Nonparametric Regressive Point Processes Based on Conditional Gaussian Processes

Siqi Liu and Milos Hauskrecht

Department of Computer Science, University of Pittsburgh

University of Pittsburgh

Introduction

We study **point process** models for **multivariate event sequences**

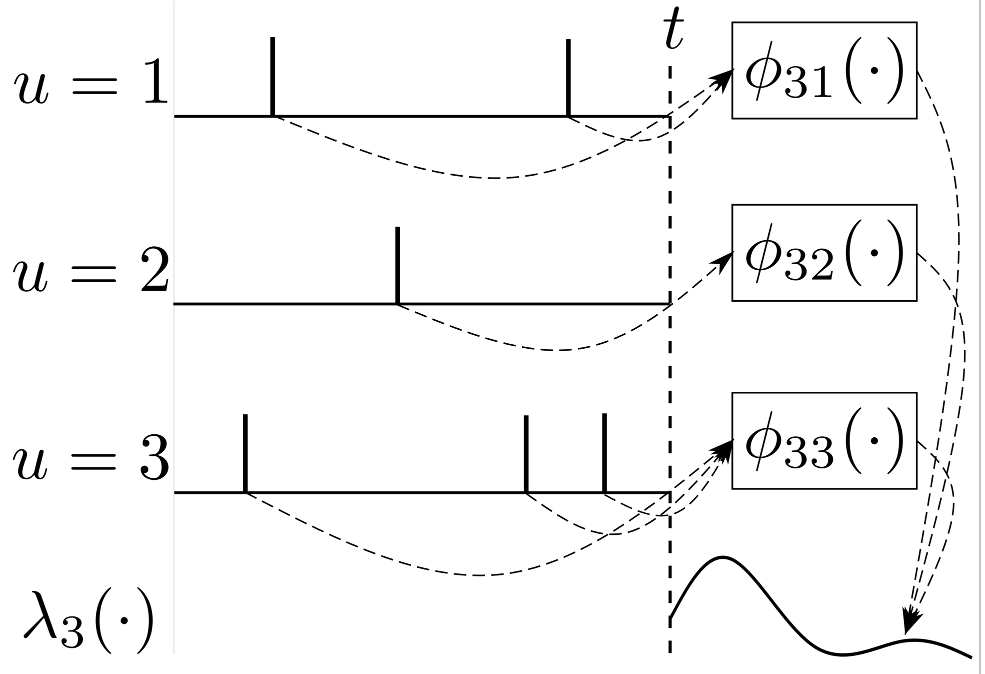
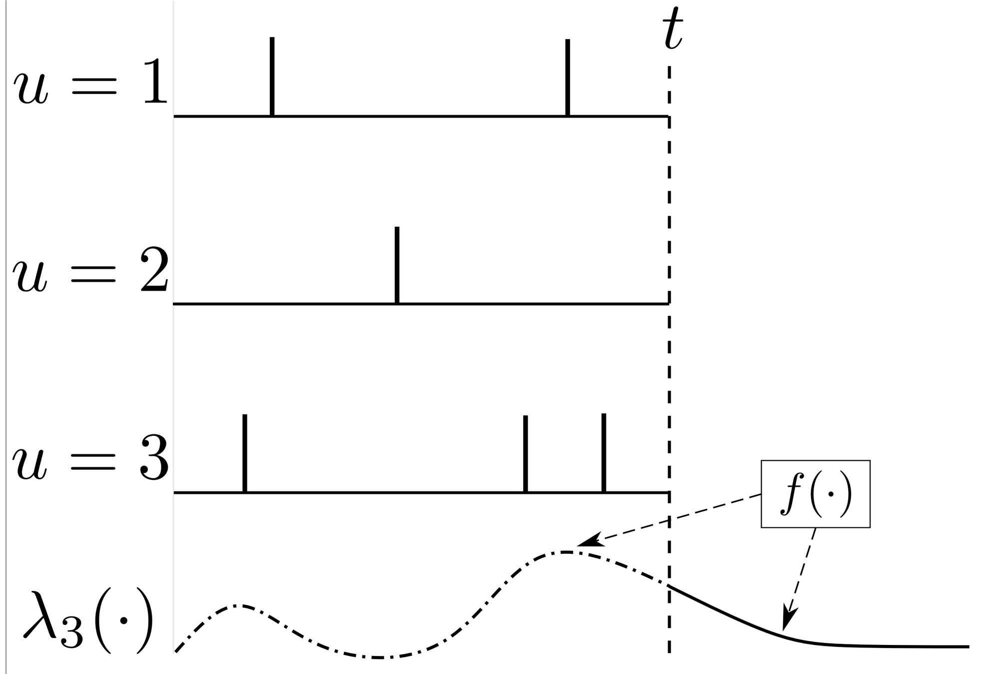
- Data consist of multiple sequences $\mathcal{D} = \{y_c\}_{c=1}^{|\mathcal{D}|}$, $y_c = \{(t_i, u_i)\}_{i=1}^{|y_c|}$
- $t_i \in \mathbb{R}$ is the time and $u_i \in \{1, \ldots, U\}$ the type of the i-th event
- For each type u of events, a conditional intensity function (CIF)

$$\lambda_u(t) = \lim_{dt \to 0^+} \frac{\mathbb{E}[N([t, t+dt))|\mathcal{H}_t]}{dt}$$

is the instantaneous rate of events at time t given the history $\mathcal{H}_t$

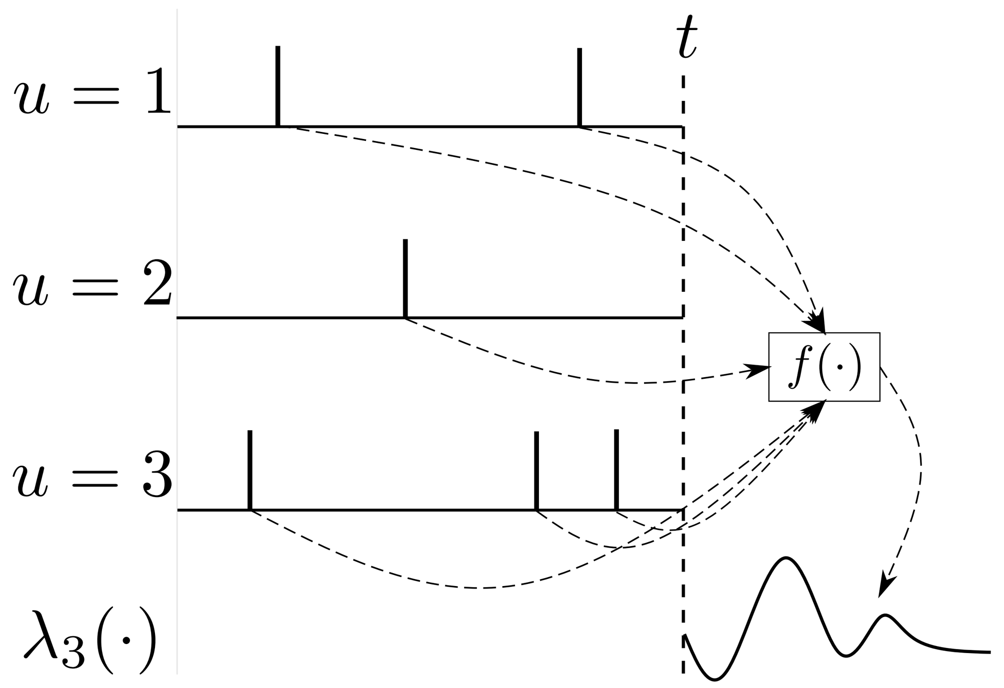
$$p(y_c) = \prod_{u=1}^{U} \left[\prod_{i=1}^{|y_c|} \lambda_u(t_i)^{\delta(u_i, u)} \exp\left(-\int_0^{T^c} \lambda_u(t) dt \right) \right] \triangleq \prod_{u=1}^{U} p_u(y_c)$$

Two types of models have been developed independently over years

Hawkes Process	**GP-Modulated Point Process**
	
$\lambda_3(t) = \mu_3 + \sum_{t_j<t} \phi_{3u_j}(t - t_j)$	$\lambda_3(t) = g(f(t)), f \sim \mathcal{GP}, g(\cdot) \geq 0$
Regressive point process	**Latent-state** point process
Once learned, applicable to **unseen** sequences	Principled way to **flexibly** model the intensity functions
Limited flexibility	Inference on the **same** sequence

GP Regressive Point Processes

We propose a model combining the advantages of the above two models



$$\lambda_3(t) = f(x(t))^2 = f(t - s_1^1(t), \ldots, t - s_U^Q(t))^2$$

- $f \sim \mathcal{GP}(\mu, K)$
- $s_u^q(t)$ is the time of the q-th (from last) event of type u before time t
- $x(t) = (t - s_u^q(t))_{u=1,q=1}^{U,Q}$ are the times since the last Q events for each type u

$$K(x(t), x'(t')) = \sum_{d=1}^{D} \underbrace{\mathbb{I}[x_d(t)]\,\mathbb{I}[x'_d(t')]}_{K_1} \underbrace{\gamma_d \exp\left(-\frac{(x_d(t) - x'_d(t'))^2}{2\alpha_d} \right)}_{K_2}$$

- $\mathbb{I}[x_d(t)] = 1$ iff the q-th (from last) event of type u exists

Conditional GPRPP

- Introduce a set of pseudo-input-points Z and their pseudo-observations $f_Z = m_Z + \epsilon_Z$ with noise ϵ_Z
- Marginalize out ϵ_Z and maximize the likelihood conditioned on m_Z

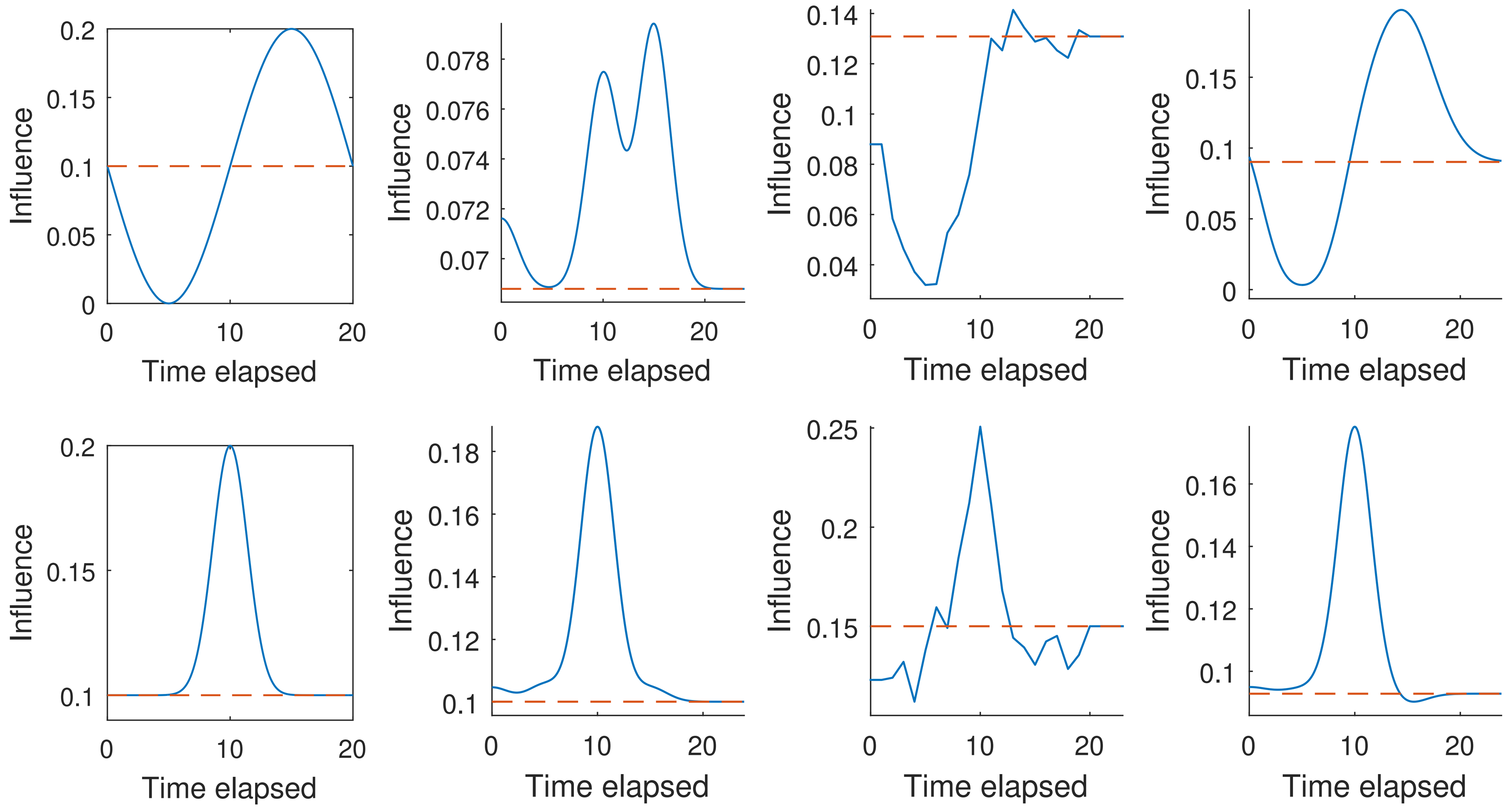
$$\begin{aligned} \ln p_{\tilde{u}}(y|m_Z) &= \ln \iint p_{\tilde{u}}(y|f_x) p(f_x|m_Z, \epsilon_Z) p(\epsilon_Z) df_x d\epsilon_Z \\ &= \ln \int p_{\tilde{u}}(y|f_x) p(f_x|m_Z) df_x \\ &= \ln \mathbb{E}[p_{\tilde{u}}(y|f_x)] \geq \mathbb{E}[\ln p_{\tilde{u}}(y|f_x)] \end{aligned}$$

$p_{\tilde{u}}$ is the density for event type $\tilde{u}$

Results

Learning the influence of past events. Solid lines are the CIFs after the occurrence of an event of the same type. Dashed lines are the baseline CIFs.

Ground truth **HP-GS**[1] **HP-LS**[2] **CGPRPP**



Test log-likelihood on MIMIC lab orders. Each dataset consists of a different set of multiple types of lab test orders on patients.

Dataset	355	60	3	95	368	354	151	550	113	140
HP-GS[1]	-3668	-4673	**-3721**	-4064	-3366	-4344	-3338	-1053	-4656	-3206
HP-GS-A[1]	-3947	-5051	-3733	-4390	-3711	-4792	-3574	-1064	-5049	-3475
HP-LS[2]	-6510	-7299	-5722	-5712	-5625	-7185	-5323	-1744	-7143	-4625
NSMMPP[3]	-3664	-4660	-3737	-3982	**-3309**	-4409	-3763	**-1039**	-4539	-3244
CGPRPP	**-3249**	**-4246**	-3759	**-3933**	-3378	**-4225**	**-3093**	-1175	**-4276**	**-2942**

[1] Xu et al. Learning Granger causality for Hawkes processes. In *International Conference on Machine Learning*, 2016.

[2] Eichler et al. Graphical modeling for multivariate Hawkes processes with nonparametric link functions. *Journal of Time Series Analysis*, 2017.

[3] Mei et al. The neural Hawkes process: A neurally self-modulating multivariate point process. In *Advances in Neural Information Processing Systems*, 2017.