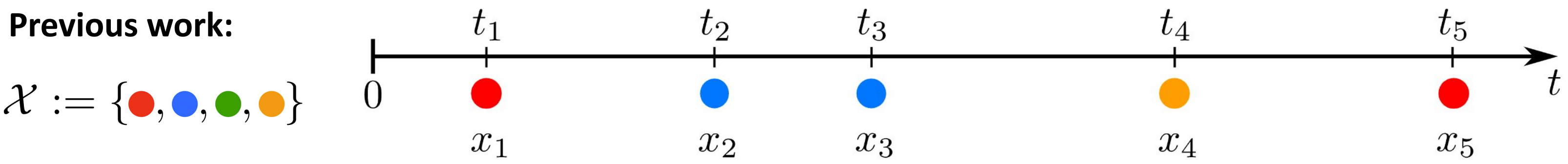
University of California, Irvine

Probabilistic Modeling for Sequences of Sets in Continuous-Time

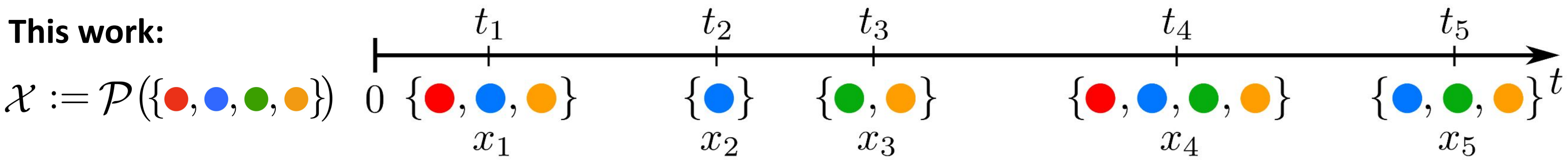
Yuxin Chang[1] Alex Boyd[2] Padhraic Smyth[1,2]
[1]Department of Computer Science [2]Department of Statistics
University of California, Irvine

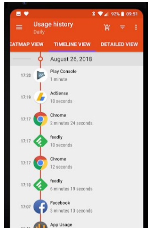
Event Data Over Continuous-Time

Previous work:



This work:





Images from Google search.

Marked Temporal Point Processes (MTPPs)

$$x \in \mathcal{X} := \{1, \ldots, K\}$$

$$\lambda(t) := \sum_x \lambda_x(t) := \sum_x \lambda_x(t \mid \mathcal{H}(t))$$

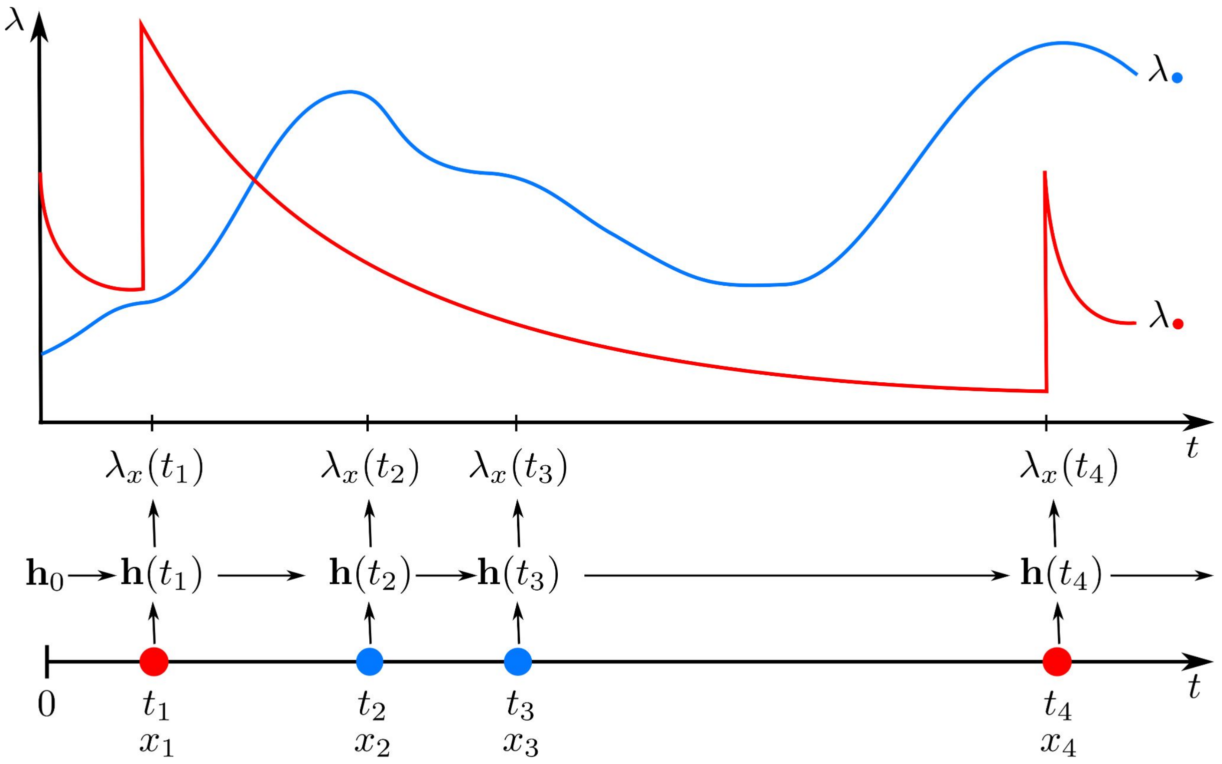
Total Intensity

Mark-specific Intensities

Can be modeled by…

- Parametric: e.g., Hawkes
- Neural: e.g., neural/transformer Hawkes

Recurrent MTPPs $\lambda_x(t) = f_x(\mathbf{h}(t))$



E.g., Recurrent MTPP (Du et al., 2016) and Neural Hawkes process (Mei and Eisner, 2017).

Sequence Log-likelihood

Challenge for sets-valued events: $|\mathcal{X}| = 2^K$

$$\log p(\mathcal{H}(T)) = \sum_{i=1}^{N} \log \lambda_{x_i}(t_i) - \int_0^T \lambda(s)ds$$

$$= \underbrace{\sum_{i=1}^{N} \log p(x_i \mid t_i)}_{\mathcal{L}_{\text{Set}}} + \underbrace{\sum_{i=1}^{N} \log \lambda(t_i) - \int_0^T \lambda(s)ds}_{\mathcal{L}_{\text{Time}}}$$

Model set structures directly

Modeling Set Structures

Dynamic Bernoulli

$$\mathcal{O}(Kd)$$

$$p(X = x \mid t, \mathbf{h}(t)) \sim \text{Bernoulli}(\boldsymbol{\theta}(t))$$

$$\theta_k(t) = \text{function}_k(\mathbf{h}(t)) \text{ for } k \in \{1, \ldots, K\}$$

Dynamic DPPs[1]

$$\mathcal{O}(K^3 + Kd)$$

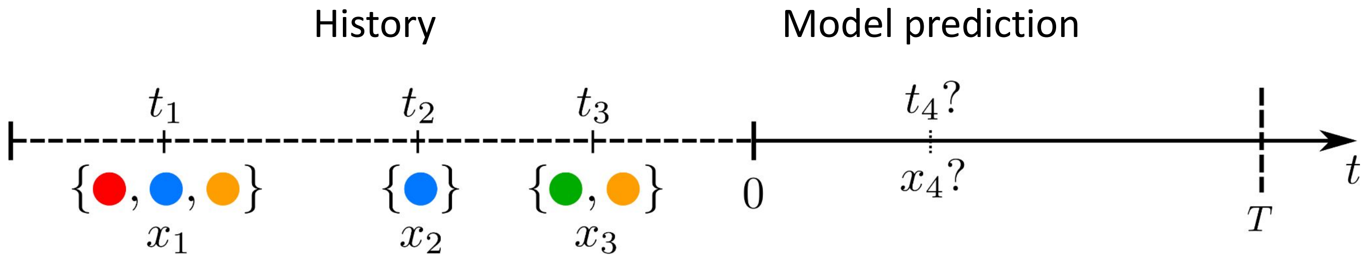
$$p(X = x \mid t, \mathbf{h}(t)) := \frac{\det(\mathbf{L}_x(t))}{\det(\mathbf{L}(t) + \mathbf{I})}$$

$$\mathbf{L}_x(t) = \text{function}(x, \mathbf{h}(t))$$

$$\mathbf{L} \in \mathbb{R}^{|K| \times |K|}, \quad \mathbf{L}_x \in \mathbb{R}^{|x| \times |x|}$$

[1]DPPs: determinantal point processes (Macchi, 1975; Kulesza el al., 2012)

Probabilistic Querying



Querying at ***item-level*** in the context of set-valued events:

$p(\bullet \text{ before } T \mid \text{history})$?

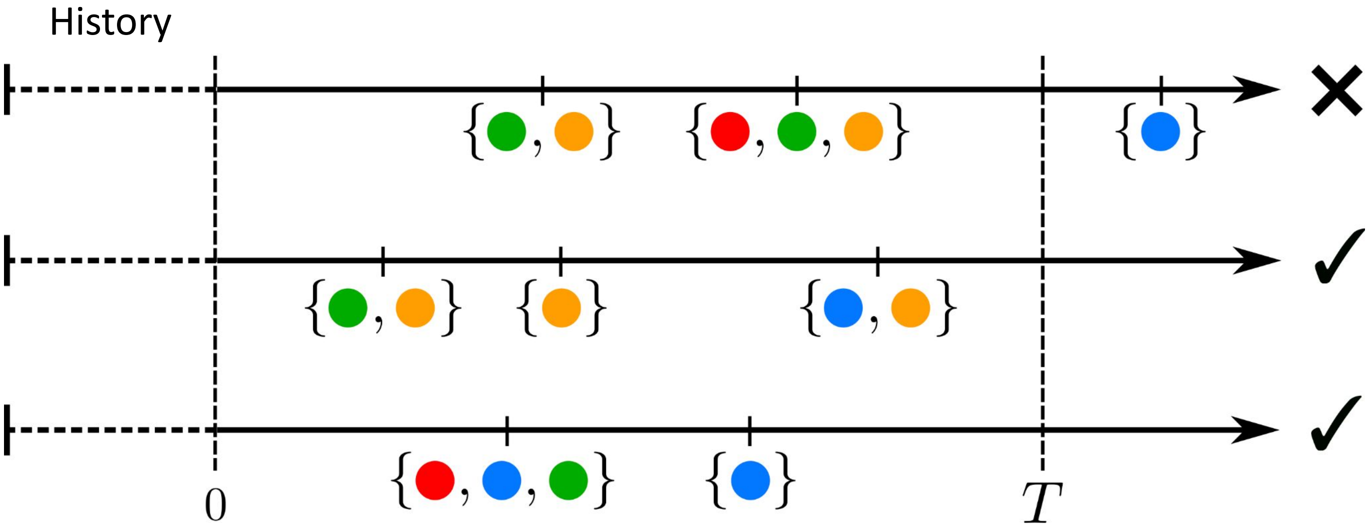
$p(\bullet \text{ before } \bullet \mid \text{history})$?

Beyond one-step-ahead prediction

$p(\text{statement in the future} \mid \text{history})$? ⟶ **Intractable!**

Answering Queries: Naive Sampling (Daley and Vere-Jones, 2003)

$p(\bullet \text{ before } T \mid \text{history})?$



Importance Sampling Estimators: Hitting Time Example

Boyd, Showalter, Mandt, Smyth. (NeurIPS 2022)
Boyd, Chang, Mandt, Smyth. (AIStats 2023)

Define proposal distribution $q \sim \mu_x(t) = \mathbb{1}(\text{no overlap in } A \text{ and } x)\lambda_x(t)$:

$$
\begin{aligned}
p(\text{hit}(A) \leq t) &= 1 - p(\text{hit}(A) > t) \\
&= 1 - \mathbb{E}_{\mathcal{H}\sim p}\left[\mathbb{1}\left(\text{no item in } A \text{ occurs in } \mathcal{H}[0,t]\right)\right] \\
&= 1 - \mathbb{E}_{\mathcal{H}\sim q}\left[\mathbb{1}\left(\text{no item in } A \text{ occurs in } \mathcal{H}[0,t]\right)\frac{p(\mathcal{H}[0,t])}{q(\mathcal{H}[0,t])}\right] \\
&= 1 - \mathbb{E}_{\mathcal{H}\sim q}\left[\frac{p(\mathcal{H}[0,t])}{q(\mathcal{H}[0,t])}\right] \\
&= \ldots \\
&= 1 - \mathbb{E}_{\mathcal{H}\sim q}\left[\exp\left(-\int_0^t \lambda(s)\hat{p}_A(s)ds\right)\right]
\end{aligned}
$$

Probability estimate by the model

Can be extended to general Boolean expressions, see paper for more details.

Experiments

Dataset	# Sequences	# Items	T_{max}	Avg. Length
Shopping	174,615	21	366	17
Music Listening	10,705	15	744	207
Movie Review	11,198	20	8,781	65
Online Courses	6,892	97	715	52

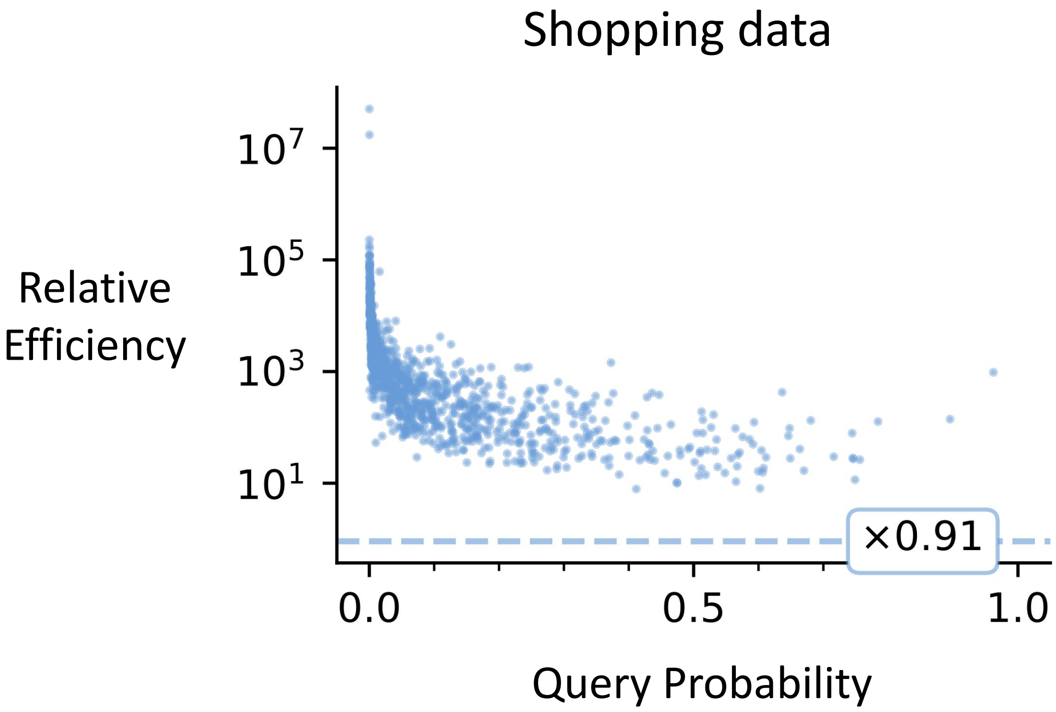
- Three main sets of experiments:
  - Querying efficiency
  - Test log-likelihood comparison (in paper)
  - Query-based model selection (in paper)

Evaluating Importance Estimators (Van der Vaart, 2000)

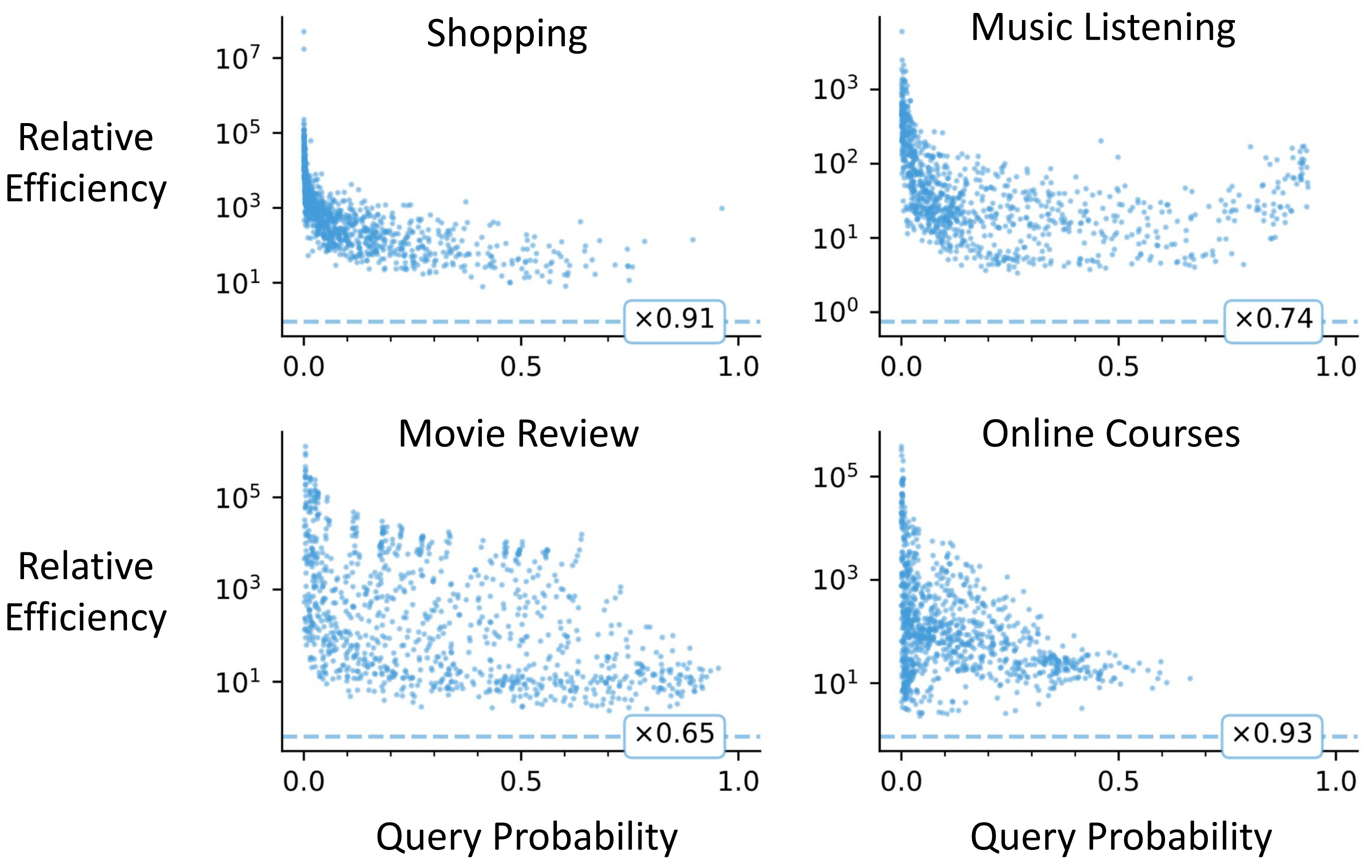
$$\text{Relative Efficiency} := \frac{\mathrm{Var}_p\left[\hat{\pi}_{\mathrm{Naive}}(\mathcal{H}(T))\right]}{\mathrm{Var}_q\left[\hat{\pi}_{\mathrm{Imp.}}(\mathcal{H}(T))\right]}$$

Intuition: # times fewer importance samples to achieve the same estimator variance.

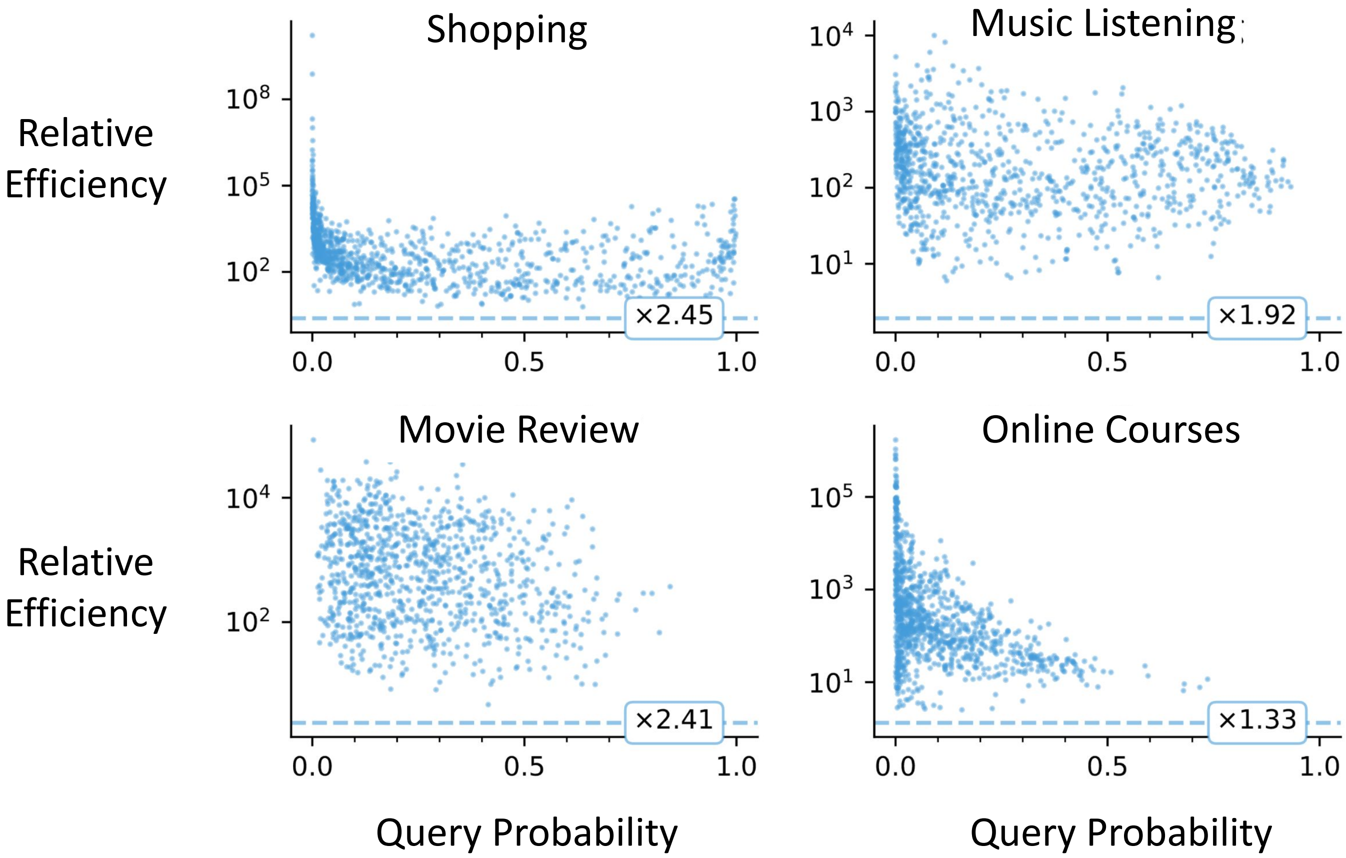
Relative Efficiency of Hitting Time Queries



Statistical Efficiency of Hitting Time Queries



Statistical Efficiency of “*A*-before-*B*” Queries



Contributions

1. New probabilistic models for set-valued data in continuous-time.
2. New querying methods to answer ***item-level*** queries for set-valued events.
3. Demonstrated significant improvements on modeling and *orders-of-magnitude* improvements on querying efficiency over alternative methods.

Poster @Today, Room 1 #53

Email: yuxinc20@uci.edu

Our paper: