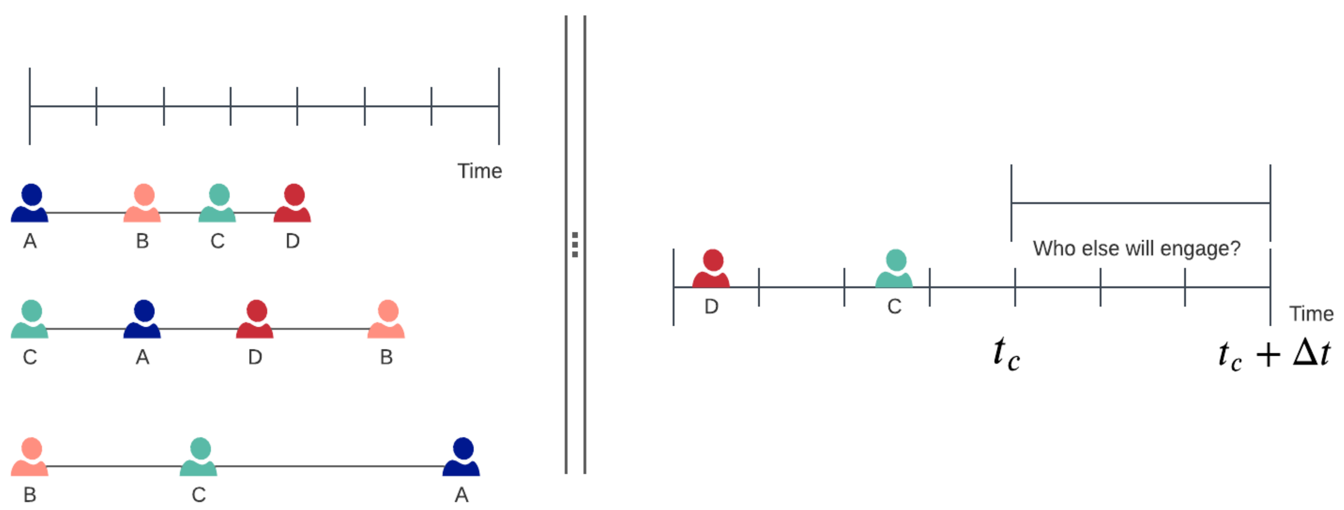
Anytime User Engagement Prediction in Information Cascades for Arbitrary Observation Periods

Akshay Aravamudan, Xi Zhang & Georgios C. Anagnostopoulos
Florida Institute of Technology

Problem Overview

- Study of information diffusion on internet is of increasing importance to several stakeholders.
- User engagement is one facet of popularity prediction.
- Temporal point processes are powerful statistical tools for event modeling on the timeline.

Goal: to predict the number of user who will engage a particular information cascade



Motivation

- User engagement prediction can help to determine product adoption, spread of political messaging, and which messages to send for content moderation to mitigate mis/dis-information.
- Existing works tackle this as either a macro-level task, which may require hand-crafted features, or a micro-level task which often lacks interpretability and/or requires training per prediction scenario (censoring time and observation period).

Contributions

- We provide a single model to predict user engagement for all censoring times t_c and forecast horizons Δt that performs competitively against state-of-the-art.
- We show our discriminative approach with split population formulation yields benefits over traditional approaches.
- We provide prediction intervals for the number of users engaging an information cascade.

Problem formulation

We build the model upon the following foundations:

- Multivariate survival processes: User engages a cascade as a survival process
- Split Population formulation: Imparts the passivity of users to cascades
- Right censoring of cascades: Data has been observed up to a time
- Discriminative objective: Train the model to predict user engagement

DANTE: Discriminative ANytime user Engagement prediction

Prediction probability

$$\mathrm{pp}(t_c, \Delta t) \triangleq \mathbb{P}\{T < t_c + \Delta t, \Delta = 1 | T \geq t_c, \mathcal{H}_{t_c}\}$$

Probability of observing an event from a user in the time interval $(t_c, t_c + \Delta t)$ given observed events in history $\mathcal{H}_{t_c}$

Discriminative likelihood

$$p(\ell \mid \mathcal{H}_{t_c}) = \mathrm{pp}(t_c, \Delta t)^{\ell} \left[1 - \mathrm{pp}(t_c, \Delta t)\right]^{1-\ell}$$

Posterior probability of user engaging ($\ell = 1$) in a cascade during $(t_c, t_c + \Delta t)$ given the cascade's history $\mathcal{H}_{t_c}$

"Anytime" formulation

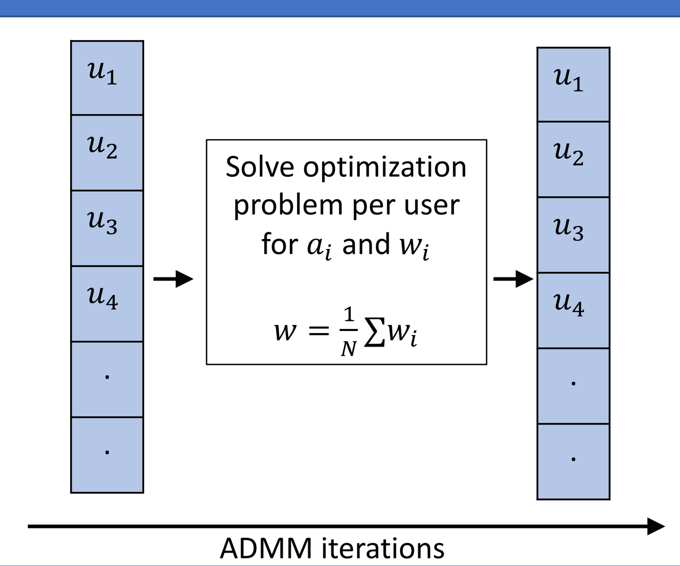
- Remove dependence on t_c and Δt

$$p_{\mathrm{LB}}(\ell \mid \mathcal{H}_{t_c}) \triangleq \left[\inf_{\substack{t_c \\ (t_c,\Delta t)\in\mathcal{S}_{\inf}}} \inf_{\Delta t} p(\ell \mid \mathcal{H}_{t_c})\right]^{\ell} \cdot \left[1 - \sup_{\substack{t_c \\ (t_c,\Delta t)\in\mathcal{S}_{\sup}}} \sup_{\Delta t} p(\ell \mid \mathcal{H}_{t_c})\right]^{1-\ell}$$

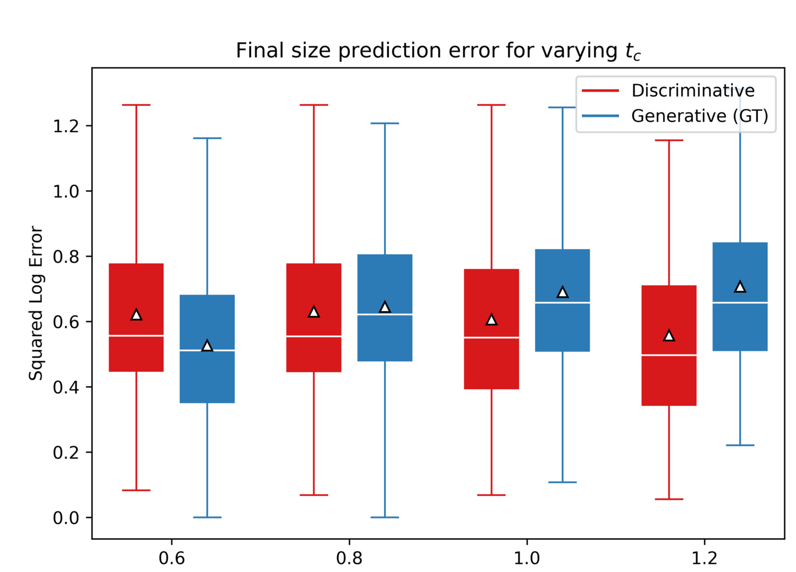
Maximizing the lower bound above enables learning for all censoring times t_c and forecast horizons $t_c + \Delta t$

Training

- Model parameters to learn, $\forall i \in \{1,2,\ldots,N\}$
  - a_i is used to construct the hazard rate.
  - w is the split population weight vector.
- Can be viewed as a multi-task learning problem per user with sub-problem objective $\tilde{E}^i(a_i, w)$, trained via Consensus ADMM.



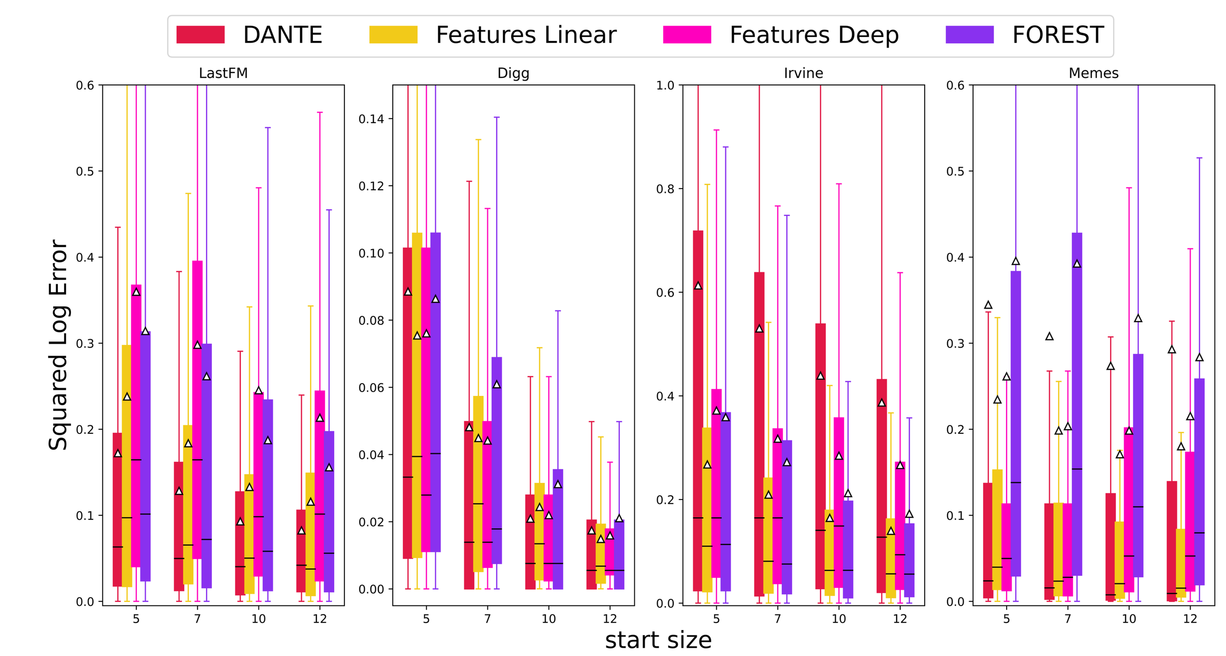
Synthetic Experiments



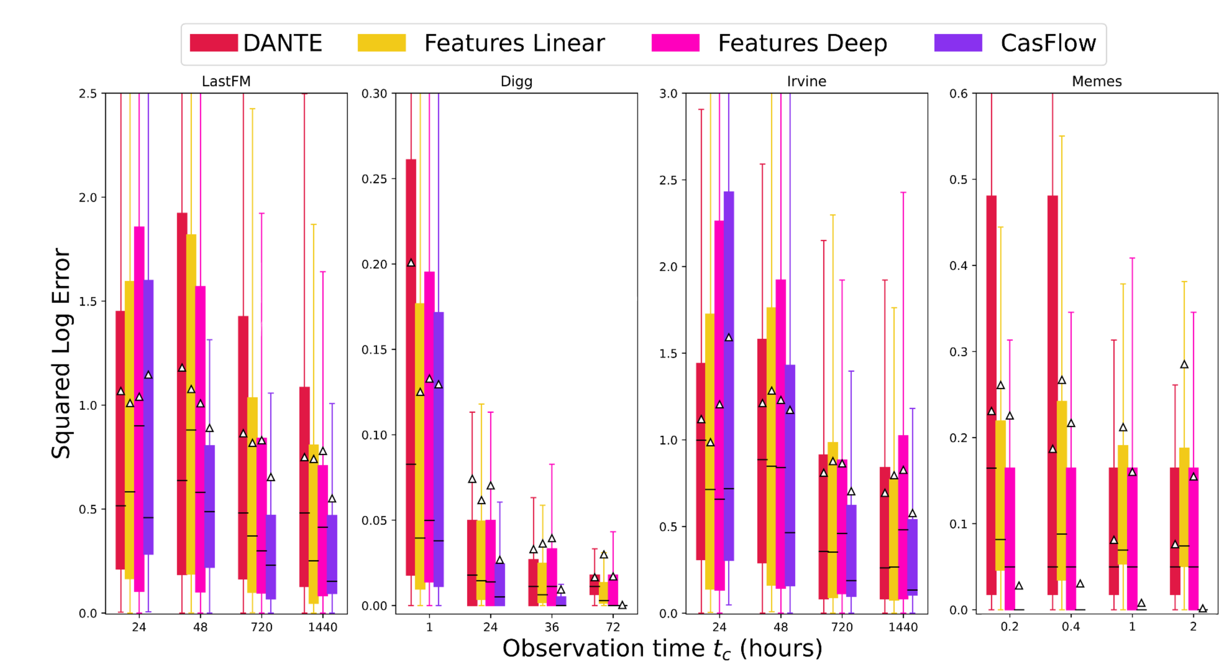
- We noticed that the discriminative model performed better on higher values of observation time.
- The discriminative model performed better in instances of smaller training datasets.

Real world experiments

$$\{SLE\}_{c=1}^{C} = (\log S_c^{t_c+\Delta t} - \log \widetilde{S}_c^{t_c+\Delta t})^2$$



Here we predict final size of the information cascade after having observed a fixed number of users. For LastFM, Memes and Digg, DANTE outperforms in terms of median and performs comparably in terms of mean.



Here we compare predictive performance against models after having observed events for a fixed amount of time t_c. Apart from CasFlow, DANTE is highly competitive with all other methods.

Conclusions

- We propose a **single** discriminative probabilistic model for all observation times $t_c$and prediction time intervals Δt.
- Our probabilistic approach renders an interpretable model that allows us to produce estimated count prediction intervals.
- Such models can easily produce answers to arbitrary user engagement based probabilistic queries.