Learning Time-Varying Coverage Functions

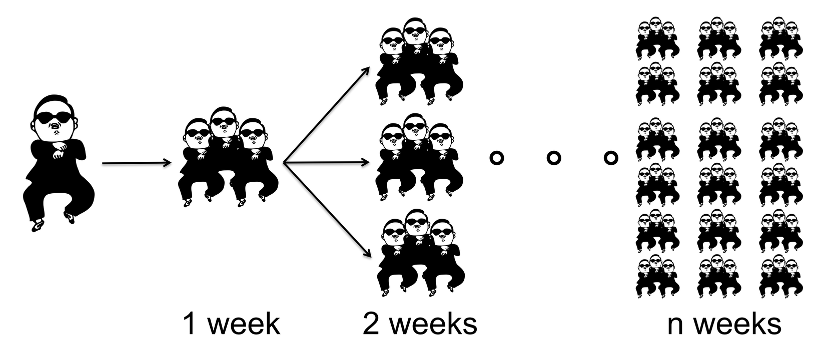
Nan Du† Yingyu Liang‡ Maria-Florina Balcan◇ Le Song†

†Georgia Institute of Technology, ‡Princeton University, ◇ Carnegie Mellon University

Georgia Institute of Technology

Motivation

- Viral Marketing: Can we predict how many people will follow several weeks later ?



- Warehouse Distribution : Can we estimate how many users will be served in the future ?

Challenges

- Users' influence (or warehouses' utilities) are unknown.
- These influence (or utilities) grow continuously in time.
- Temporal dynamics governing such growth are latent and unknown.
- Only temporal event information, e.g., when a user shares a video, when an order arrives at a warehouse, etc. might be available.

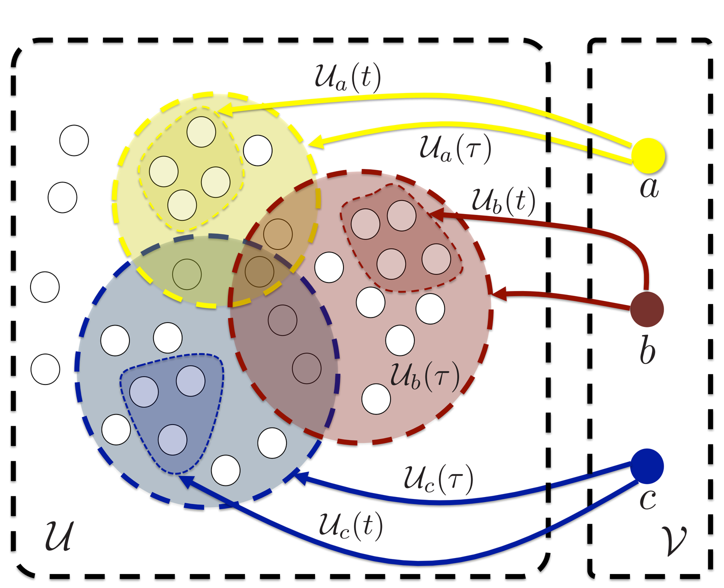
Can we learn the users' influence (or warehouses' utilities) as a function of time from these limited observations ?

Time-Varying Coverage Function

- A time-varying coverage function is a temporal combinatorial function over a finite set $\mathcal{V}$ of items, defined as

$$f(\mathcal{S},t) = Z\cdot\mathbb{P}\left(\bigcup_{s\in\mathcal{S}}\mathcal{U}_s(t)\right), \quad \text{for all } \mathcal{S}\in 2^{\mathcal{V}}, \text{ where}$$

- The ground set $\mathcal{U}$ with
  - σ-algebra $\mathscr{A}$
  - Probability measure $\mathbb{P}$
  - Normalization constant Z
- $\mathcal{U}_s(t)\subseteq\mathcal{U}$: the set covered by item $s\in\mathcal{V}$ at time t
- $\mathcal{U}_s(t)\subseteq\mathcal{U}_s(\tau)$ for all $t\leqslant\tau$ and $s\in\mathcal{V}$

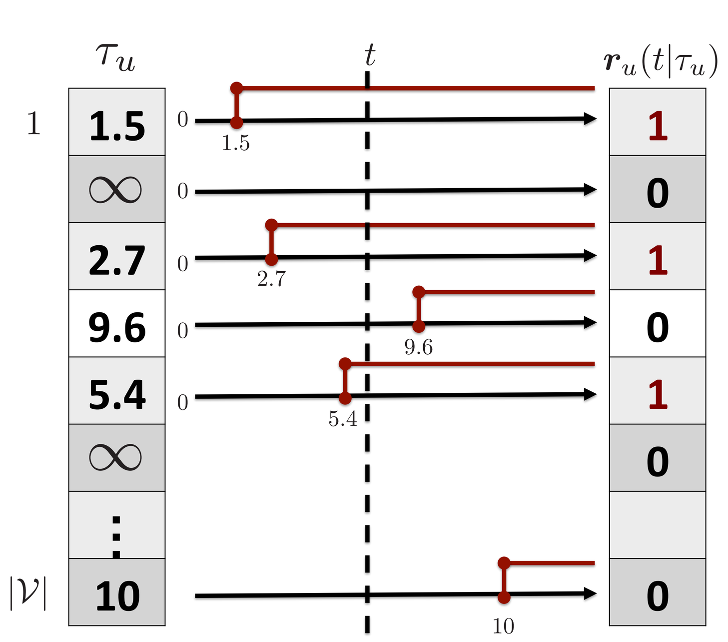


Representation

- $\forall u\in\mathcal{U}$, $|\mathcal{V}|$-dimensional vector τ_u^s = time being covered by s.
- $\boldsymbol{r}_u(t)$: $\mathbb{R}_+\mapsto\{0,1\}^{|\mathcal{V}|}$ indicates whether u is covered at time t by each $s\in\mathcal{V}$.
- **Lemma** we can represent

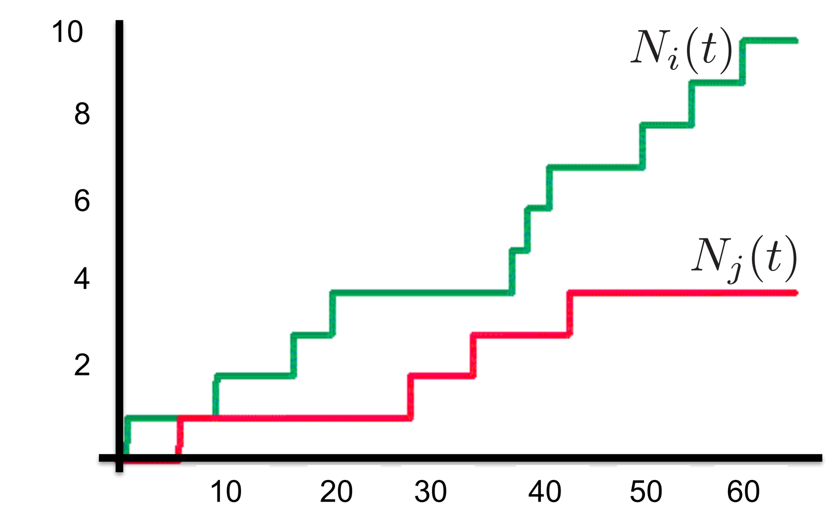
$$f(\mathcal{S},t) = Z\cdot\mathbb{E}_{\tau\sim\mathbb{Q}(\tau)}\left[\phi(\chi_{\mathcal{S}}^\top\boldsymbol{r}(t|\tau))\right]$$

where $\phi(x) := \min\{x,1\}$, and $\boldsymbol{r}(t|\tau)$ is a multidimensional step function based on τ.



Model Formulation by Counting Process

- $N(t) = \Lambda(t) + M(t)$ where
  - $N(t)$: counting process
  - $\Lambda(t)$: cumulative intensity function
  - $M(t)$: zero-mean martingale
- Assume each source set $\mathcal{S}_i$ induces a counting process $N_i(t)$
  - $N_i(t) = f(\mathcal{S}_i,t) + M_i(t)$
  - $f(\mathcal{S}_i,t) = \int_0^t a(\mathcal{S}_i,\tau)d\tau$
  - $a(\mathcal{S}_i,t)$: intensity function



Parametrization

- Kernel Smoothing : convolve the non-smooth intensity $a(\mathcal{S},t)$ with $K(t) = k(t/\sigma)/\sigma$ where σ is the bandwidth to get a smoothed intensity

$$a^K(\mathcal{S},t) = K(t)\star(df(\mathcal{S},t)/dt) = Z\cdot\mathbb{E}_{\tau\sim\mathbb{Q}(\tau)}[K(t-t(\mathcal{S},\tau))].$$

where $t(\mathcal{S},\tau)$ is the time when function $\phi(\chi_{\mathcal{S}}^\top\boldsymbol{r}(t|\tau))$ jumps from 0 to 1.

- Random Approximation : $\mathbb{Q}'(\tau)/C\le\mathbb{Q}(\tau)\le C\mathbb{Q}'(\tau)$, $\{\tau_i\}\overset{i.i.d.}{\sim}\mathbb{Q}'(\tau)$

$$\mathcal{A} = \left\{a_{\boldsymbol{w}}^K(\mathcal{S},t) = \sum_{i=1}^W w_i K(t-t(\mathcal{S},\tau_i)) \;:\; \boldsymbol{w}\geqslant 0, \frac{Z}{C}\leqslant\|\boldsymbol{w}\|_1\leqslant ZC\right\}.$$

- **Lemma** If $W = \tilde{O}(Z^2/(\epsilon\sigma)^2)$, with probability $\ge 1-\delta$, $\exists\tilde{a}\in\mathcal{A}$ such that $\mathbb{E}_{\mathcal{S}}\mathbb{E}_t\left[(a(\mathcal{S},t)-\tilde{a}(\mathcal{S},t))^2\right] = \mathbb{E}_{\mathcal{S}\sim\mathbb{P}(\mathcal{S})}\int_0^T\left[(a(\mathcal{S},t)-\tilde{a}(\mathcal{S},t))^2\right]dt/T = O(\epsilon^2+\sigma^4)$

Learning Algorithm : TCoverageLearner

- Given m *i.i.d.* counting processes, $\mathcal{D}^m := \{(\mathcal{S}_1,N_1(t)),\dots,(\mathcal{S}_m,N_m(t))\}$ up to observation time T, the log-likelihood is

$$\ell(\mathcal{D}^m|a) = \sum_{i=1}^m\left\{\int_0^T\{\log a(\mathcal{S}_i,t)\}\,dN_i(t) - \int_0^T a(\mathcal{S}_i,t)\,dt\right\}.$$

- Plugging the parametrization $a_{\boldsymbol{w}}^K(\mathcal{S},t)$ to solve

$$\min_{\boldsymbol{w}}\sum_{i=1}^m\left\{\boldsymbol{w}^\top\boldsymbol{g}_i - \sum_{t_{ij}<T}\log\left(\boldsymbol{w}^\top\boldsymbol{k}(t_{ij})\right)\right\} \quad \text{subject to} \quad \boldsymbol{w}\geqslant 0,\ \|\boldsymbol{w}\|_1\leqslant 1,$$

where we define t_{ij} as the j-th event occurs in the i-th process.

$$\boldsymbol{g}_{ik} = \int_0^T K(t-t(\mathcal{S}_i,\tau_k))\,dt \quad\text{and}\quad \boldsymbol{k}_l(t_{ij}) = K(t_{ij}-t(\mathcal{S}_i,\tau_l)).$$

- With Gaussian RBF kernel $\boldsymbol{g}_{ik} = \frac{1}{2}\left\{\operatorname{erfc}\left(-\frac{t(\mathcal{S}_i,\tau_k)}{\sqrt{2}\sigma}\right) - \operatorname{erfc}\left(\frac{T-t(\mathcal{S}_i,\tau_k)}{\sqrt{2}\sigma}\right)\right\}$.
- Sample $\{\tau_i\}\overset{i.i.d.}{\sim}\mathbb{Q}'(\tau)$ from training data
  - N_s = number of counting processes induced by $s\in\mathcal{V}$.
  - $\mathcal{J}_s$ = collection of all the jumping time before T
  - in probability $|\mathcal{J}_s|/|\mathcal{V}|N_s$, uniformly sample τ_i^s from $\mathcal{J}_s$; else, $\tau_i^s=\infty$.

Sample Complexity

Suppose $W = \tilde{O}\left(Z^2\left[\left(\frac{ZT}{\epsilon}\right)^{5/2} + \left(\frac{ZT}{\epsilon a_{\min}}\right)^{5/4}\right]\right)$ and $m = \tilde{O}\left(\frac{ZT}{\epsilon}[W+\epsilon_\ell]\right)$.

Then with probability $\ge 1-\delta$ over the random sample of $\{\tau_i\}_{i=1}^W$, we have that for any $0\le t\le T$,

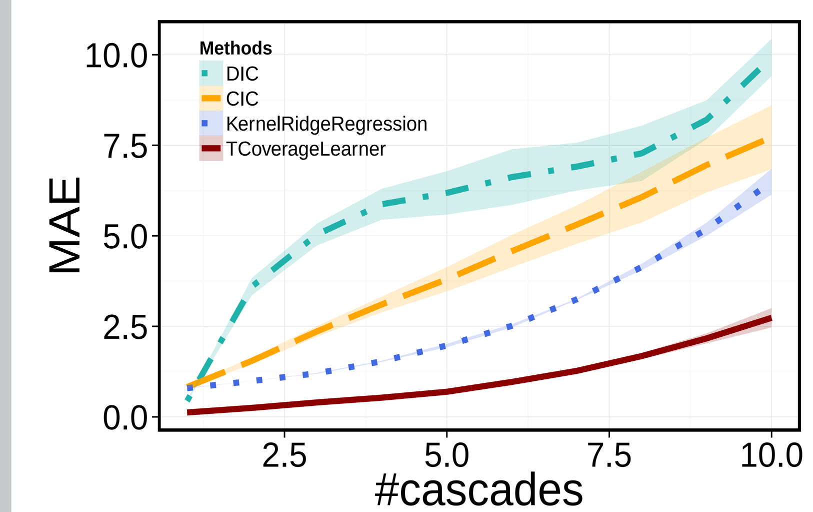
$$\mathbb{E}_{\mathcal{S}}\left[\hat{f}(\mathcal{S},t) - f(\mathcal{S},t)\right]^2\le\epsilon.$$

Experimental Evaluation : Competitors

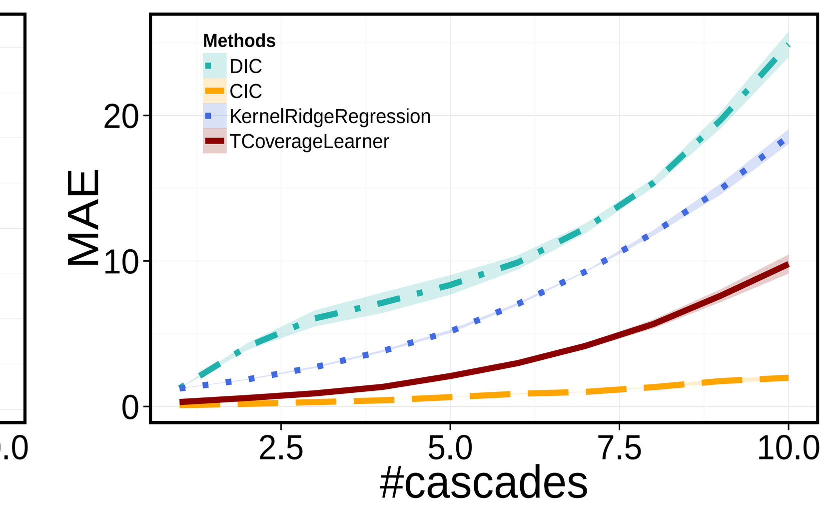
- **C**ontinuous-time **I**ndependent **C**ascade model with exponential pairwise transmission function (CIC).
- **D**iscrete-time **I**ndependent **C**ascade model (DIC).
- Kernel Ridge Regression

Experimental Evaluation : Synthetic Data

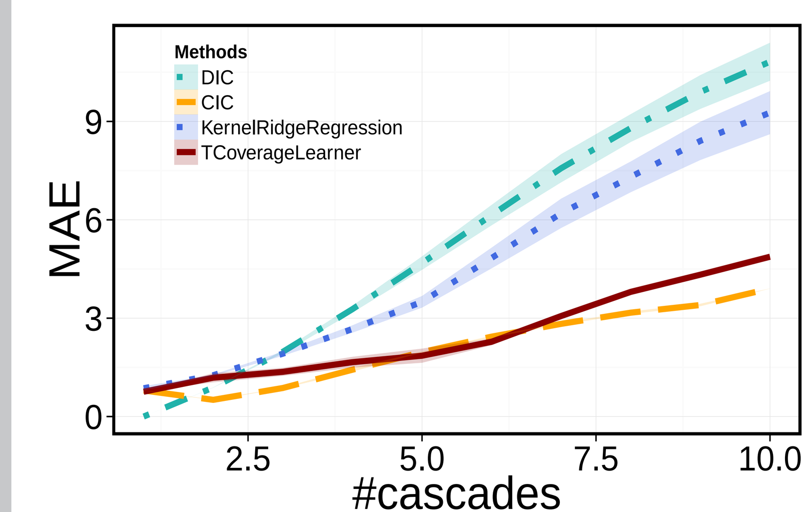
Robustness to model mis-specifications



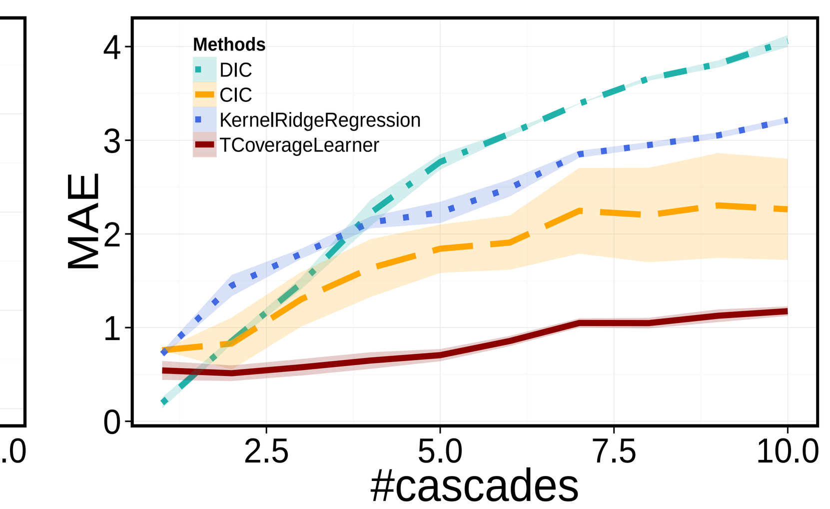
(a) Weibull Family (CIC)



(b) Exponential (CIC)

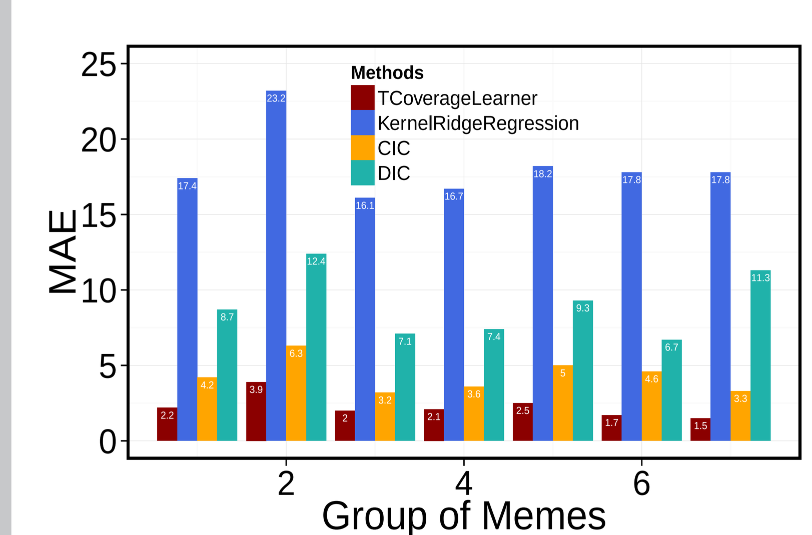


(c) DIC

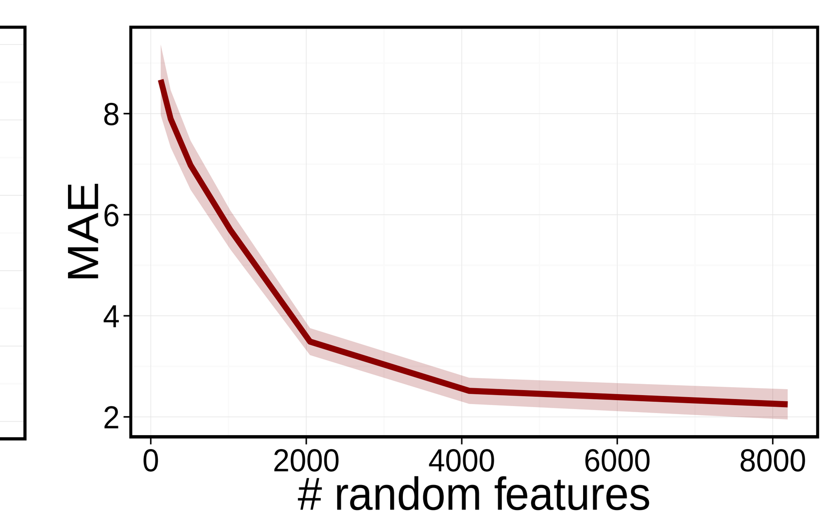


(d) Linear-Threshold

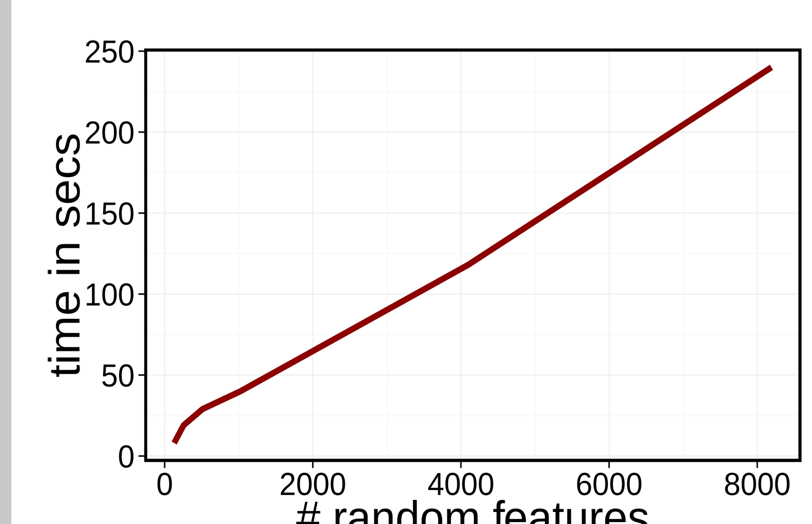
Experimental Evaluation : Real Data



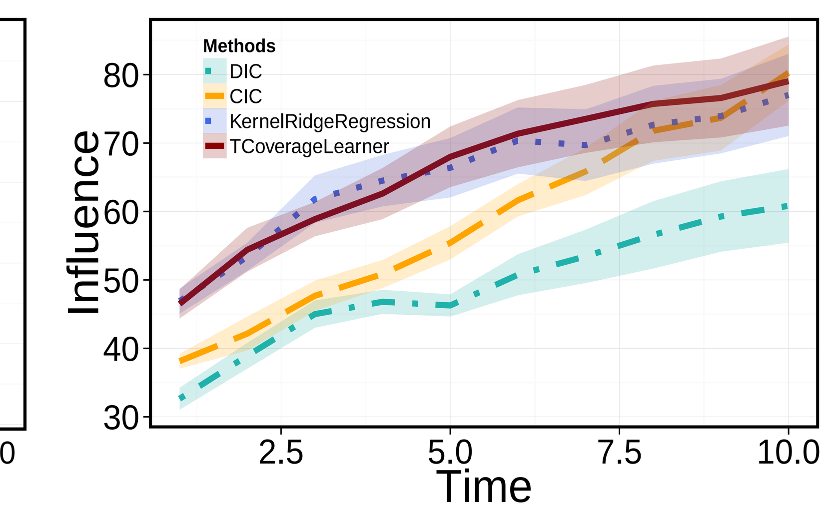
(a) MAE on real data



(b) Effect of random features



(c) Runtime



(d) Influence maximization

Mail: dunan@gatech.edu