Conjugate Bayesian Two-step Change Point Detection for Hawkes Process

Zeyue Zhang[1,2], Xiaoling Lu[1,2], Feng Zhou[1,3*]

[1]Center for Applied Statistics and School of Statistics, Renmin University of China
[2]Innovation Platform, Renmin University of China
[3]Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing
{zhangzeyue, xiaolinglu, feng.zhou}@ruc.edu.cn

中国人民大學
RENMIN UNIVERSITY OF CHINA

Background

Background

- Point process data: widely used in finance[1], neuroscience[11], and social networks[7], etc.
- Hawkes processes[5]: ability to model self-exciting and clustering behaviors.
- Real-world data often exhibit dynamic changes over time [8, 10].
- Change Point Detection (CPD): Identifies shifts in underlying process parameters to address time-varying dynamics.
- Existing CPD Methods Limitations: Many methods lack analytical solutions, making them computationally inefficient[2, 6].

Existing CPD Methods and Our Contribution

Our Contribution

- CoBay-CPD Proposal: A conjugate Bayesian two-step method for Hawkes processes using data augmentation, improving accuracy and efficiency in change point detection.
- Analytical Gibbs Sampler: Enables closed-form sampling, reducing computational burden.
- Experimental Results: Demonstrates accurate and timely detection, proving practical for dynamic event modeling across various scenarios.

Hawkes Process with Inhibition

Conditional Intensity Function of Hawkes Process:

$$\lambda^*(t) = \mu + \sum_{t_i<t} \phi(t - t_i) \quad (1)$$

Traditional Hawkes processes capture only excitatory interactions.
Nonlinear Inhibition: To incorporate inhibition, we use a nonlinear model[9]:

$$\lambda^*(t) = \bar{\lambda}\sigma(h(t)), \quad h(t) = \mu + \sum_{t_i<t} \phi(t - t_i).$$

Flexible Influence Function:

$$\phi(\cdot) = \sum_{b=1}^{B} w_b \tilde{\phi}_b(\cdot) \quad (2)$$

Formulation of $h(t)$ and Probability Density Function:

$$h(t) = \mu + \sum_{t_i<t} \phi(t - t_i) = \mu + \sum_{t_i<t} \sum_{b=1}^{B} w_b \tilde{\phi}_b(t - t_i) = \mathbf{w}^\top \mathbf{\Phi}(t) \quad (3)$$

$$p(t_{1:N}|\mathbf{w}, \bar{\lambda}) = \prod_{i=1}^{N} \bar{\lambda}\sigma(h(t_i)) \exp\left(-\int_0^T \bar{\lambda}\sigma(h(t))dt\right) \quad (4)$$

Non-conjugate Bayesian CPD

Estimation step: The likelihood for the timestamps $t_{\tau_m:m}$ after the change point:

$$p(t_{\tau_m:m}|\mathbf{w},\bar{\lambda}) = \prod_{i=\tau_m}^{m} \bar{\lambda}\sigma(h(t_i)) \exp\left(-\int_{t_{\tau_m}}^{t_m} \bar{\lambda}\sigma(h(t))dt\right). \quad (5)$$

According to Bayes' theorem, the posterior of model parameters is expressed as:

$$p(\mathbf{w},\bar{\lambda}|t_{\tau_m:m}) \propto p(t_{\tau_m:m}|\mathbf{w},\bar{\lambda})p(\mathbf{w})p(\bar{\lambda}), \quad (6)$$

where we choose the prior of $\mathbf{w}$ as Gaussian $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0},\mathbf{K})$ and the prior of $\bar{\lambda}$ as an uninformative improper prior $p(\bar{\lambda}) \propto 1/\bar{\lambda}$.

Prediction step: we leverage the posterior of model parameters to compute the predictive distribution of the next timestamp as:

$$p(t_{m+1}|t_{\tau_m:m}) = \iint p(t_{m+1}|t_{\tau_m:m},\mathbf{w},\bar{\lambda})p(\mathbf{w},\bar{\lambda}|t_{\tau_m:m})d\mathbf{w}d\bar{\lambda}. \quad (7)$$

This formula calculates the distribution of the next timestamp t_{m+1} given the observed data points.

Non-conjugate Bayesian CPD

Approximation Method:

1. Use MCMC to sample from the posterior distribution of parameters.
2. Apply the thinning algorithm to sample $\{t_{m+1}^{(k)}\}$, forming a confidence interval. If t_{m+1} falls within it, no change point is detected; otherwise, a change point is inferred.

Challenge:

- For non-conjugate Bayesian CPD, the MCMC algorithm in step 1 lacks analytical solutions, reducing computational efficiency and timeliness.

Conjugate Bayesian CPD (CoBay-CPD)

Data Augmentation: Pólya-Gamma variables and marked Poisson processes.
Augmented Likelihood: After augmentation, the likelihood becomes:

$$p(t_{\tau_m:m}, \boldsymbol{\omega}, \Pi | \mathbf{w}, \bar{\lambda}) = \prod_{i=\tau_m}^{m} [\lambda(t_i, \omega_i) e^{f(\omega_i, h(t_i))}] p_\lambda(\Pi | \bar{\lambda}) \prod_{(\omega, t) \in \Pi} e^{f(\omega, -h(t))} \quad (8)$$

Gibbs Sampling: With conditional conjugacy, we derive closed-form conditional distributions:

$$p(\boldsymbol{\omega} | t_{\tau_m:m}, \mathbf{w}) = \prod_{i=\tau_m}^{m} p_{\mathrm{PG}}(\omega_i | 1, h(t_i)), \quad (9a)$$

$$\Lambda(t, \omega | t_{\tau_m:m}, \mathbf{w}, \bar{\lambda}) = \bar{\lambda} \sigma(-h(t)) p_{\mathrm{PG}}(\omega | 1, h(t)), \quad (9b)$$

$$p(\bar{\lambda} | t_{\tau_m:m}, \Pi) = p_{\mathrm{Ga}}(\bar{\lambda} | N_m + R, T_m), \quad (9c)$$

$$p(\mathbf{w} | t_{\tau_m:m}, \boldsymbol{\omega}, \Pi) = \mathcal{N}(\mathbf{w} | \mathbf{m}, \boldsymbol{\Sigma}). \quad (9d)$$

allowing an efficient Gibbs sampler for posterior sampling.

Baselines and Metrics

Baselines: We compare CoBay-CPD with Bayesian CPD methods addressing non-conjugate inference for Hawkes processes:

- **SMCPD**[4]: Combines BCPD and Sequential Monte Carlo (SMC) .
- **SVCPD**[3]: Combines BCPD and Stein variational inference.
- **SVCPD+Inhibition**: Extends SVCPD to include inhibitory effects in a nonlinear Hawkes process.

Metrics:

- **False Negative Rate (FNR):** Measures the probability of missing a change point, calculated as $1 - \frac{\text{True Positives}}{\text{True Positives+False Negatives}}$.
- **False Positive Rate (FPR):** Measures the probability of incorrectly identifying stable points as change points, calculated as $1 - \frac{\text{True Negatives}}{\text{False Positives+True Negatives}}$.
- **Mean Square Error (MSE):** Assesses prediction accuracy for the next timestamp, calculated as $\frac{1}{n}\sum_{i=1}^{n}(\bar{t}_i^{(k)} - t_i)^2$.
- **Running Time (RT):** Evaluates the efficiency of each method by runtime.

Synthetic Data and Results

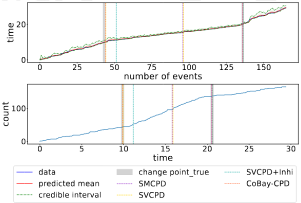
Synthetic Data: The synthetic dataset includes three concatenated Hawkes process segments with varying intensity bounds.

CoBay-CPD Superiority:

- Uses nonlinear Hawkes processes with excitation and inhibition, enhancing model expressiveness.
- Employs Gibbs sampling for accurate parameter estimation.
- Comparisons (SMCPD, SVCPD, SVCPD+Inhibition) rely on linear models and variational methods, reducing accuracy and flexibility.

Table 1: The FNR, FPR, MSE and RT of CoBay-CPD and other baselines on the synthetic dataset.

Model	FNR(↓)	FPR(% ↓)	MSE(↓)	RT(minute ↓)
SMCPD	0.38 ± 0.41	0.76 ± 0.26	0.07 ± 0.01	5.50 ± 0.31
SVCPD	0.50 ± 0.35	0.76 ± 0.26	0.06 ± 0.00	7.78 ± 0.01
SVCPD+Inhi	0.33 ± 0.24	0.60 ± 0.00	0.16 ± 0.01	23.09 ± 0.60
CoBay-CPD	$\mathbf{0.13 \pm 0.22}$	$\mathbf{0.46 \pm 0.26}$	$\mathbf{0.05 \pm 0.00}$	$\mathbf{4.62 \pm 0.10}$



Real Data and Results

WannaCry Cyber Attack: Over 200,000 computers infected worldwide in 2017, data includes 208 traffic log observations with timestamps.
NYC Vehicle Collisions: Dataset with approximately 1.05 million records; data from Oct.14th, 2017 was used in experiments.

Table 2: The FNR, FPR, MSE and RT of CoBay-CPD and other baselines on real-world datasets.

Model	WannaCry				NYC Vehicle Collisions			
	FNR(↓)	FPR(↓)	MSE($\times 10^2$ ↓)	RT(minute ↓)	FNR(↓)	FPR(% ↓)	MSE(↓)	RT(minute ↓)
SMCPD	0.38 ± 0.06	0.02 ± 0.01	3.59 ± 0.08	11.65 ± 0.07	0.56 ± 0.16	2.46 ± 0.55	0.02 ± 0.00	24.67 ± 0.26
SVCPD	0.34 ± 0.12	0.01 ± 0.01	3.47 ± 0.06	9.72 ± 0.06	0.58 ± 0.36	1.00 ± 0.43	0.02 ± 0.00	19.30 ± 0.09
SVCPD+Inhi	0.54 ± 0.09	$\mathbf{0.00 \pm 0.00}$	3.54 ± 0.06	29.76 ± 2.54	0.22 ± 0.16	1.55 ± 0.36	0.17 ± 0.01	64.47 ± 1.36
CoBay-CPD	$\mathbf{0.21 \pm 0.04}$	0.05 ± 0.02	$\mathbf{3.42 \pm 0.00}$	$\mathbf{6.24 \pm 0.49}$	$\mathbf{0.13 \pm 0.16}$	$\mathbf{0.89 \pm 0.16}$	$\mathbf{0.01 \pm 0.00}$	$\mathbf{8.70 \pm 0.26}$

Results Summary

- **WannaCry Data Result:** CoBay-CPD performs best wrt FNR, FPR balance, MSE, and runtime.
- **NYC Data Result:** CoBay-CPD performs best wrt FNR, FPR, MSE, and runtime.

Ablation Study and Stress tests

Ablation Study: Number of Basis Functions, Confidence Interval, Confidence Interval.

Table 3: Ablation study. The FNR, FPR, MSE and RT of CoBay-CPD with different hyperparameters.

Metric	Number of Basis Functions			Confidence Interval			Prior Covariance		
	1	2	3	95%	90%	85%	$\sigma^2 = 0.01$	$\sigma^2 = 0.5$	$\sigma^2 = 10$
FNR(↓)	0.38 ± 0.41	0.38 ± 0.22	**0.13 ± 0.22**	0.50 ± 0.00	**0.13 ± 0.22**	0.25 ± 0.25	**0.13 ± 0.22**	**0.13 ± 0.22**	0.50 ± 0.00
FPR(% ↓)	1.07 ± 0.50	0.91 ± 0.30	**0.61 ± 0.00**	**0.46 ± 0.26**	**0.46 ± 0.26**	1.83 ± 0.43	0.76 ± 0.26	**0.46 ± 0.26**	0.91 ± 0.30
MSE(↓)	**0.05 ± 0.00**	**0.05 ± 0.00**	**0.05 ± 0.00**	**0.04 ± 0.00**	0.05 ± 0.00	**0.04 ± 0.00**	**0.04 ± 0.00**	0.05 ± 0.00	0.05 ± 0.01
RT(minute ↓)	**1.57 ± 0.03**	2.61 ± 0.08	3.62 ± 0.10	5.03 ± 0.02	4.62 ± 0.10	**4.50 ± 0.11**	4.74 ± 0.02	4.62 ± 0.10	**4.41 ± 0.10**

Stress Tests: number of change points, number of change points, closeness between adjacent change points(Δt).

Table 5: The FNR, FPR and MSE of CoBay-CPD and other baselines on synthetic dataset with different number of change points.

Model	1			2			3		
	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)
SMCPD	0.33 ± 0.47	0.63 ± 0.63	0.08 ± 0.03	0.38 ± 0.41	0.76 ± 0.26	0.07 ± 0.01	0.67 ± 0.19	1.23 ± 0.35	0.08 ± 0.01
SVCPD	0.67 ± 0.47	0.63 ± 0.95	0.06 ± 0.01	0.50 ± 0.35	0.76 ± 0.26	0.06 ± 0.00	0.50 ± 0.32	1.74 ± 0.65	0.15 ± 0.05
SVCPD+Inhi	0.33 ± 0.47	1.88 ± 0.63	0.08 ± 0.01	0.33 ± 0.24	0.60 ± 0.00	0.16 ± 0.01	0.28 ± 0.23	1.84 ± 0.50	0.09 ± 0.00
CoBay-CPD	**0.00 ± 0.00**	**0.43 ± 0.60**	**0.04 ± 0.00**	**0.13 ± 0.22**	**0.46 ± 0.26**	**0.05 ± 0.00**	**0.11 ± 0.14**	**0.31 ± 0.50**	**0.07 ± 0.01**

Table 6: The FNR, FPR and MSE of CoBay-CPD and other baselines on synthetic dataset with different difference between adjacent $\bar{\lambda}$'s ($\Delta\bar{\lambda}$).

Model	0.1			1			5		
	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)
SMCPD	1.00 ± 0.00	**1.20 ± 0.00**	0.06 ± 0.01	0.50 ± 0.50	0.70 ± 0.70	0.06 ± 0.02	0.33 ± 0.47	0.63 ± 0.63	0.08 ± 0.03
SVCPD	1.00 ± 0.00	2.41 ± 0.98	0.05 ± 0.01	0.83 ± 0.37	3.29 ± 1.05	0.06 ± 0.01	0.67 ± 0.47	0.63 ± 0.95	0.06 ± 0.01
SVCPD+Inhi	**0.67 ± 0.47**	1.41 ± 0.83	0.06 ± 0.00	0.33 ± 0.47	1.17 ± 0.52	0.06 ± 0.00	0.33 ± 0.47	1.88 ± 0.63	0.08 ± 0.01
CoBay-CPD	1.00 ± 0.00	1.61 ± 0.57	**0.02 ± 0.00**	**0.25 ± 0.43**	**0.35 ± 0.60**	**0.03 ± 0.00**	**0.00 ± 0.00**	**0.43 ± 0.60**	**0.04 ± 0.00**

Table 7: The FNR, FPR and MSE of CoBay-CPD and other baselines on synthetic dataset with different closeness between two change points (Δt).

Model	5			10			15		
	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)	FNR(↓)	FPR(% ↓)	MSE(↓)
SMCPD	0.42 ± 0.34	**0.75 ± 0.75**	0.03 ± 0.01	0.67 ± 0.24	**0.23 ± 0.52**	0.05 ± 0.01	0.17 ± 0.24	1.00 ± 0.83	0.07 ± 0.01
SVCPD	0.42 ± 0.19	1.24 ± 0.56	0.03 ± 0.01	0.75 ± 0.25	0.46 ± 0.65	0.05 ± 0.01	**0.08 ± 0.19**	3.01 ± 1.15	0.06 ± 0.01
SVCPD+Inhi	0.58 ± 0.19	1.24 ± 1.33	0.05 ± 0.01	0.25 ± 0.38	**0.23 ± 0.52**	0.05 ± 0.00	0.17 ± 0.24	2.01 ± 1.33	0.06 ± 0.00
CoBay-CPD	**0.33 ± 0.24**	1.00 ± 0.70	**0.01 ± 0.00**	**0.00 ± 0.00**	0.93 ± 0.65	**0.02 ± 0.00**	**0.08 ± 0.19**	**0.80 ± 0.57**	**0.03 ± 0.00**

Conclusion

- Introduced a novel conjugate Bayesian two-step change point detection method for Hawkes processes, addressing the non-conjugate inference challenge.
- Used data augmentation to convert the problem to a conditionally conjugate form, allowing for efficient Gibbs sampling.
- Outperformed existing methods in accuracy and efficiency for change point detection.
- Contributions offer significant potential for advancing event-driven time series analysis across diverse applications.

[1] Emmanuel Bacry, Iacopo Mastromatteo, and Jean-François Muzy.
Hawkes processes in finance.
Market Microstructure and Liquidity, 1(01):1550005, 2015.

[2] David M Blei, Alp Kucukelbir, and Jon D McAuliffe.
Variational inference: A review for statisticians.
Journal of the American Statistical Association, 112(518):859–877, 2017.

[3] Gianluca Detommaso, Hanne Hoitzing, Tiangang Cui, and Ardavan Alamir.
Stein variational online changepoint detection with applications to hawkes processes and neural networks.
arXiv preprint arXiv:1901.07987, 2019.

[4] Arnaud Doucet, Adam M Johansen, et al.
A tutorial on particle filtering and smoothing: Fifteen years later.
Handbook of nonlinear filtering, 12(656-704):3, 2009.

[5] Alan G Hawkes.
Spectra of some self-exciting and mutually exciting point processes.
Biometrika, 58(1):83–90, 1971.

[6] Radford M Neal.
Probabilistic inference using markov chain monte carlo methods.
1993.

[7] Julio Cesar Louzada Pinto, Tijani Chahed, and Eitan Altman.
Trend detection in social networks using Hawkes processes.
In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 1441–1448. ACM, 2015.

[8] Haoyun Wang, Liyan Xie, Yao Xie, Alex Cuozzo, and Simon Mak.
Sequential change-point detection for mutually exciting point processes.
Technometrics, 65(1):44–56, 2023.

[9] Feng Zhou, Quyu Kong, Zhijie Deng, Jichao Kan, Yixuan Zhang, Cheng Feng, and Jun Zhu.
Efficient inference for dynamic flexible interactions of neural populations.
Journal of Machine Learning Research, 23(211):1–49, 2022.

[10] Feng Zhou, Zhidong Li, Xuhui Fan, Yang Wang, Arcot Sowmya, and Fang Chen.
Fast multi-resolution segmentation for nonstationary Hawkes process using cumulants.
International Journal of Data Science and Analytics, 10:321–330, 2020.

[11] Feng Zhou, Yixuan Zhang, and Jun Zhu.
Efficient inference of flexible interaction in spiking-neuron networks.
In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

Thanks!