Multivariate Hawkes Processes for Large-scale Inference

Rémi Lemonnier[1, 2] Kevin Scaman[1] Argyris Kalogeratos[1]

[1]CMLA – ENS Cachan, CNRS, Université Paris-Saclay, France

[2]numberly, Paris, France

Abstract

We present a framework for fitting multivariate Hawkes processes for large-scale problems, both in the number of events in the observed history n and the number of event types d (i.e. dimensions). The proposed *Scalable Low-Rank Hawkes Process* (SLRHP) framework introduces a low-rank approximation of the kernel matrix that allows to perform the nonparametric learning of the d^2 triggering kernels in at most $O(ndr^2)$ operations, where r is the rank of the approximation ($r \ll d, n$). This comes as a major improvement to the existing state-of-the-art inference algorithms that require $O(nd^2)$ operations. Furthermore, the low-rank approximation allows SLRHP to learn representative patterns of interaction between event types, which is usually valuable for the analysis of complex processes in real-world networks.

Background

1. Motivations

Applications of Hawkes processes to large-scale problems

- *Finance*: modeling order book and buying order arrivals.
- *Biology*: modeling occurrences of genes in DNA chains.
- *Social interactions studies*: modeling videos shares/views, or tweets.

Previous work on large-scale inference

- Markovian nonparametric estimation using the memoryless property of exponential kernels: complexity $O(nd^2)$. Lemonnier and Vayatis [2014]
- Learning a low-rank mutual excitation matrix while fixing the temporal excitation pattern: complexity $O(n^2 d)$.Tran et al. [2015], Xu et al. [2016]

2. Problem statement

Hawkes processes on graphs

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a directed network of d nodes and $A \in \{0,1\}^{d\times d}$ its adjacency matrix. We consider a multivariate Hawkes process (MHP) $N(t) = \{N_u(t) : u = 1, ..., d, t \geq 0\}$ such that the *mutual excitations* take place along the edges of $\mathcal{G}$. For an event history $\mathcal{H}: (u_m, t_m)_{m=1}^n$, the rate of occurrence of node u is given by:

$$\lambda_u(t) = \mu_u(t) + \sum_{m: t_m < t} A_{u_m u}\, g_{u_m u}(t - t_m). \tag{1}$$

Model considerations

Without further assumptions, the inference requires:

- learning of d^2 triggering kernels that encode cross-excitements,
- $O(n^2)$ computations of $g_{u_m u}(t - t_m)$,

which yields a prohibitive $O(n^2 d^2)$ complexity.

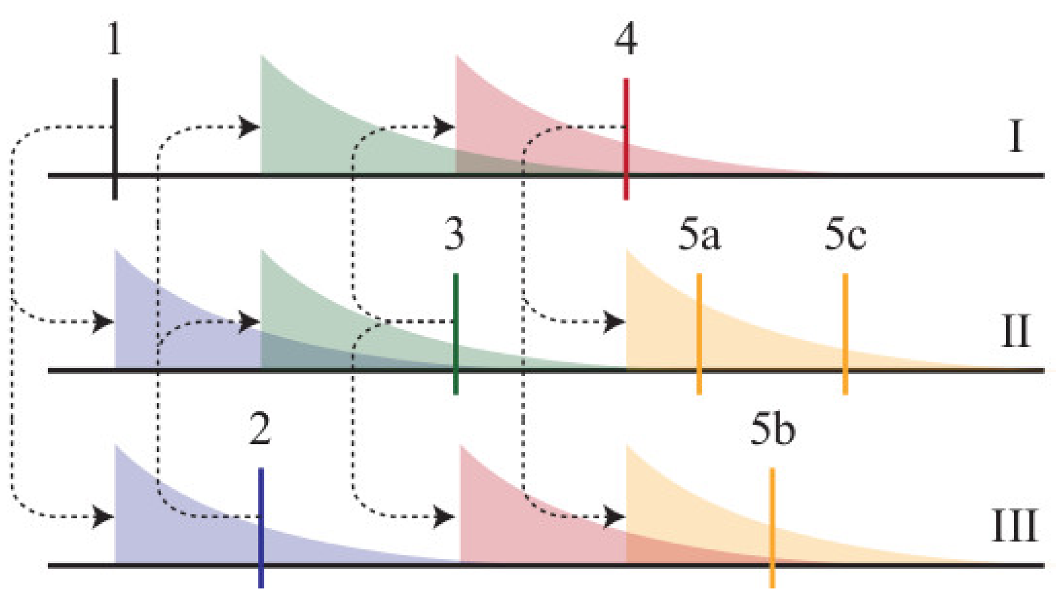


Figure: Mutually-exciting dynamics of Hawkes processes.

Theoretical Results

3. The proposed decompositions

Low-rank decomposition $O(d^2) \to O(d)$

The natural occurrence rates μ_u and triggering kernels g_{vu} of Eq. 1 are defined via the low-rank approximations:

$$\mu_u(t) = \textstyle\sum_{i=1}^r P_{ui}\, \tilde{\mu}_i(t) \text{ and } g_{vu}(t) = \sum_{i,j=1}^r P_{ui}\, P_{vj}\, \tilde{g}_{ji}(t), \tag{2}$$

where u, v are event types and $P \in \mathbb{R}_+^{d\times r}$ is the projection matrix from the original d-dimensional to the low-dimensional space.

Markovian decomposition $O(n^2) \to O(n)$

The natural occurrence rate and the kernel function are approximated by a sum of K exponential triggering functions with $\gamma, \delta > 0$ fixed hyperparameters:

$$\widehat{\mu}_i^K(t) = \textstyle\sum_{k=0}^K \beta_{i,k}\, e^{-k\gamma t} \text{ and } \widehat{g}_{ji}^K(t) = \sum_{k=1}^K \alpha_{ji,k}\, e^{-k\delta t}. \tag{3}$$

4. Log-likelihood formulation

The log-likelihood of the model can be rewritten as follows:

$$\widehat{\mathcal{L}}(P, \mathcal{H}; \alpha) = \sum_{h,m} \ln\left(\sum_{u,v,i,j,k} P_{ui}\, P_{vj}\, \alpha_{ji,k}\, D_{h,m,u,v,k} \right) - \sum_{h,u,v,i,j,k} P_{ui}\, P_{vj}\, \alpha_{ji,k}\, B_{h,u,v,k}, \tag{4}$$

where B and D are two sparse tensors that can be computed in $O(nd)$.

5. Optimization algorithm

Optimization of Hawkes parameters α

Using *self-concordant barriers* Nesterov et al. [1994]

Optimization of projections matrices P

Let p be a reshaping of the projection matrix P to a vector (linearized) and $2\,\Xi_{ui,vj}^{hm} = \sum_k (\alpha_{ji,k} D_{h,m,u,v,k} + \alpha_{ij,k} D_{h,m,v,u,k})$, $2\,\Psi_{ui,vj} = \sum_{h,k} (\alpha_{ji,k} B_{h,u,v,k} + \alpha_{ij,k} B_{h,v,u,k})$.

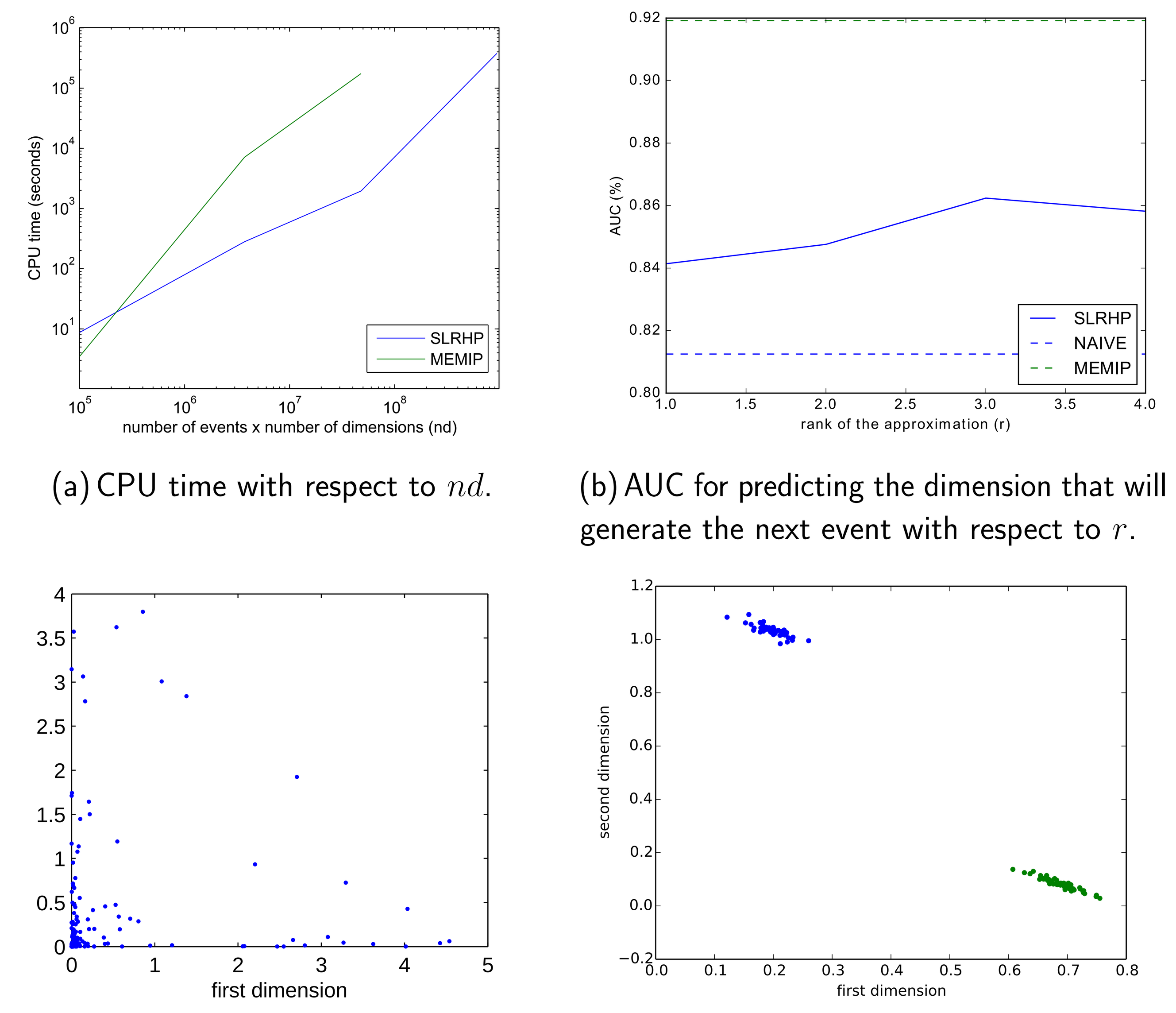
Proposition 1. *The log-likelihood is non-decreasing under the update:*

$$p_{ui}^{t+1} = p_{ui}^t \left(\sum_{h,m} \frac{(\Xi^{hm} p^t)_{ui}}{p^{t\top} \Xi^{hm} p^t (\Psi p^t)_{ui}} \right)^{1/2}. \tag{5}$$

Furthermore, if p_{ui} is a stable fixed point of Eq. 5, then p_{ui} is a local maximum of the log-likelihood.

Experiments

6. Results



(a) CPU time with respect to nd.

(b) AUC for predicting the dimension that will generate the next event with respect to r.

(c) Learned embedding when $r = 2$.

(d) Learned embedding for simulated example.

7. Empirical conclusions

- **Highly scalable** approach able to scale to datasets one order of magnitude larger than the state-of-the-art.
- **Little loss of accuracy** compared to state-of-the-art competitors.

References

Remi Lemonnier and Nicolas Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate Hawkes processes. In *Machine Learning and Knowledge Discovery in Databases*, pages 161–176. Springer, 2014.

Yurii Nesterov, Arkadii Semenovich Nemirovskii, and Yinyu Ye. *Interior-point polynomial algorithms in convex programming*, volume 13. SIAM, 1994.

Long Tran, Mehrdad Farajtabar, Le Song, and Hongyuan Zha. Netcodec: Community detection from individual activities. In *Proceedings of the SIAM International Conference on Data Mining*, SIAM ICDM, pages 91–99, 2015.

Hongteng Xu, Mehrdad Farajtabar, and Hongyuan Zha. Learning granger causality for Hawkes processes. In *Proceedings of the International Conference on Machine Learning*, ICML, pages 1717–1726, 2016.