Recent Developments in Learning Hawkes Processes

Hongteng Xu

Department of ECE
Duke University

hongteng.xu@duke.edu

March 11, 2023

Overview

- **Event sequence:** $S = \{(t_i, d_i)\}_{i=1}^{I}$, $d_i \in \mathcal{D}$.
- **Counting processes:** $N = \{N_d(t)\}_{d=1}^{D}$. $N_d(t)$ is the number of type-d events occurring till time t.
- **Intensity function:** The expected instantaneous happening rate of type-u events given the history.

$$\lambda_d(t) = \frac{\mathbb{E}[dN_d(t)|\mathcal{H}_t]}{dt},\ \mathcal{H}_t = \{(t_i, d_i)|t_i < t, d_i \in \mathcal{D}\}.$$

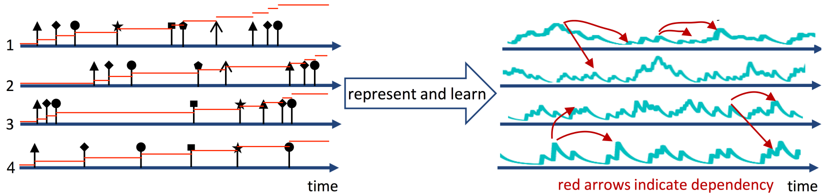


Figure: Event sequences and intensity functions.

Hawkes Processes

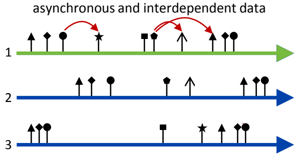
A multi-dimensional Hawkes process has a particular form of intensity:

$$\begin{aligned}\lambda_d(t) &= \mu_d + \sum\nolimits_{d'=1}^{D} \int_0^t \phi_{dd'}(t-s)dN_{d'}(s) \\ &= \mu_d + \sum\nolimits_{(t_i,d_i)\in\mathcal{H}_t} \phi_{dd_i}(t-t_i).\end{aligned} \tag{1}$$

- $\boldsymbol{\mu} = [\mu_d]$: the exogenous fluctuation of the system.
- $\boldsymbol{\Phi} = [\phi_{dd'}(t)]$: the impact functions measuring the endogenous triggering pattern of the type-d' events on the type-d' events.
- $\sum_{(t_i,d_i)\in\mathcal{H}_t} \phi_{dd_i}(t-t_i)$: the accumulated endogenous intensity caused by history.

Typical Examples

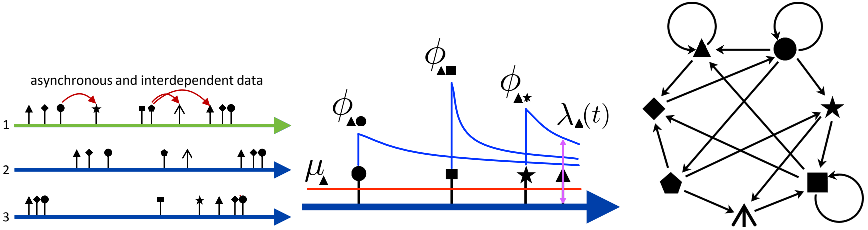
Scene	Entities	Sequences	Task
Patient admission	Diseases	Patients' admission records	Disease network
Job hopping	Companies	LinkedIn users' job history	Competition network
Online shopping	Items	Buying/rating behaviors	Recommendation



(a) Observations

Typical Examples

Scene	Entities	Sequences	Task
Patient admission	Diseases	Patients' admission records	Disease network
Job hopping	Companies	LinkedIn users' job history	Competition network
Online shopping	Items	Buying/rating behaviors	Recommendation



(a) Observations (b) Intensity per entity (c) Endogenous graph

Figure: Illustration of Hawkes process model.

Learning Methods

- **Maximum Likelihood Estimation (MLE)**
- **Least Squares Estimation (LS)**
- Wiener-Hopf Equations
- Cumulants-based Methods
- ...

Maximum Likelihood Estimation (MLE)

Given conditional intensity function,

- For each sequence $S_m = \{(t_i, d_i)\}_{i=1}^{I}$, the conditional probability of event is

$$p((t,d)|\mathcal{H}_t) = \lambda_d(t)\exp\left(-\sum_{d'=1}^{D}\int_{t_0}^{t}\lambda_{d'}(s)ds\right). \tag{2}$$

- For a set of event sequences $\mathcal{S} = \{S_m\}_{m=1}^{M}$. $S_m = \{(t_i^m, d_i^m)\}_{i=1}^{I}$, the log-likelihood function is

$$\mathcal{L}(\boldsymbol{\theta};\mathcal{S}) = \sum_{m=1}^{M}\left\{\sum_{i=1}^{I}\log\lambda_{d_i^m}(t_i^m) - \sum_{d=1}^{D}\int_0^T\lambda_d(s)ds\right\}. \tag{3}$$

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta};\mathcal{S}). \tag{4}$$

Least Squares Estimation (LS)

Assuming $\phi_{dd'}(t) = a_{dd'}\kappa(t)$, we can learn Hawkes processes as a linear predictor:

$$\begin{aligned}\lambda_d(t) &= \mu_d + \sum\nolimits_{t_i<t} a_{dd_i}\kappa(t-t_i)\\ &= \boldsymbol{e}_d^\top \boldsymbol{\mu} + vec(\boldsymbol{K}(t))^\top vec(\boldsymbol{A}) = \boldsymbol{x}_d^\top(t)\boldsymbol{\theta}.\end{aligned} \tag{5}$$

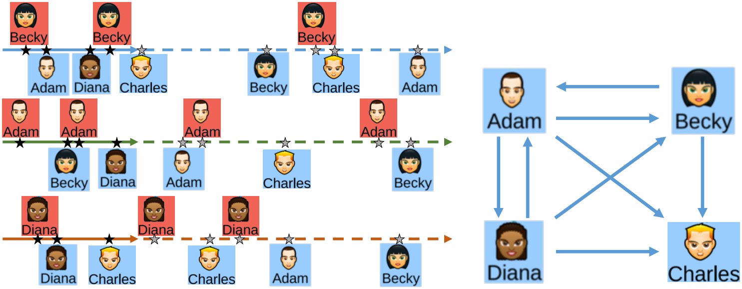
$\boldsymbol{\theta} = [\boldsymbol{\mu}; \mathrm{vec}(\boldsymbol{A})] \in \mathbb{R}^{D(1+D)}$, $\boldsymbol{x}_d(t) = [\boldsymbol{e}_d; \mathrm{vec}(\boldsymbol{K}(t))]$.
$\boldsymbol{K}(t) = [k_{dd'}(t)] \in \mathbb{R}^{D\times D}$, $k_{dd'}(t) = \sum_{(t_i,d_i)\in\mathcal{H}_t,\ d_i=d'} \kappa(t-t_i)$.

$$\begin{aligned}&R(\boldsymbol{\theta}; \boldsymbol{N}) = \mathbb{E}\Big[\frac{1}{t^2}\Big(N(t) - \int_0^t \lambda(s)ds\Big)^2\Big]\\ &= \frac{1}{MI}\sum\nolimits_{m=1}^M \sum\nolimits_{i=1}^I \frac{1}{(t_i^m)^2}\Big|N_{d_i^m}^m(t_i^m) - \int_0^{t_i^m} \lambda_{d_i^m}(s)ds\Big|^2 = \|\boldsymbol{N} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2.\end{aligned} \tag{6}$$

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} R(\boldsymbol{\theta}; \boldsymbol{N}).$$

Challenges of real-world data

1. Extremely-short observations
2. Multiple exogenous sources



(a) Extremely-short and multi-source data (b) Endogenous graph

Figure: Illustration the challenges.

How to learn the endogenous triggering pattern robustly?

Strategy 1: Single source + HP

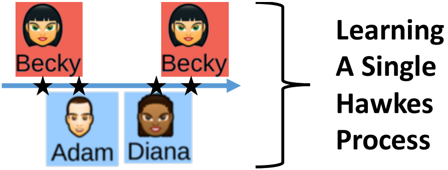


Figure: Single source data + HP.

Strategy 1: Single source + HP

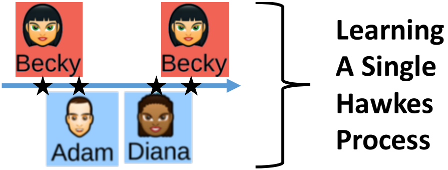


Figure: Single source data + HP.

$$\begin{aligned}
&\min_{\boldsymbol{\theta}_{single}} R(\boldsymbol{\theta}_{single}; \boldsymbol{N}_{single}),\\
&\boldsymbol{\theta}_{single} = [\boldsymbol{\mu}; vec(\boldsymbol{A})],\\
&R(\theta_{single}; \boldsymbol{N}_{single}) = \|\boldsymbol{N}_{single} - \boldsymbol{X}_{single}\boldsymbol{\theta}_{single}\|_2^2
\end{aligned} \tag{7}$$

Risk: Over-fitting

Strategy 2: Multi-source + HP

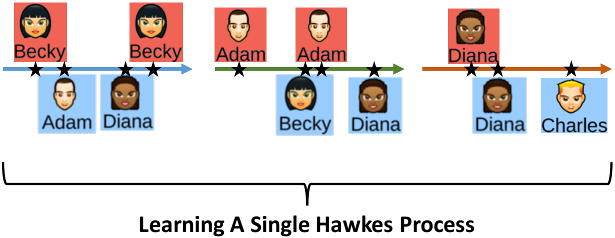


Figure: Multi-source data + HP.

Strategy 2: Multi-source + HP

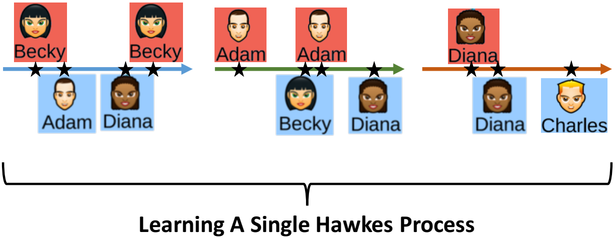


Figure: Multi-source data + HP.

$$\min_{\boldsymbol{\theta}_{single}} R(\boldsymbol{\theta}_{single}; \boldsymbol{N}_{multi}),$$
$$\boldsymbol{\theta}_{single} = [\boldsymbol{\mu}; vec(\boldsymbol{A})], \quad \boldsymbol{N}_{multi} = [\boldsymbol{N}^1; ...; \boldsymbol{N}^M] \tag{8}$$
$$R(\theta_{single}; \boldsymbol{N}_{multi}) = \|\boldsymbol{N}_{multi} - \boldsymbol{X}_{multi}\boldsymbol{\theta}_{single}\|_2^2 = \sum_{m=1}^{M} \|\boldsymbol{N}^m - \boldsymbol{X}^m\boldsymbol{\theta}_{single}\|_2^2$$

Risk: Model misspecification

Strategy 3: Multi-source + MHP

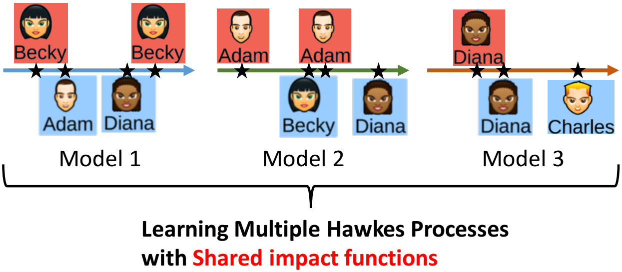


Figure: Multi-source data + Multi-HPs.

Strategy 3: Multi-source + MHP

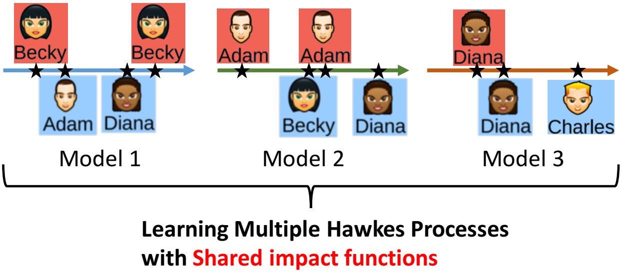


Figure: Multi-source data + Multi-HPs.

$$\begin{aligned} &\min_{\boldsymbol{\theta}_{multi}} R(\boldsymbol{\theta}_{multi}; \boldsymbol{N}_{multi}), \\ &\boldsymbol{\theta}_{multi} = [\boldsymbol{\mu}^1; ...; \boldsymbol{\mu}^M; vec(\boldsymbol{A})], \quad \boldsymbol{N}_{multi} = [\boldsymbol{N}^1; ...; \boldsymbol{N}^M] \\ &R(\theta_{multi}; \boldsymbol{N}_{multi}) = \|\boldsymbol{N}_{multi} - \boldsymbol{X}_{multi}\boldsymbol{\theta}_{multi}\|_2^2 \end{aligned} \tag{9}$$

Question: Can we do better?

Our strategy: Superposition + HP

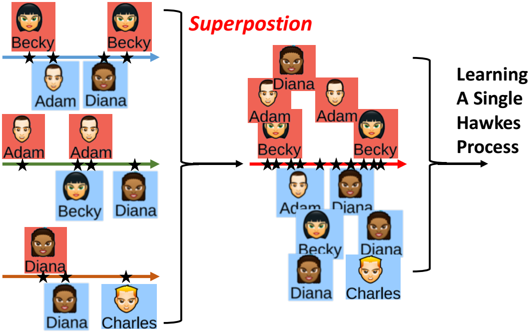


Figure: Superposed data + HP.

Our strategy: Superposition + HP

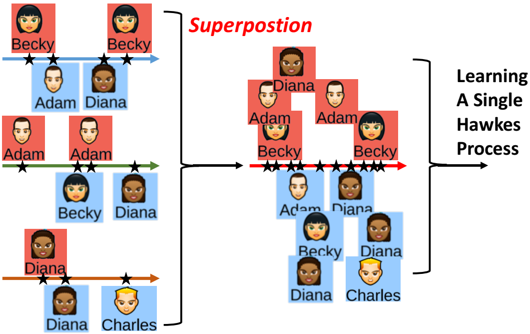


Figure: Superposed data + HP.

$$\boldsymbol{N}_{super}(t) = \sum\nolimits_{m=1}^{M} \boldsymbol{N}^m(t)$$

$$\min_{\boldsymbol{\theta}_{super}} R(\boldsymbol{\theta}_{super}; \boldsymbol{N}_{super}), \quad \boldsymbol{\theta}_{super} = [\sum\nolimits_{m=1}^{M} \boldsymbol{\mu}^m; vec(\boldsymbol{A})], \tag{10}$$

$$R(\theta_{super}; \boldsymbol{N}_{super}) = \frac{1}{M^2} \|\boldsymbol{N}_{super} - \boldsymbol{X}_{super}\boldsymbol{\theta}_{super}\|_2^2$$

The property of superposed Hawkes processes

Theorem (Property 1)

The superposition of M independent Hawkes processes, where $N^m(t) \sim HP(\boldsymbol{\mu}^m, \boldsymbol{\Phi})$ for $m = 1, ..., M$, is still a Hawkes process, where $N(t) = \sum_{m=1}^{M} N^m(t)$ and $N(t) \sim HP(\sum_{m=1}^{M} \boldsymbol{\mu}^m(t), \boldsymbol{\Phi})$.

The property of superposed Hawkes processes

Theorem (Property 1)

The superposition of M independent Hawkes processes, where $N^m(t) \sim HP(\boldsymbol{\mu}^m, \boldsymbol{\Phi})$ for $m = 1, ..., M$, is still a Hawkes process, where $N(t) = \sum_{m=1}^{M} N^m(t)$ and $N(t) \sim HP(\sum_{m=1}^{M} \boldsymbol{\mu}^m(t), \boldsymbol{\Phi})$.

Proof.

$$
\begin{aligned}
\lambda_d(t) =& \frac{\mathbb{E}[dN_d(t)|\mathcal{H}_t]}{dt} = \sum_{m=1}^{M} \frac{\mathbb{E}[dN_d^m(t)| \cup_{l=1}^{M} \mathcal{H}_t^l]}{dt} \\
=& \sum_{m=1}^{M} \frac{\mathbb{E}[dN_d^m(t)|\mathcal{H}_t^m]}{dt} = \sum_{m=1}^{M} \lambda_d^m(t). \\
=& \sum_{m=1}^{M} \Big(\mu_d^m(t) + \sum_{(t_i^m, d_i^m)\in\mathcal{H}_t^m} \phi_{dd_i^m}(t - t_i^m)\Big) \\
=& \sum_{m=1}^{M} \mu_d^m(t) + \sum_{(t_i, d_i)\in\mathcal{H}_t} \phi_{dd_i}(t - t_i),
\end{aligned}
\tag{11}
$$

□

Why and when does our strategy work?

Theorem (Property 2)

Suppose that we have M independent and stationary D-dimensional Hawkes processes with shared impact functions, i.e., $\{HP(\boldsymbol{\mu}^m, \boldsymbol{A})\}_{m=1}^M$, where the parameters are bounded as $\|\boldsymbol{\mu}^m\|_2^2 \leq B_\mu$ and $\|vec(\boldsymbol{A})\|_2^2 \leq B_A$. Each of them has an observed event sequence with I events. Then the bound on the excess risk $\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})]$ is tighter than that of $\mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]$ when the upper bound of $\|\sum_{m=1}^M \boldsymbol{\mu}^m\|_2^2$, denoted as $B_{\Sigma\mu}$, satisfies

$$\begin{aligned} B_{\Sigma\mu} \leq & MB_\mu + D(M+D)B_\mu \log\Big(1 + \frac{MI}{D(M+D)}\Big) \\ & - D(1+D)B_\mu \log\Big(1 + \frac{MI}{D(1+D)}\Big). \end{aligned} \tag{12}$$

Here $\boldsymbol{\theta}^*$ represents the ground truth of parameters.

Proof of Property 2

For the linear predictor $\hat{\boldsymbol{\theta}}$ learned by minimizing the squared loss $R(\boldsymbol{\theta}) = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|_2^2$, where $\boldsymbol{\theta} \in \{\boldsymbol{\theta} \in \mathbb{R}^C : \|\boldsymbol{\theta}\|_2^2 \leq B\}$ and the M observations $\boldsymbol{y} = [y_1; ...; y_M]$ satisfy $y_i \in \{y : |y| \leq Y\}$, we have [Shamir(2015)]

$$\mathbb{E}[R(\hat{\boldsymbol{\theta}}) - R(\boldsymbol{\theta}^*)] \leq \mathcal{O}\Big(\frac{B + CY^2 \log(1 + \frac{M}{C})}{C}\Big). \tag{13}$$

Additionally, we have $\lim_{t\rightarrow\infty} \frac{N_d(t)}{t} = \frac{\mu_d}{1-\|\boldsymbol{A}\|_2}$ and $\|\boldsymbol{A}\|_2 \ll 1$ for the stationarity.

Table: Plug our models to (13).

Parameters	Multi-source+MHP	Superposition+HP
# Samples, M	MI	MI
# Variables, C	$D(M+D)$	$D(1+D)$
$\sup \lvert y\rvert^2$, Y^2	$\mathcal{O}(B_\mu)$	$\mathcal{O}(B_\mu)$
$\sup \|\boldsymbol{\theta}\|_2^2$, B	$B_A + MB_\mu$	$B_A + B_{\Sigma\mu}$

Proof of Property 2

$$\mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]$$
$$\leq \mathcal{O}\Big(\frac{B_A + MB_\mu + D(M+D)B_\mu \log(1 + \frac{MI}{D(M+D)})}{MI}\Big),$$

$$\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})]$$
$$\leq \mathcal{O}\Big(\frac{B_A + B_{\Sigma\mu} + D(1+D)B_\mu \log(1 + \frac{MI}{D(1+D)})}{MI}\Big).$$

Proof of Property 2

$$\begin{aligned}&\mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]\\ &\leq \mathcal{O}\Big(\frac{B_A + MB_\mu + D(M+D)B_\mu \log(1+\frac{MI}{D(M+D)})}{MI}\Big),\end{aligned}$$

$$\begin{aligned}&\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})]\\ &\leq \mathcal{O}\Big(\frac{B_A + B_{\Sigma\mu} + D(1+D)B_\mu \log(1+\frac{MI}{D(1+D)})}{MI}\Big).\end{aligned}$$

$$\begin{aligned}&\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})] \leq \mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]\\ &\rightarrow\\ &B_{\Sigma\mu} + D(1+D)B_\mu \log(1+\frac{MI}{D(1+D)})\\ &\leq MB_\mu + D(M+D)B_\mu \log(1+\frac{MI}{D(M+D)}).\end{aligned} \tag{14}$$

Typical Infeasible Condition

Lemma (Typical Infeasible Condition)

For the Hawkes processes with the same exogenous intensity and endogenous impact functions, the superposition-based strategy is inefficient, $\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})] \geq \mathbb{E}[R_{single}(\hat{\boldsymbol{\theta}}) - R_{single}(\boldsymbol{\theta}^*)]$.

Typical Infeasible Condition

Lemma (Typical Infeasible Condition)

For the Hawkes processes with the same exogenous intensity and endogenous impact functions, the superposition-based strategy is inefficient, $\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})] \geq \mathbb{E}[R_{single}(\hat{\boldsymbol{\theta}}) - R_{single}(\boldsymbol{\theta}^*)]$.

Proof.

$\boldsymbol{\mu}^1 = \ldots = \boldsymbol{\mu}^M = \boldsymbol{\mu}$ and $\|\sum_{m=1}^M \boldsymbol{\mu}^m\|_2^2 = M^2\|\boldsymbol{\mu}\|_2^2 \leq M^2 B_\mu = B_{\Sigma\mu}$.

$$\mathbb{E}[R_{single}(\hat{\boldsymbol{\theta}}) - R_{single}(\boldsymbol{\theta}^*)] \leq \mathcal{O}\Big(\frac{B_A + B_\mu + D(1+D)B_\mu \log(1+\frac{MI}{D(1+D)})}{MI}\Big)$$

$$\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})]$$
$$\leq \mathcal{O}\Big(\frac{B_A + M^2 B_\mu + D(1+D)B_\mu \log(1+\frac{MI}{D(1+D)})}{MI}\Big).$$

$\square$

Typical Feasible Condition

Lemma (Typical Feasible Condition)

For the Hawkes processes with complementary exogenous intensities, i.e., $\{HP(\boldsymbol{\mu}^m, \boldsymbol{\Phi})\}_{m=1}^M$ *and* $supp(\boldsymbol{\mu}^m) \cap supp(\boldsymbol{\mu}^{m'}) = \emptyset$ *for all* $m \neq m'$*, the superposition-based strategy always provides us with benefits on efficiency, i.e.,* $\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})] \leq \mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]$.

Typical Feasible Condition

Lemma (Typical Feasible Condition)

For the Hawkes processes with complementary exogenous intensities, i.e., $\{HP(\boldsymbol{\mu}^m, \boldsymbol{\Phi})\}_{m=1}^M$ and $supp(\boldsymbol{\mu}^m) \cap supp(\boldsymbol{\mu}^{m'}) = \emptyset$ for all $m \neq m'$, the superposition-based strategy always provides us with benefits on efficiency, i.e., $\mathbb{E}[R_{super}(\hat{\boldsymbol{\theta}}_{super}) - R_{super}(\boldsymbol{\theta}^*_{super})] \leq \mathbb{E}[R_{multi}(\hat{\boldsymbol{\theta}}_{multi}) - R_{multi}(\boldsymbol{\theta}^*_{multi})]$.

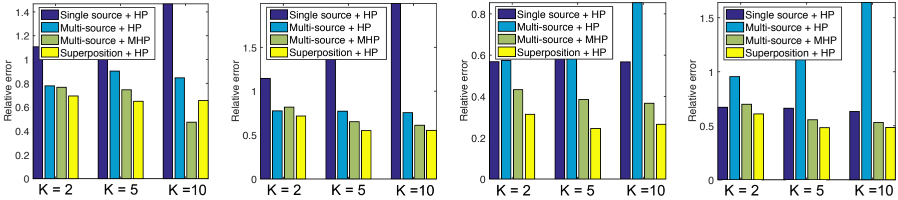
Proof.

$\|\sum_{m=1}^M \boldsymbol{\mu}^m\|_2^2 = \sum_{m=1}^M \|\boldsymbol{\mu}^m\|_2^2 \leq MB_\mu = B_{\Sigma\mu}$. Plugging the upper bound into the condition (12), we have

$$\begin{aligned} MB_\mu \leq & MB_\mu + D(M+D)B_\mu \log\Big(1 + \frac{MI}{D(M+D)}\Big) \\ & - D(1+D)B_\mu \log\Big(1 + \frac{MI}{D(1+D)}\Big). \end{aligned}$$

□

Given K D-dimensional Hawkes process models, we generate 20 event sequences for each. These Hawkes processes share the same impact functions, which are parameterized as an infectivity matrix $\boldsymbol{A} \in \mathbb{R}^{D\times D}$. The exogenous intensity $\boldsymbol{\mu}$ of each Hawkes process is a random sparse vector, in which only one element is nonzero (Imitation of behaviors in social networks).



(a) LS, $D = 5$ (b) LS, $D = 10$ (c) MLE, $D = 5$ (d) MLE, $D = 10$

Figure: Estimation errors of impact functions obtained by various methods.

Applications to Cold-Start of Recommendation Systems

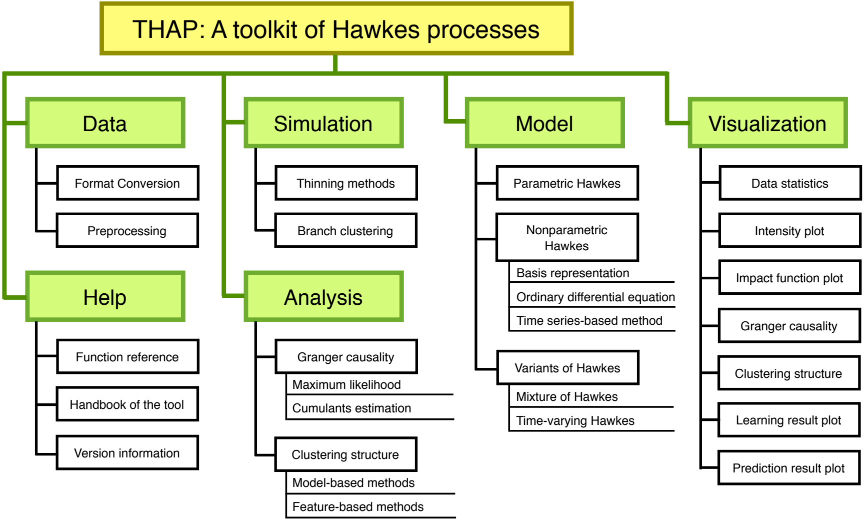
Given users' buying-and-rating behaviors (< 3) from January 2014 to April 2014 (Amazon product data), we aim to predict (recommend) items for them. Because during this period only one or two buying behaviors happened, this is a typical cold-start problem.

$$d_{next} = \arg\max_{d \in \mathcal{D}} \sum\nolimits_{(t_i, d_i) \in \mathcal{H}_t} a_{dd_i} \exp(-w(t - t_i)). \tag{15}$$

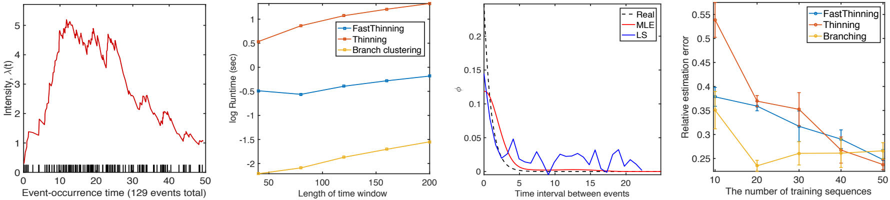
Table: Summary of the performance for various methods.

Method		MostPopular			BPR			FPMC			Multi-source+MHP			Superposition+HP		
Metric		$P@N$	$R@N$	$F_1@N$	$P@N$	$R@N$	$F_1@N$	$P@N$	$R@N$	$F_1@N$	$P@N$	$R@N$	$F_1@N$	$P@N$	$R@N$	$F_1@N$
Top5	Baby	0.145	0.726	0.242	0.306	1.532	0.511	**0.484**	**2.419**	**0.806**	0.339	1.694	0.565	0.306	1.532	0.511
	Garden	0.277	1.385	0.462	0.646	3.231	1.077	0.277	1.385	0.462	0.739	3.692	1.231	**1.046**	**5.231**	**1.744**
	Pet	0.517	2.585	0.862	0.526	2.632	0.877	0.517	2.585	0.862	0.780	3.900	1.300	**0.864**	**4.323**	**1.441**
Top10	Baby	0.234	2.339	0.425	**0.379**	**3.790**	**0.689**	0.307	3.065	0.557	0.218	2.177	0.396	0.282	2.822	0.513
	Garden	0.246	2.462	0.448	0.431	4.308	0.783	0.308	3.077	0.559	0.646	6.461	1.174	**0.800**	**8.000**	**1.454**
	Pet	0.371	3.712	0.675	0.428	4.276	0.778	0.470	4.700	0.854	0.549	5.498	1.000	**0.630**	**6.297**	**1.145**
Top20	Baby	0.335	6.694	0.638	0.294	5.887	0.561	**0.339**	**6.774**	**0.645**	0.194	3.871	0.369	0.254	5.081	0.484
	Garden	0.369	7.385	0.703	0.431	8.615	0.821	0.300	6.000	0.571	0.439	8.769	0.835	**0.508**	**10.154**	**0.967**
	Pet	0.374	7.472	0.712	0.465	9.305	0.886	0.371	7.425	0.707	0.338	6.767	0.645	**0.489**	**9.774**	**0.931**

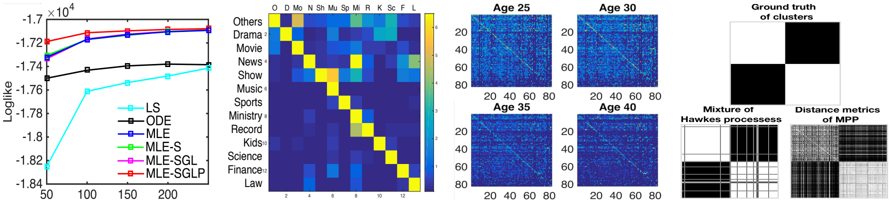
THAP: Package Architecture



Functions and Applications



(a) Data, intensity (b) Simulation time (c) Impact function (d) Estimation errors



(e) Log-likelihood (f) Granger causality (g) Infectivity (h) Clustering

Figure: Visualization of typical functions achieved by *THAP*

Comparisons with existing toolkits

Table: Models and algorithms of Hawkes processes in different toolkits.

Model	Type	Parametric		Nonparametric	
	Impact function	Exponential	Gaussian	Smooth	Discrete
Simulator	Branch clustering	★■♣	★♣	★♣	
	(Fast) Thinning	★♦♣♠	★♣♠	★♣♠	
Learning	MLE(+Regularizer)	★■♣♠	★♣♠	★♣♠	
	MLE + ODE	★♣♠	★♣♠	★♣♠	★♣♠
	Least-squares	★			★
Analysis	Granger causality	★■♣♠	★♣♠	★♣♠	♠
	Mixture model	★	★	★	
	Distance metric	★	★	★	★
	Time-Varying HP	★♣	★♣	★♣	

★ = Proposed *THAP* [Xu and Zha(2017)b]
♦ = *R-hawkes* [Da Fonseca and Zaatour(2014)]
■ = *pyhawkes* [Linderman and Adams(2014)]
♣ = *PtPack* [Du(2016)]
♠ = *tick* [Bacry et al.(2017)]

Summary

- Hawkes process is a powerful tool to capture the time-dependent mechanism hidden in real-world data
- Robust learning from imperfect (real-world) observations is an important issue. Data-based solutions have potentials to suppress, even solve it.
- A Matlab-based toolkit for learning Hawkes processes is developed for the education and the research in the field of statistical machine learning.
- Link of THAP: `https://github.com/HongtengXu/Hawkes-Process-Toolkit`
- Homepage: `https://sites.google.com/view/hongtengxu`

References

Emmanuel Bacry, Martin Bompaire, Stéphane Gaïffas, and Soren Poulsen.
tick: a python library for statistical learning, with an emphasis on time-dependent modeling. *arXiv:1707.03003*, 2017.

José Da Fonseca and Riadh Zaatour.
Hawkes process: Fast calibration, application to trade clustering, and diffusive limit. *Journal of Futures Markets*, 2014.

Michael Eichler, Rainer Dahlhaus, and Johannes Dueck.
Graphical modeling for multivariate hawkes processes with nonparametric link functions. *Time Series Analysis*, 2017.

Alan Hawkes.
Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society. Series B*, 1971.

Shamir, Ohad.
The sample complexity of learning linear predictors with the squared loss. *Journal of Machine Learning Research*, 2015.

Scott Linderman and Ryan Adams.
Discovering latent network structure in point process data. In *ICML*, 2014.

Dixin Luo, Hongteng Xu, Yi Zhen, Xia Ning, Hongyuan Zha, Xiaokang Yang, and Wenjun Zhang.
Multi-task multi-dimensional hawkes processes for modeling event sequences. In *IJCAI*, 2015.

Yosihiko Ogata.
On lewis' simulation method for point processes. *IEEE Transactions on Information Theory*, 1981.

Nan Du.
Modeling, Learning, and Inference of High-Dimensional Asynchronous Event Data. *PhD Thesis*, 2016.

Hongteng Xu and Hongyuan Zha.
A dirichlet mixture model of hawkes processes for event sequence clustering. In *NIPS*, 2017.

Hongteng Xu and Hongyuan Zha.
THAP: a Matlab toolkit for learning with Hawkes processes. *arXiv:1708.09252*, 2017.

Hongteng Xu, Dixin Luo, and Hongyuan Zha.
Learning hawkes processes from short doubly-censored event sequences. In *ICML*, 2017.

Ke Zhou, Hongyuan Zha, and Le Song.
Learning triggering kernels for multi-dimensional hawkes processes. In *ICML*, 2013.

Hongteng Xu, Lawrence Carin, and Hongyuan Zha.
Learning registered point processes from idiosyncratic observations. *arXiv:1710.01410*, 2017.

Hongteng Xu, Dixin Luo, Xu Chen, and Lawrence Carin.
Benefits from superposed Hawkes processes *arXiv:1710.01410*, 2017.

The End