A Federated Framework for Marked Point Processes

Zhiyuan Zhao, Zheng Dong

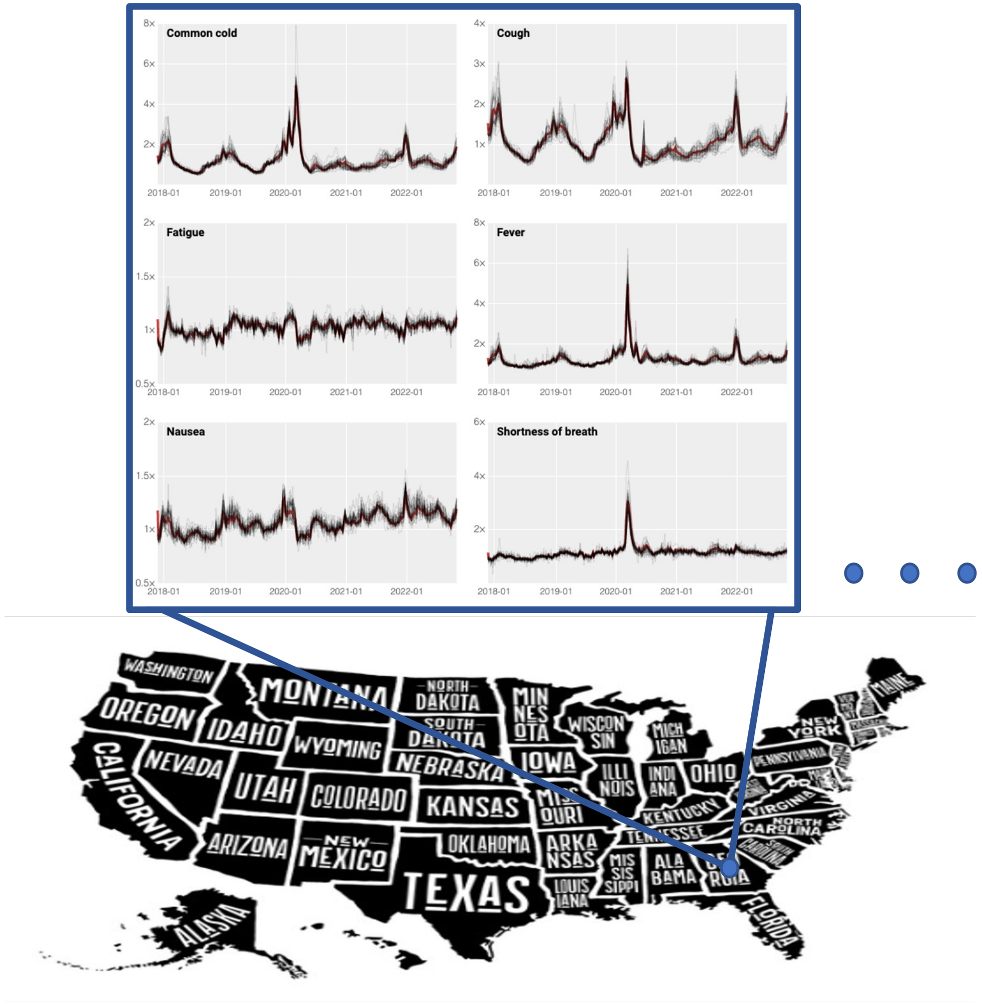
Georgia Institute of Technology

CSE 8803 Project Presentation

COVID-19

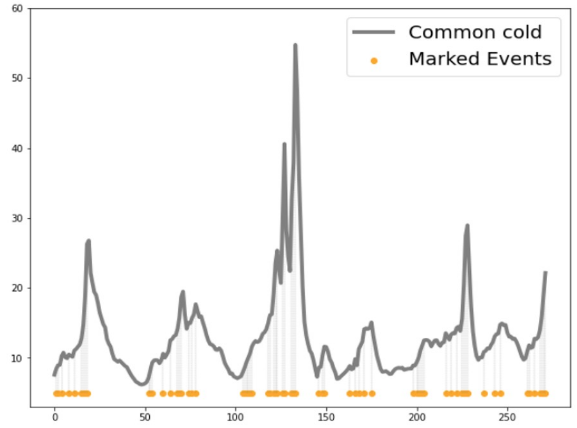
- COVID-19 Cases & Deaths
- Daily acativity & Mobility Data
- Google Symptom Search Trend Data
- ……

Google Search Trend Data

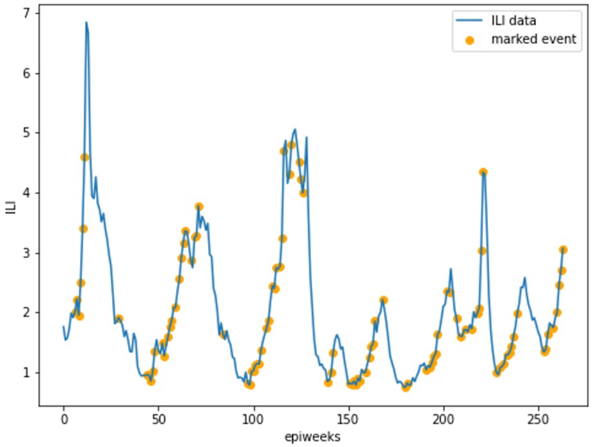


- Nationwide and State-level data (50 States & D.C.)
- A search trend of each symptom in each state and nationwide from 2017 to 2022

Self-exciting Point Processes



We find "symptom outbreak" using Google search trend data



We validate the found "symptom outbreaks" using ILI data

- Conditional intensity function:

$$\lambda_m(t) = \frac{\mathbb{E}[\mathbb{N}_m([t, t+dt])|\mathcal{H}_t]}{dt}$$

m: the m^{th} symptom.

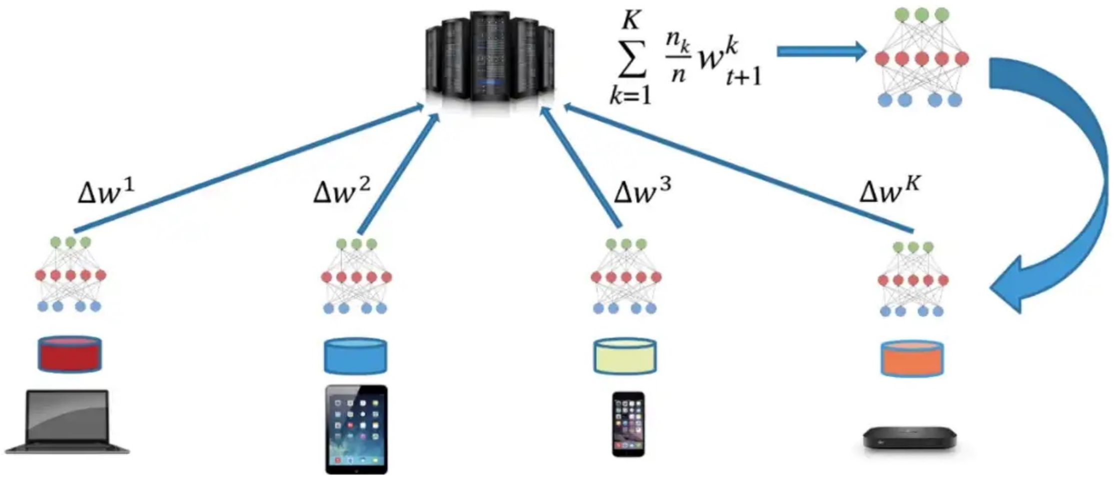
- Self-exciting point process:

$$\lambda_m(t) = \mu + \sum_{(t_i, m_i) \in \mathcal{H}_t} k(t_i, t, m_i, m)$$

- Log-likelihood based on λ_m parameterized by $\boldsymbol{\theta}$:

$$\ell(\mathcal{H}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \log \lambda_{m_i}(t_i) - \sum_{m=1}^{M} \int_0^T \lambda_m(t) dt$$

Federated Learning



- Weighted empirical risk:

$$F(\boldsymbol{\theta}) = \sum_{k=1}^{K} p_k F_k(\boldsymbol{\theta}) = \sum_{k=1}^{K} \frac{s_k}{\sum_{i=1}^{K} s_i} F_k(\boldsymbol{\theta})$$

- Federated Averaging (*FedAvg*) :

1. Get $\boldsymbol{\theta}_k^{t+1}$ by updating $F_k(\boldsymbol{\theta}^t)$
2. $\boldsymbol{\theta}^{t+1} = \frac{1}{K} \sum_k \boldsymbol{\theta}_k^{t+1}$

- Federated Proximal (*FedProx*):

1. Get $\boldsymbol{\theta}_k^{t+1}$ by updating
   $h_k(\boldsymbol{\theta}, \boldsymbol{\theta}^t) = F_k(\boldsymbol{\theta}^t) + \frac{a}{2} ||\boldsymbol{\theta} - \boldsymbol{\theta}^t||^2$
2. $\boldsymbol{\theta}^{t+1} = \frac{1}{K} \sum_k \boldsymbol{\theta}_k^{t+1}$

Final Model

Georgia Institute of Technology

Minimize:

$$\underbrace{F(\boldsymbol{\theta}) = \sum_{k=1}^{K} p_k F_k(\boldsymbol{\theta})}_{\text{Federated Learning}} = -\sum_{k=1}^{K} \frac{s_k}{\sum_{i=1}^{K} s_i} \underbrace{\ell_k\left(\mathcal{H}^{(k)}, \boldsymbol{\theta}\right)}_{\text{Point Process Model}}$$

Experiment Setup

Georgia Institute of Technology

- Model Comparison:
  1. *FedProx* with point process
  2. *FedAvg* with point process
  3. Vanilla point process

- In model training, we fit the point process model for different symptom peaks using:
  1. State-level symptom data during epiweeks from 2017 to 2021 for *FedProx* and *FedAvg*, treat each state as a client.
  2. Nationwide symptom data during epiweeks from 2017 to 2021 for vanilla point process.

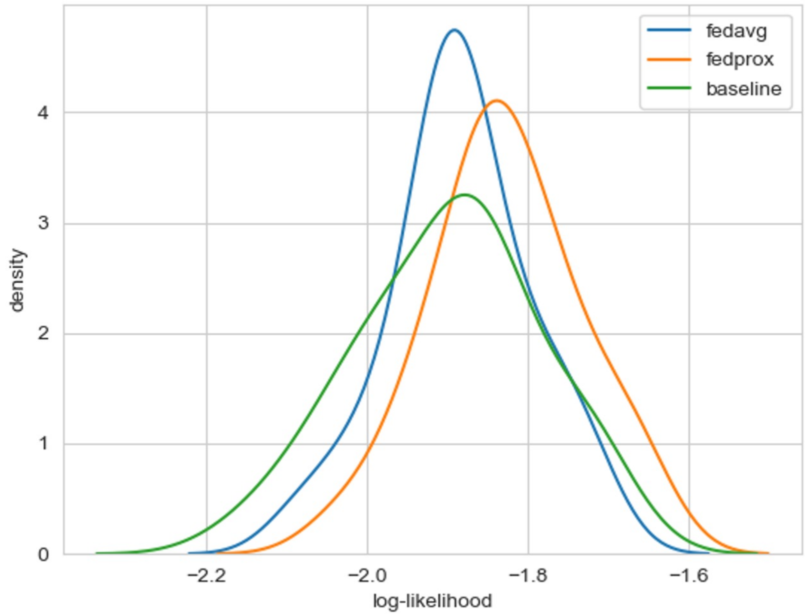
- In model testing, we test our learned models (By *FedProx, FedAvg,* and vanilla point process) using symptom data in each state during epiweeks from 2021 to 2022.
- **Criteria**: log-likelihood of observing the testing events.

Results

Observation 1: Federated paradigm outperforms vanilla point process in validation log-likelihood

	Baseline	FedAvg	FedProx
Log-likelihood	$-1.904_{\pm 0.004}$	$-1.856_{\pm 0.028}$	$-1.849_{\pm 0.022}$

Observation 2: *FedProx* has better 'worst case' guarantee than other two methods



Conclusion

What we found:

- Federated learning frameworks provide better performance than vanilla point process
- Federated learning frameworks, particularly *FedProx*, have better worst case guarantee

Future work:

- Alternative data resources
- Better prediction model
- Advanced Federated Learning framework
- ……