{n_w} \right\rceil$ *and* $T_{sub} := |X^j(\Lambda)|$.

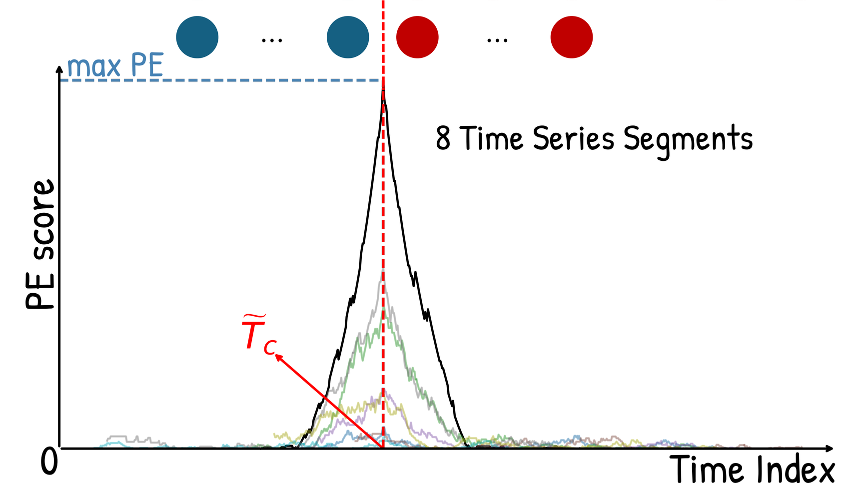
Note: n_{st} is the stride length, i.e., the number of time points the sliding window jumps forward at each step. $2n_w$ is the length of the sliding window. Λ denotes the index of the time series segment of $\mathbf{X}^j$.

Union of Parent Sets in Non-Stationary Time Series



Time Series Segments





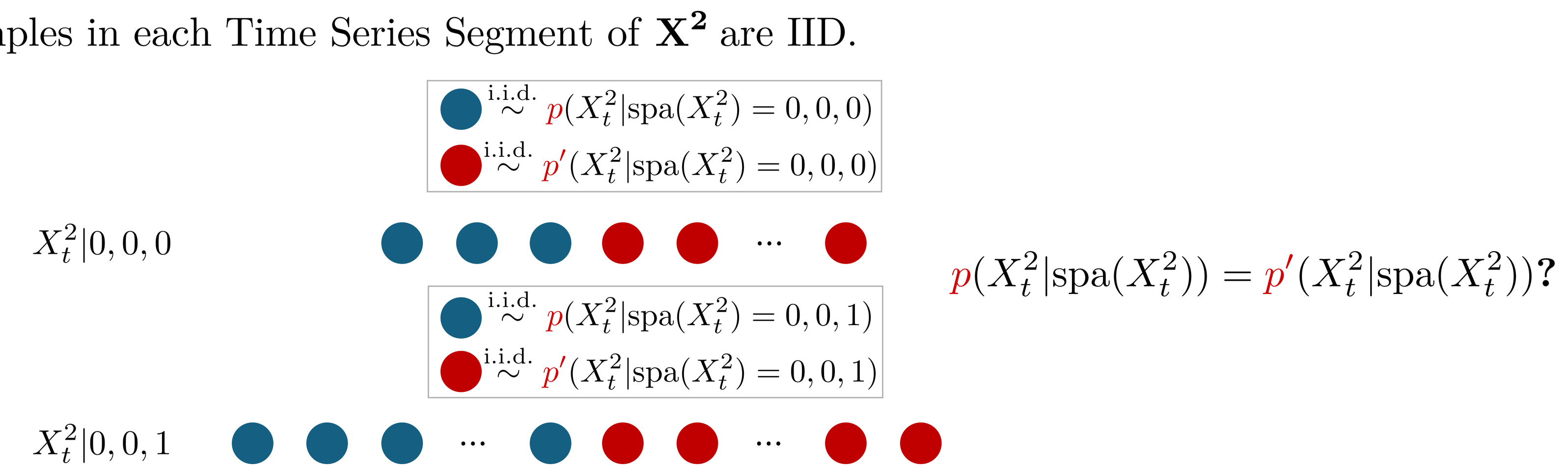
Theorem (adapted from [1]). *Let $\text{SPA}(X_t^j)$ denote the union parent set of X_t^j and $\widehat{\text{SPA}}(X_t^j)$ denote the estimated union parent set obtained from PCMCI [2] algorithm on time series $\mathbf{X}^j$ with a Mechanism-Shift SCM. Under certain assumptions and with an oracle (infinite sample size limit), we have that:*

$$\text{SPA}(X_t^j) \subseteq \widehat{\text{SPA}}(X_t^j)$$

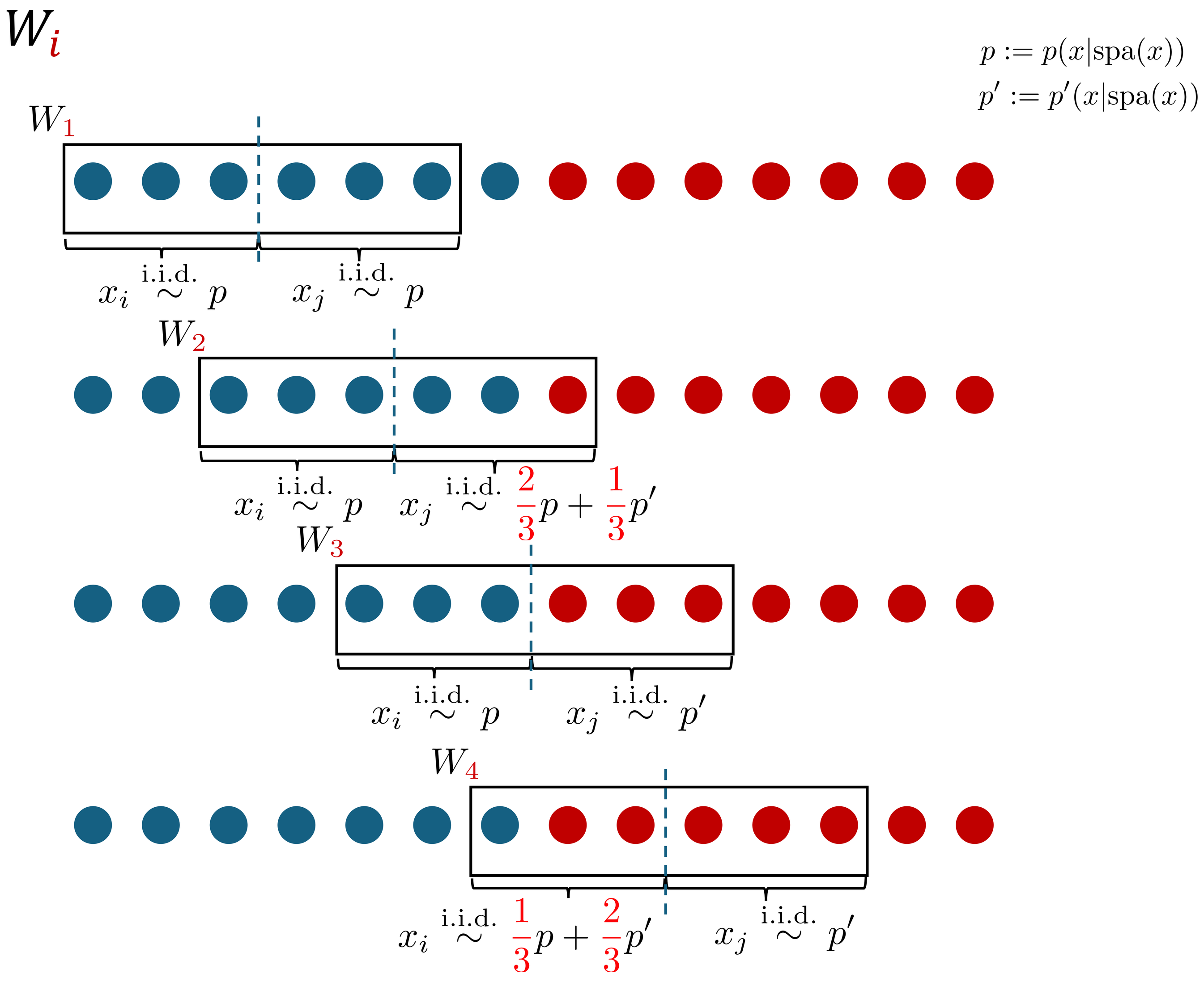
Causal-RuLSIF: Change Point Detection

Definition [Time Series Segments]. *A collection of non-overlapping, non-empty subsets of a univariate, discrete-valued time series $\mathbf{X}^j \in V$ governed by a Mechanism-Shift SCM, where all segments share the same union parent set $\text{SPA}(X_t^j)$ over a fixed domain, but differ in their configurations $\text{spa}(X_t^j)$.*

Samples in each Time Series Segment of $\mathbf{X}^2$ are IID.



Dynamic α-Relative Pearson Divergence PE_{α_i} via Sliding Window W_i

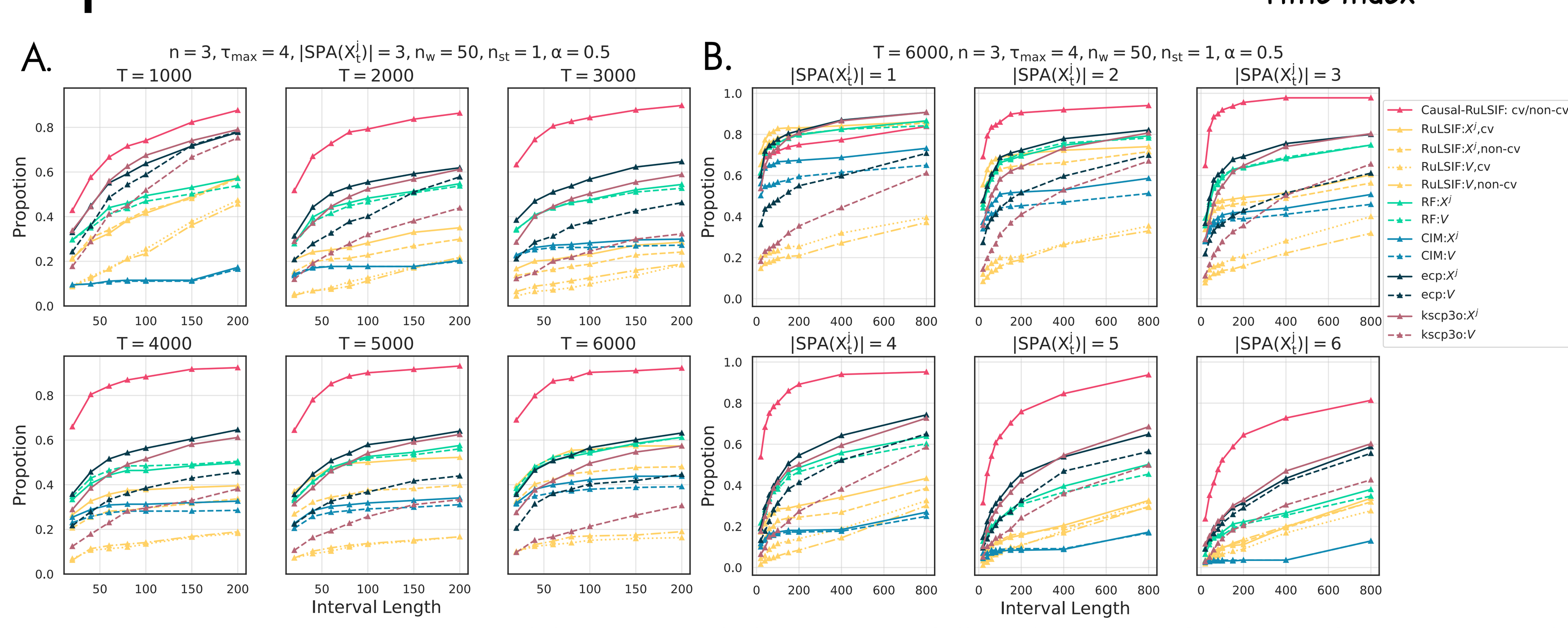


One of the two sample sets from each half of the sliding window follows a mixture distribution of the form $(1-\alpha_i)p + \alpha_i p'$, assuming one change point per window.

$\alpha_i \propto$ the concentration of p' in such mixture distribution which depends on the window index i.

Instead of a fixed hyperparameter α, we have a dynamic parameter α_i that corresponds to the sliding window W_i.

Experiments



1. Proportion (Y-axis): Counts as 1 if the estimated change point falls within a specified interval (X-axis) around the true change point. The proportion is calculated over all true change points.
2. In the legend, CV: cross-validation is used to select RuLSIF hyperparameters; non-CV: no cross-validation. X^j: a univariate time series component in V; V: the whole multivariate time series.
3. Baseline Algorithms: RuLSIF: Liu et al. (2013), CIM (Change in Mean): Vostrikova(1981), RF(ChangeForest): Londschien et al. (2023), ecp: James and Matteson (2013), kscp3o: Zhang et al. (2017). Each marker represents the average accuracy over 100 random trials.

In Figure A, Causal-RuLSIF is tested with different time series lengths T, and its performance improves as the time series length T increases, empirically supporting its consistency.

In Figure B, Causal-RuLSIF is tested with different sizes of $\text{SPA}(X_t^j)$. It is less accurate when $|\text{SPA}(X_t^j)| = 1$ as variables depend only on their own past X_{t-1}^j, with no cross-series correlation. When $|\text{SPA}(X_t^j)| > 1$, Causal-RuLSIF outperforms others by capturing shifts in causal mechanisms. However, its performance drops as $|\text{SPA}(X_t^j)|$ increases due to the exponential growth in the total number of parent configurations, which reduces the effective sample size.

Case Study: A Real-world Air Pollution Dataset

This dataset monitors the amount of PM_{10} (coarse particles with a diameter between 2.5 and 10 micrometers) in the air. The 3-variate time series data records the hourly concentration of PM_{10} across three counties in California—Fresno, Mono and Monterey—from Jan to June 2023.

During the first half of 2023, the causal mechanism of PM_{10} in Fresno is likely to shift on May 8, 2023. While the PM_{10} levels in Fresno and Monterey are not influenced by other counties, a causal link from Monterey to Mono has emerged after February 2, 2023.

Table 1: Causal-RuLSIF in PM_{10} dataset

X	$\widetilde{T}_c^j$	$\widehat{\text{PA}}(X^j_{t<\widetilde{T}_c^j})$; $\widehat{\text{PA}}(X^j_{t\geq\widetilde{T}_c^j})$
$\mathbf{X}^{\text{Fr}}$	05/08/23	$\{X^{\text{Fr}}_{t-1}\}$; $\{X^{\text{Fr}}_{t-1,t-2,t-3}\}$
$\mathbf{X}^{\text{Mono}}$	02/01/23	$\{X^{\text{Mono}}_{t-1}\}$; $\{X^{\text{Mono}}_{t-1,t-2}, X^{\text{Mont}}_{t-3}\}$
$\mathbf{X}^{\text{Mont}}$	04/04/23	$\{X^{\text{Mont}}_{t-1}\}$;$\{X^{\text{Mont}}_{t-1}\}$

References

[1] Gao, S., Addanki, R., Yu, T., Rossi, R. and Kocaoglu, M., 2023. Causal discovery in semi-stationary time series. *Advances in Neural Information Processing Systems, 36*, pp.46624-46657.
[2] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D., 2019. Detecting and quantifying causal associations in large nonlinear time series datasets. Science advances, 5(11), p.eaau4996.
[3] Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H. and Sugiyama, M., 2013. Relative density-ratio estimation for robust distribution comparison. Neural computation, 25(5), pp.1324-1370.
[4] Liu, S., Yamada, M., Collier, N. and Sugiyama, M., 2013. Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 43, pp.72-83.