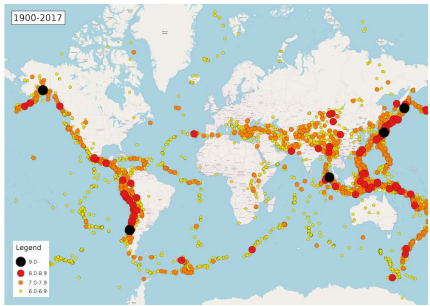
Hierarchical Contrastive Learning for Temporal Point Processes

AAAI2023(Oral)

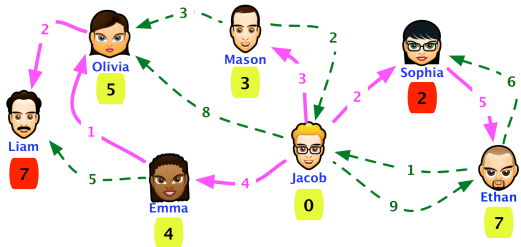
Qingmei Wang[1], Minjie Cheng[1], Shen Yuan[1], Hongteng Xu[1,2]

[1]Gaoling School of Artificial Intelligence, Renmin University of China
[2]Beijing Key Laboratory of Big Data Management and Analysis Methods

Event Sequences in Real World



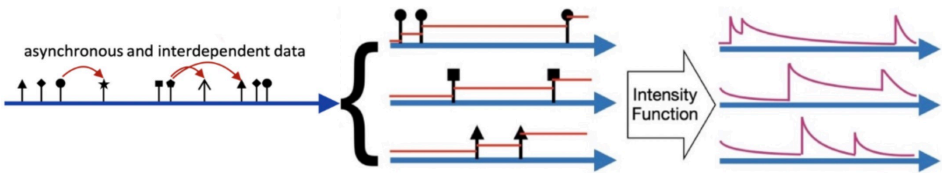
(a) The locations and the intensities of the earthquakes from 1900 to 2017.



(b) User behaviors on social networks.

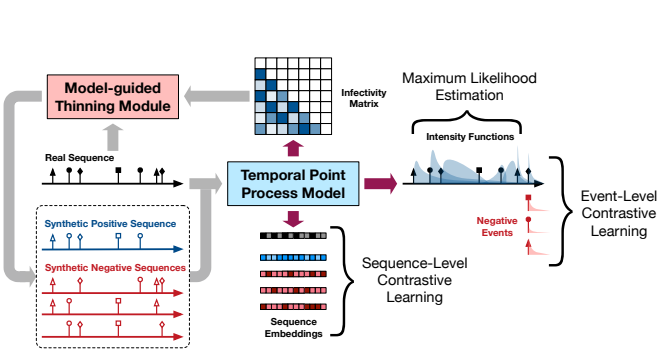
Event Sequences Modeling and Learning Temporal Point Processes

- Asynchronous and interdependent event sequences: $\boldsymbol{s} = \{(t_i, c_i)\}_{i=1}^{L}$

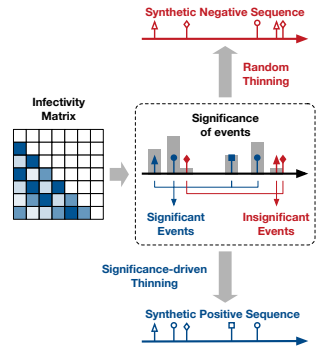


- Inspired by the EM algorithm, we can learn the TPP model $\{\lambda_c(t)\}_{c=1}^{C}$ by maximum likelihood estimation (MLE)
- **High risk of overfitting for imperfect (e.g., sparse, incomplete) event sequence**.

Hierarchical Contrastive Learning of TPPs



(c) Hierarchical contrastive learning



(d) Model-guided thinning

- Hierarchical contrastive learning method to regularize the MLE of TPPs.

$$\max_{f_\theta} \underbrace{\mathcal{L}(\boldsymbol{s};\theta)}_{\text{Log-likelihood}} + \underbrace{\gamma_1 \sum_{i=1}^{L} \mathcal{L}_{\text{event}}(\boldsymbol{\lambda}(t_i))}_{\text{Event-level contrastive loss}} + \underbrace{\gamma_2 \mathcal{L}_{\text{seq}}(\boldsymbol{s}, \boldsymbol{s}_{(P)}, \{\boldsymbol{s}_{k,(N)}\}_{k=1}^{K})}_{\text{Sequence-level contrastive loss}}, \quad (1)$$

Comparison Experiments

Table 2: Comparisons for various methods on learning TPPs from different datasets

Models	Data	Metrics	Methods					
			MLE+Reg	MLE+DA	DIS	INITIATOR	NCE-TPP	MLE+HCL
HP	Hawkes	Log-Like	-0.06 (0.05)	-0.52 (0.34)	-6.6 (1.11)	-0.22 (0.02)	-0.10 (0.08)	**-0.04 (0.05)**
		Type-Acc	0.38 (0.01)	0.38 (0.01)	0.32 (0.05)	0.35 (0.02)	0.33 (0.02)	**0.40 (0.00)**
	Missing	Log-Like	-0.06 (0.00)	-1.09 (0.00)	-3.38 (0.00)	-0.53 (0.04)	-0.04 (0.01)	**-0.02 (0.00)**
		Type-Acc	**0.42 (0.00)**	0.41 (0.01)	0.40 (0.01)	0.38 (0.02)	0.41 (0.01)	**0.42 (0.02)**
	Bookorder	Log-Like	-2.60 (0.40)	-3.28 (0.58)	-1.64 (0.36)	-1.71 (0.25)	-1.60 (0.38)	**-1.57 (0.30)**
		Type-Acc	0.57 (0.00)	0.57 (0.01)	**0.62 (0.01)**	0.60 (0.02)	**0.62 (0.00)**	**0.62 (0.00)**
	StackOverflow	Log-Like	-0.77 (0.01)	-2.31 (0.12)	-0.96 (0.07)	-0.79 (0.03)	-0.74 (0.04)	**-0.72 (0.01)**
		Type-Acc	0.45 (0.02)	0.43 (0.02)	0.40 (0.05)	0.49 (0.02)	**0.51 (0.06)**	0.50 (0.01)
	Retweet	Log-Like	-8.94 (0.20)	-10.73 (2.20)	—	-8.89 (0.11)	-8.92 (0.05)	**-8.84 (0.02)**
		Type-Acc	0.60 (0.01)	0.59 (0.00)	0.58 (0.02)	0.62 (0.05)	**0.66 (0.03)**	**0.66 (0.04)**
THP	Hawkes	Log-Like	0.11 (0.03)	-1.23 (0.43)	-0.68 (0.13)	0.03 (0.05)	0.12 (0.01)	**0.14 (0.02)**
		Type-Acc	0.38 (0.00)	0.26 (0.00)	0.34 (0.03)	0.24 (0.02)	**0.40 (0.01)**	0.38 (0.00)
	Missing	Log-Like	-0.47 (0.01)	-1.32 (0.21)	-0.75 (0.16)	-1.08 (0.10)	-0.50 (0.02)	**-0.34 (0.08)**
		Type-Acc	0.41 (0.01)	0.27 (0.00)	0.41 (0.00)	0.40 (0.00)	**0.42 (0.01)**	0.41 (0.00)
	Bookorder	Log-Like	-1.69 (0.31)	-4.50 (1.79)	-1.64 (0.36)	-1.70 (0.43)	-1.60 (0.30)	**-1.58 (0.21)**
		Type-Acc	0.62 (0.00)	0.62 (0.01)	0.62 (0.00)	0.53 (0.01)	**0.64 (0.00)**	**0.64 (0.00)**
	StackOverflow	Log-Like	**-0.77 (0.00)**	-2.31 (0.12)	-0.96 (0.07)	-0.89 (0.10)	**-0.77 (0.02)**	-0.79 (0.01)
		Type-Acc	0.43 (0.00)	0.42 (0.02)	0.49 (0.03)	0.40 (0.05)	0.39 (0.02)	**0.44 (0.01)**
	Retweet	Log-Like	-7.35 (0.35)	-9.14 (0.46)	—	-10.20 (0.53)	-7.33 (0.29)	**-7.27 (0.18)**
		Type-Acc	0.53 (0.00)	0.50 (0.01)	0.54 (0.00)	0.53 (0.01)	0.54 (0.00)	**0.56 (0.00)**

* "—" means the learning method fails to converge. The best results are bolden. In each cell, the averaged performance is shown, and the parentheses contains the standard deviation.

Summary

- The proposed HCL provides a new regularizer for the scheme of MLE.
- The complexity of the proposed model-guided thinning method is $O(N)$ and can be $O(1)$ in parallel, while the complexity of Ogata's thinning method is $O(N^2)$.
- Try to combine the proposed HCL method with other learning framework in the future and find theoretical supports for HCL method.