The Hyperedge Event Model

Bomin Kim

Department of Statistics
Pennsylvania State University

July 29, 2018

Joint Statistical Meetings 2018

Collaborators


- Aaron Schein

  College of Information and Computer Sciences, UMass Amherst
- Bruce Desmarais

  Department of Political Science, Pennylvania State University
- Hanna Wallach

  Microsoft Research NYC

Motivations: Hyperedge Event Model

- Hyperedge
  edges including one sender and multiple receivers or one receiver and multiple senders.
- Event
  timestamped events
- Model
  statistical framework to jointly

"who interacts with whom, and when?"

Generative Process: "Who Interacts with Whom"

For each edge or event $e = 1, \ldots, E$,

- Receiver intensity for every sender-receiver pair $(i, j)_{i \neq j}$

$$\lambda_{iej} = \boldsymbol{b}^T \boldsymbol{x}_{iej},$$

where $\boldsymbol{x}_{iej}$ is a set of receiver selection features or covariates.

- Every sender selects candidate receivers from non-empty multivariate Bernoulli distribution $\boldsymbol{u}_{ie} \sim \text{MB}_G(\lambda_{ie1}, \ldots, \lambda_{ieA})$

$$P(\boldsymbol{u}_{ie} | \boldsymbol{b}, \boldsymbol{x}_{iej}) \propto \exp\left(\log(I(||\boldsymbol{u}_{ie}||_1 > 0)) + \sum_{j \neq i} \lambda_{iej} u_{iej} \right)$$

Generative Process: "and When"

- Timing rate for each sender

$$\mu_{ie} = g^{-1}(\boldsymbol{\eta}^T \boldsymbol{y}_{ie})$$

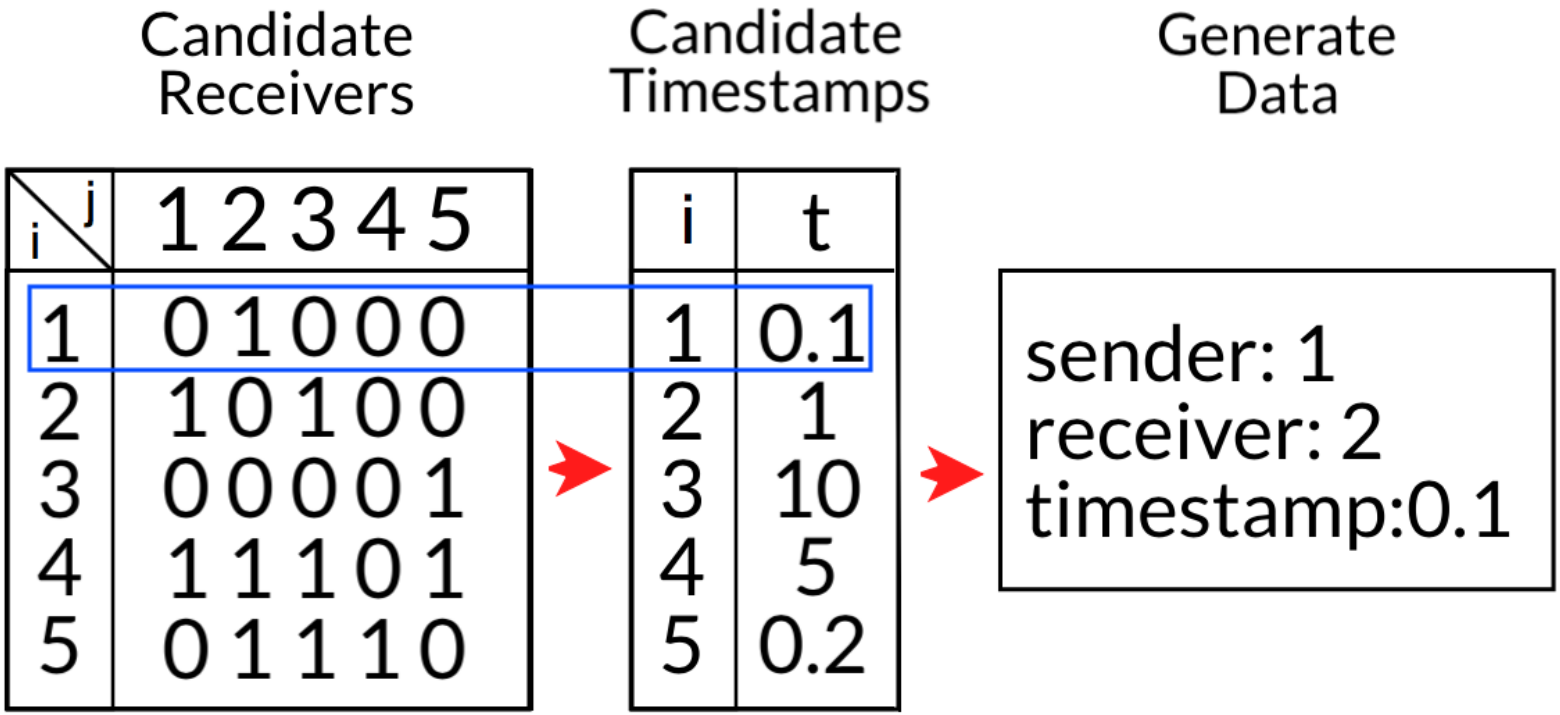
- Generalized linear model (GLM) such that time increment τ_{ie} satisfy

$$E(\tau_{ie}) = \mu_{ie} \text{ and } V(\tau_{ie}) = V(\mu_{ie})$$

- Select the sender-receiver-set with the smallest time increment

$$\begin{aligned} s_e &= \text{argmin}_i(\tau_{ie}), \\ \boldsymbol{r}_e &= \boldsymbol{u}_{s_e e}, \\ t_e &= t_{e-1} + \tau_{s_e e}. \end{aligned}$$

Generative Process: Sender, Receivers, and Timestamps



Application: Montgomery County Government Email Data

- Coefficients for receiver selection features
- Coefficients for event timing features

Results: Exploratory Analysis

- Coefficients for receiver selection features
- Coefficients for event timing features

Comparison: Lognormal vs. Exponential

- Covariates

Conclusions

- Covariates