Feature Driven and Point Process Approaches for Popularity Prediction

Swapnil Mishra, Marian-Andrei Rizoiu, Lexing Xie
The Australian National University and Data61, Australia

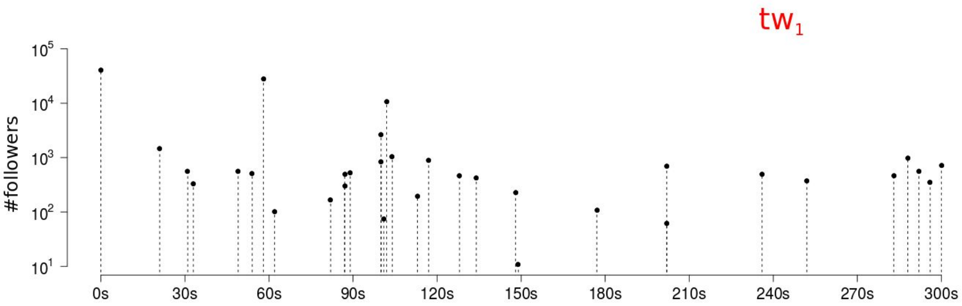
Australian National University

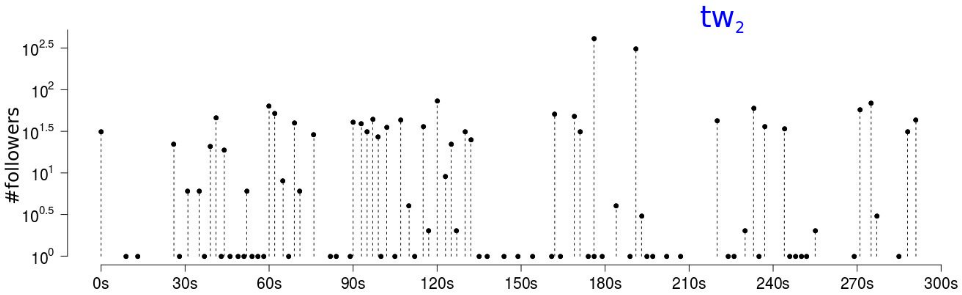
DATA 61

CSIRO

Leonard Nimoy, Spock of 'Star Trek,' Dies at 83

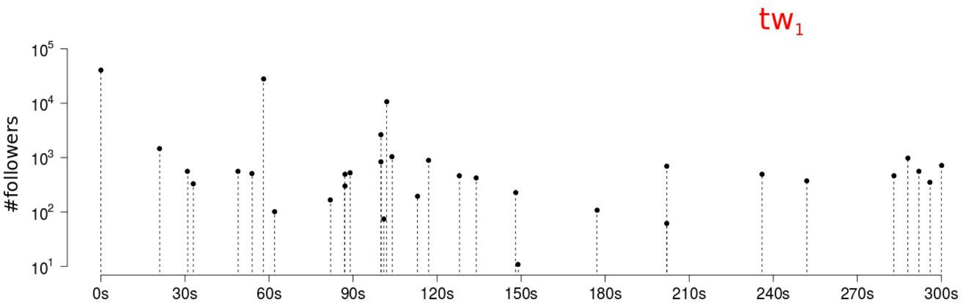
http://www.nytimes.com/2015/02/27/...

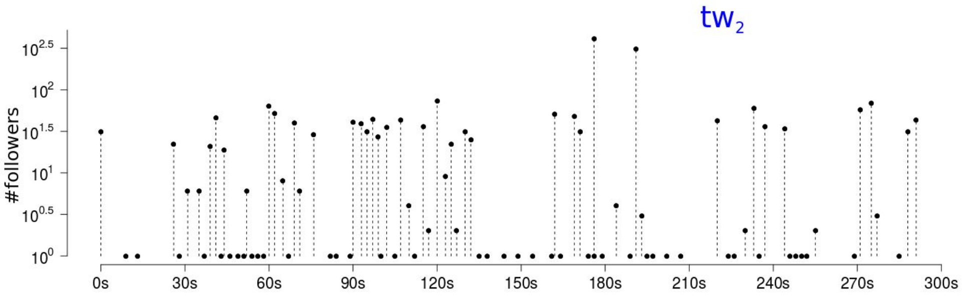




Leonard Nimoy, Spock of ‘Star Trek,’ Dies at 83

http://www.nytimes.com/2015/02/27/...





Which cascade will gather more attention?

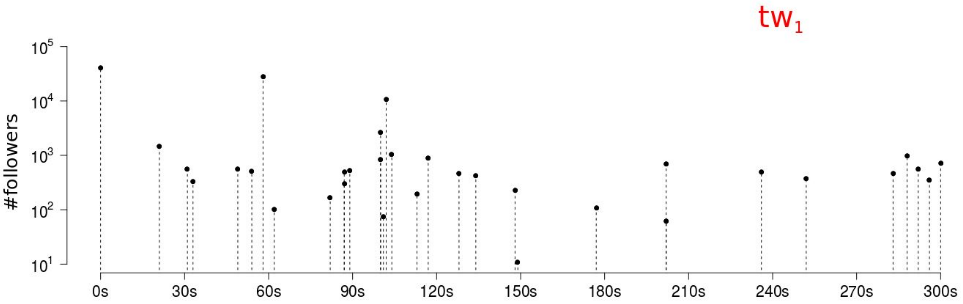
A. tw_1

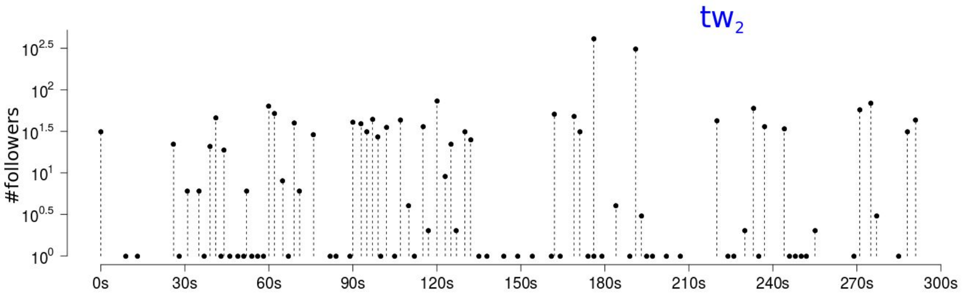
B. tw_2

C. same(∓5%)

Leonard Nimoy, Spock of ‘Star Trek,’ Dies at 83

http://www.nytimes.com/2015/02/27/...





Which cascade will gather more attention?

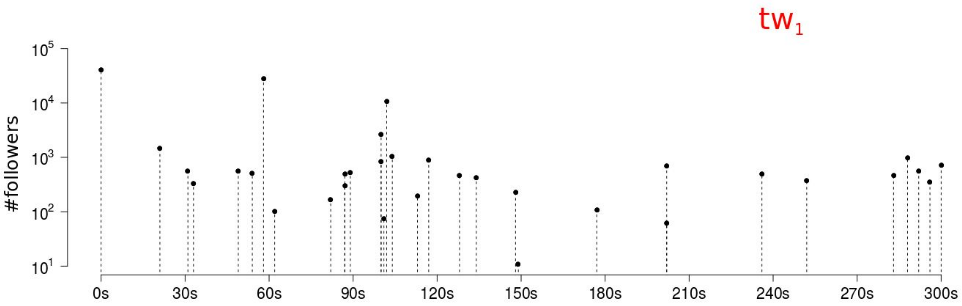
A. tw_1

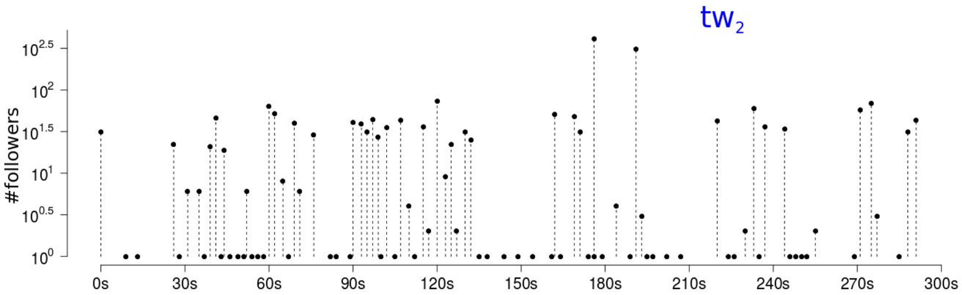
B. tw_2

C. same(∓5%)

Leonard Nimoy, Spock of 'Star Trek,' Dies at 83

http://www.nytimes.com/2015/02/27/...





Which cascade will gather more attention?

A. tw_1

B. tw_2

C. same (∓5%)

Prediction is difficult!

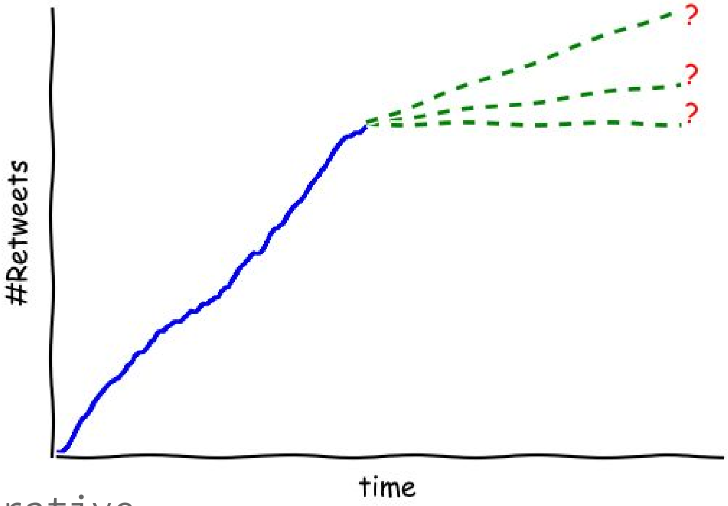
How to use event times and user features?

Popularity

Problem: predict the number of retweets on Twitter

Why do we care about it??

- Monetisation
- Personalization



Our Contributions

- Bridge gap:
  - Approaches: feature-driven vs generative
  - Problem setting: regression vs classification
- One new benchmark dataset:
  - Features
  - Event times

Outline

- Popularity prediction problem
- Self-exciting point process
- Feature-driven approaches
- Results

Existing Solutions

Approach:

Feature Driven: Cheng(2014), Martin(2016), Pinto(2013)

Generative: Zhao(2015), Ding(2015), Shen(2014)

Problem Setting:

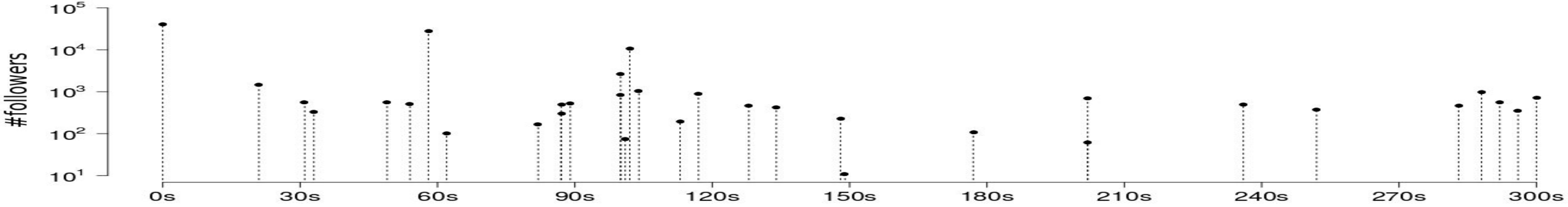
Regression: Zhao(2015), Shen(2014)

Classification: Cheng(2014), Shamma(2011)

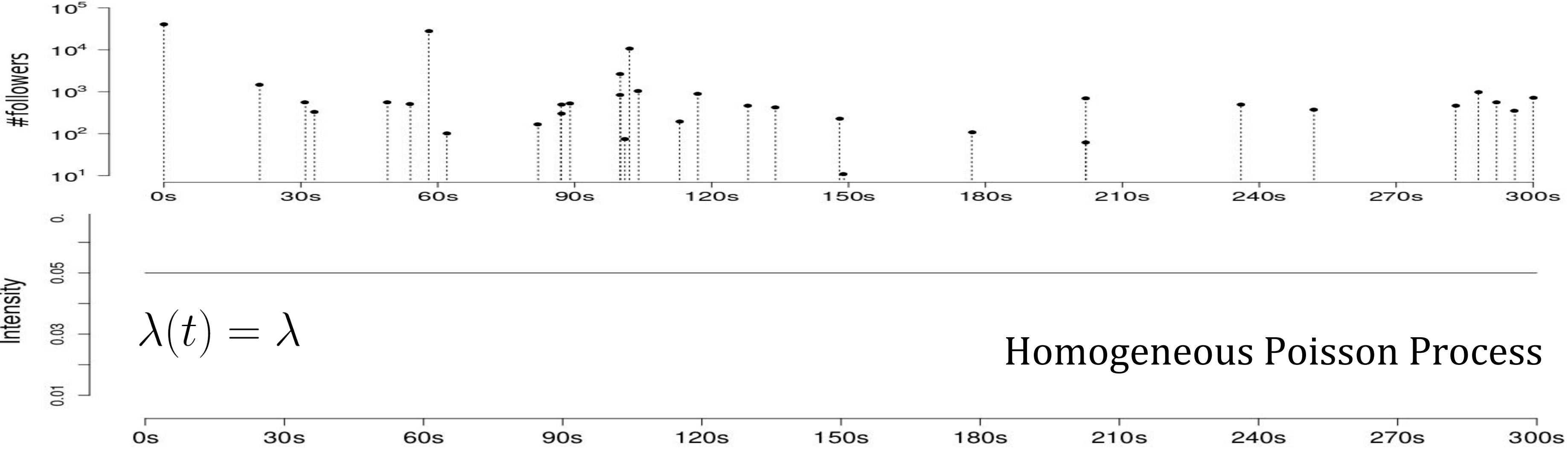
Open Questions:

- Can we have best of both generative and feature-driven models?
- How useful are features computed over data available through Public APIs?

Self-Exciting Point Processes



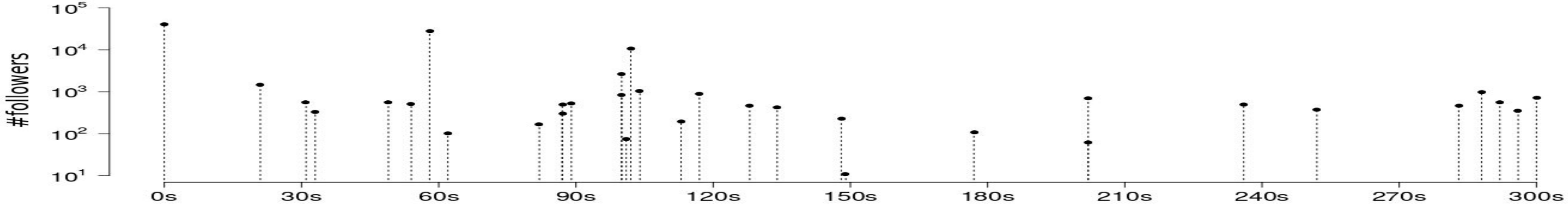
Self-Exciting Point Processes

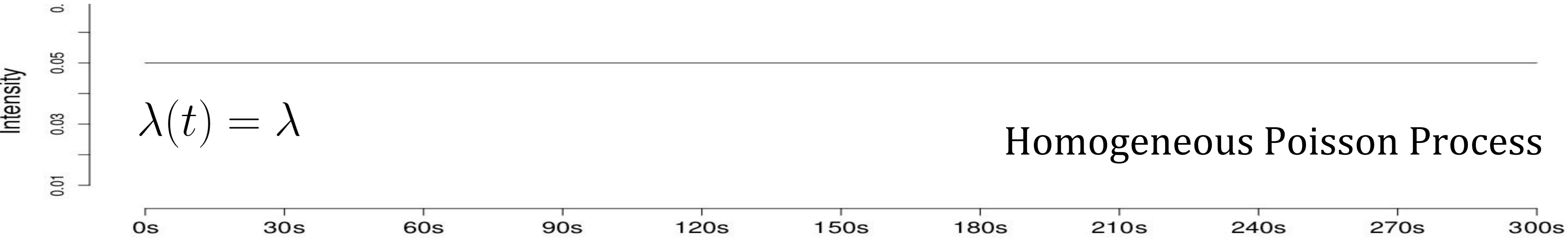


$$\lambda(t) = \lambda$$

Homogeneous Poisson Process

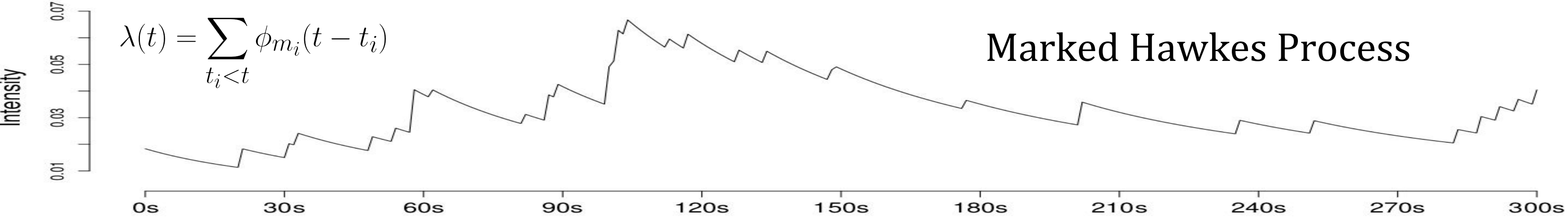
Self-Exciting Point Processes





$$\lambda(t) = \lambda$$

Homogeneous Poisson Process



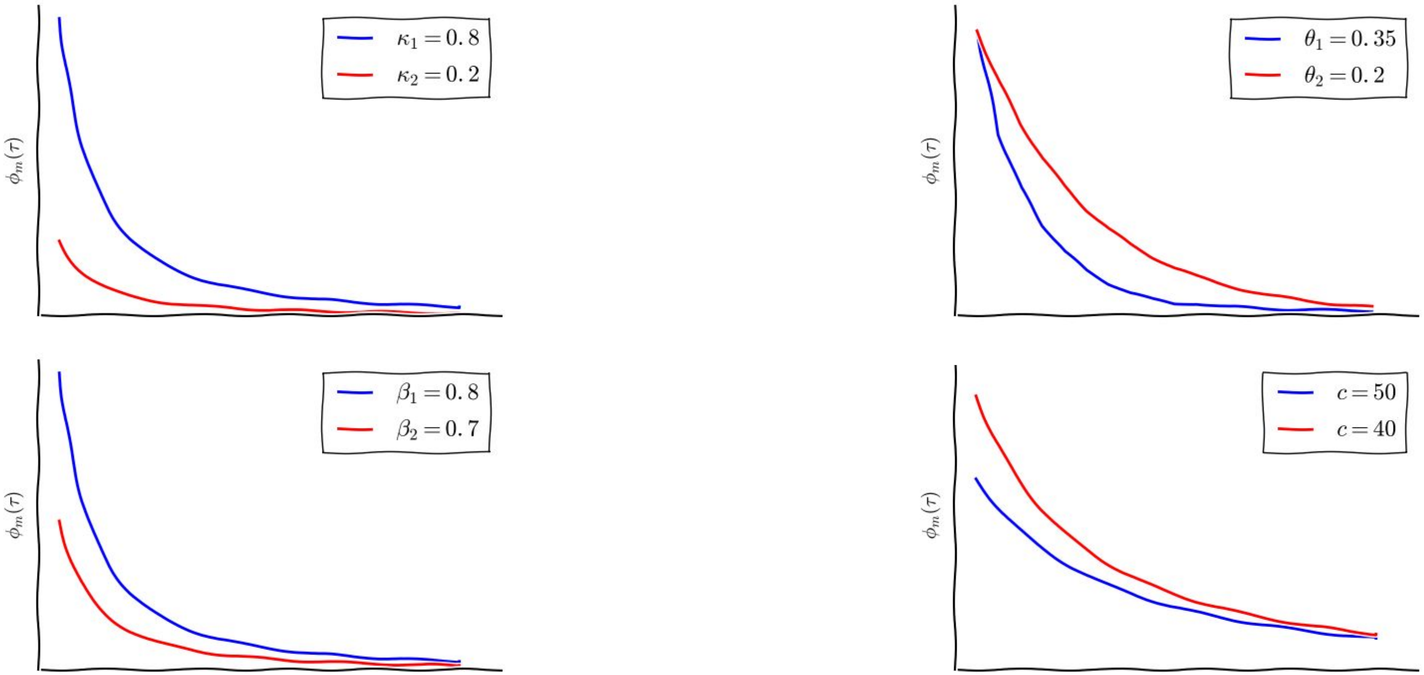
$$\lambda(t) = \sum_{t_i < t} \phi_{m_i}(t - t_i)$$

Marked Hawkes Process

Kernel for Marked Hawkes

content virality   user influence   memory

$$\phi_m(\tau) = \kappa m^{\beta} (\tau + c)^{-(1+\theta)}$$



Estimating Marked Hawkes Proc.

$$\mathcal{L}(\kappa, \beta, c, \theta) = \sum_{i=2}^{n} \log \kappa + \sum_{i=2}^{n} \log \left(\sum_{t_j < t_i} \frac{(m_j)^{\beta}}{(t_i - t_j + c)^{1+\theta}} \right) - \kappa \sum_{i=1}^{n} (m_i)^{\beta} \left[\frac{1}{\theta c^{\theta}} - \frac{(T + c - t_i)^{-\theta}}{\theta} \right]$$

To have $n^* < 1$, we use IPOPT (Wächter2006)

$$n^* = \kappa \left(\frac{\alpha - 1}{\alpha - \beta - 1} \right) \frac{1}{\theta c^{\theta}},$$

$$\text{for } \beta < \alpha - 1 \text{ and } \theta > 0$$

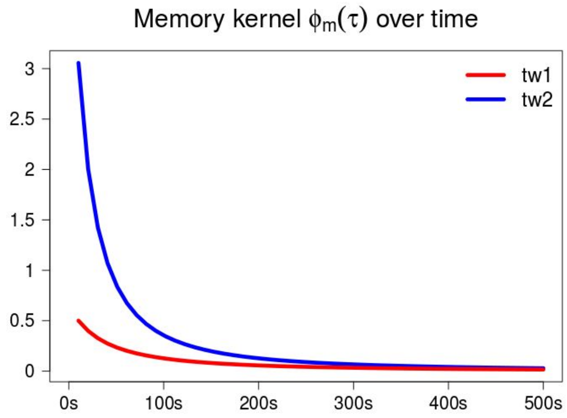
Estimating Marked Hawkes Proc.

$$\mathcal{L}(\kappa, \beta, c, \theta) = \sum_{i=2}^{n} \log \kappa + \sum_{i=2}^{n} \log \left(\sum_{t_j < t_i} \frac{(m_j)^{\beta}}{(t_i - t_j + c)^{1+\theta}} \right) - \kappa \sum_{i=1}^{n} (m_i)^{\beta} \left[\frac{1}{\theta c^{\theta}} - \frac{(T + c - t_i)^{-\theta}}{\theta} \right]$$

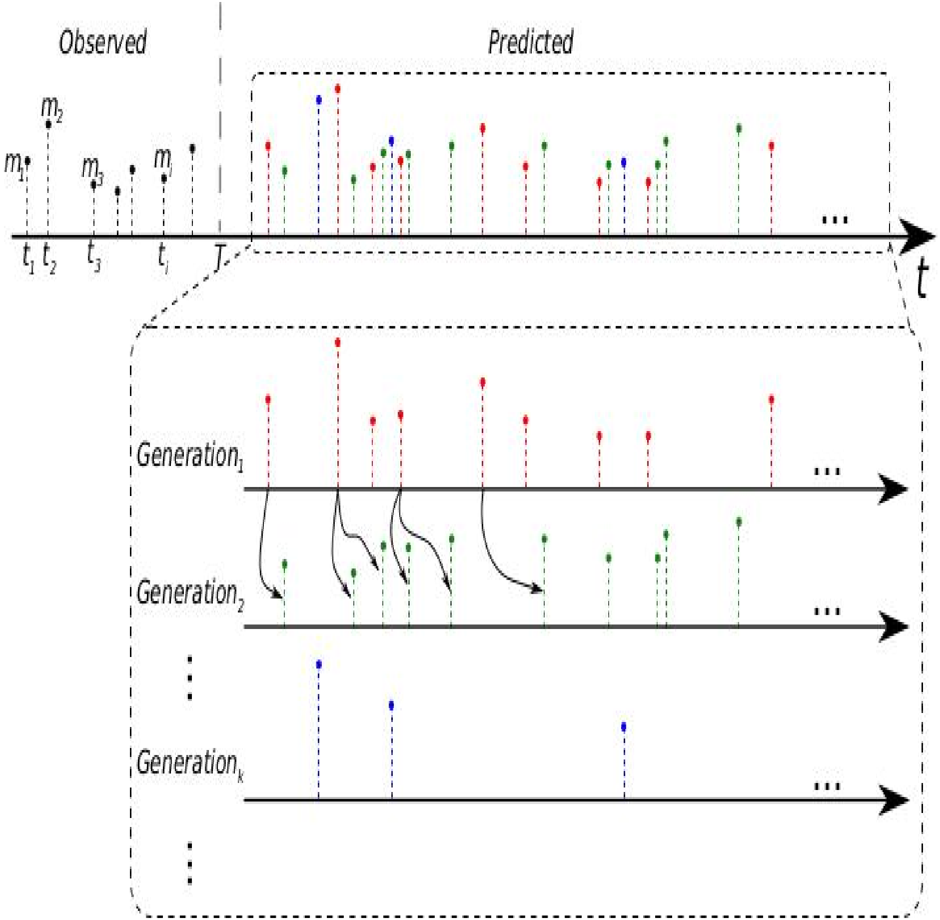
To have $n^* < 1$, we use IPOPT(Wächter2006)

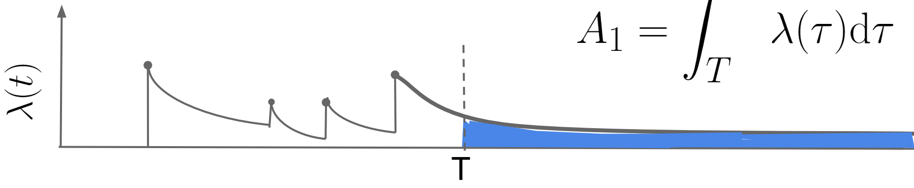
$$n^* = \kappa \left(\frac{\alpha - 1}{\alpha - \beta - 1} \right) \frac{1}{\theta c^{\theta}},$$

$$\text{for } \beta < \alpha - 1 \text{ and } \theta > 0$$



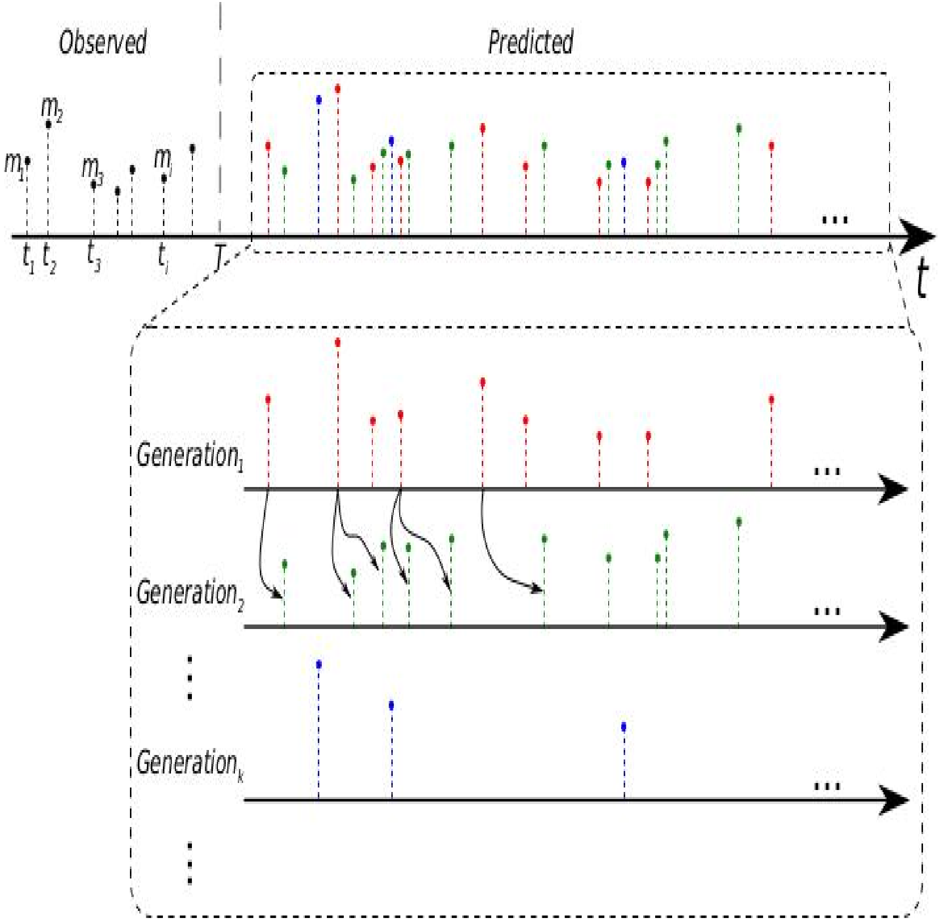
From Event Series to Predictions

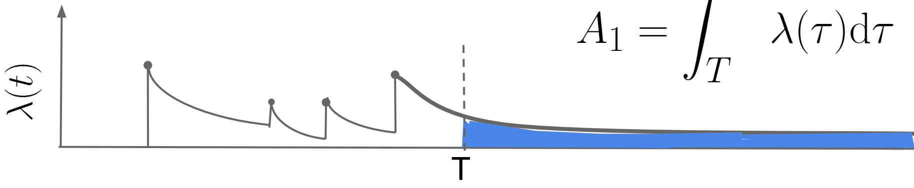




$$A_1 = \int_T^\infty \lambda(\tau)\mathrm{d}\tau$$

From Event Series to Predictions





$$A_1 = \int_T^\infty \lambda(\tau) \mathrm{d}\tau$$

$$N_\infty = n_{obs} + \sum_{i=1}^{\infty} A_i$$

$$= n_{obs} + \left(\frac{A_1}{1 - n^*} \right)$$

Further Improving Prediction

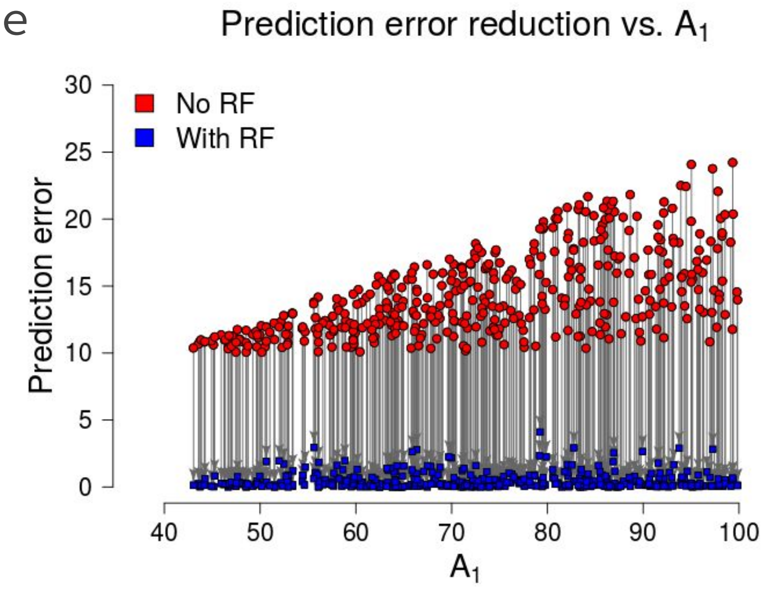
Limitations for predictions:

- #followers approximates influence
- Assumes fixed parametric kernel
- Local minima in parameter estimation

Further Improving Prediction

Limitations for predictions:

- #followers approximates influence
- Assumes fixed parametric kernel
- Local minima in parameter estimation



Predictive Layer:

$$\hat{N}_\infty = n_{obs} + \omega \left(\frac{A_1}{1 - n^*} \right)$$

$$\omega = RandomForest\left(c, \theta, A_1, n^*\right)$$

Outline

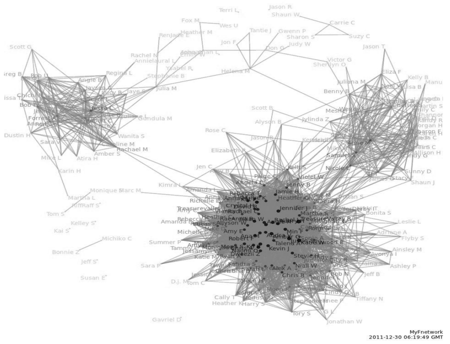
- Popularity prediction problem
- Self-exciting point process
- Feature-driven approaches
- Results

Features

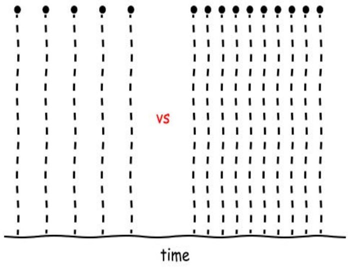
Criterion for selection:

- can be computed on data from Public API’s
- shown to perform well

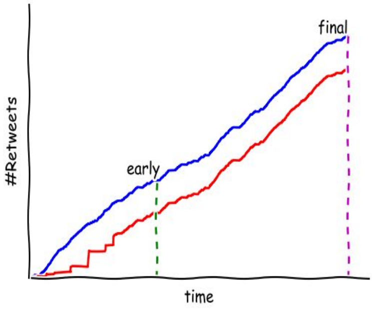
User Features (Cheng2014, Martin 2016)



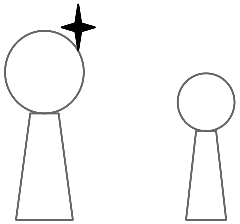
Temporal Features (Cheng2014)



Early volume (Szabo2010, Pinto2013)



Past Success (Bakshy2011, Martin2016)



Datasets

Tweet-1Mo

- Provided by Zhao et.al(SEISMIC)
- One month tweets
- Random Sample of 30.5K cascades > 50

NEWS

- April’15 to July’15 english news
- 49.7 million tweets
- 110K cascades > 50

Datasets

Tweet-1Mo

- Provided by Zhao et.al(SEISMIC)
- One month tweets
- Random Sample of 30.5K cascades > 50

NEWS

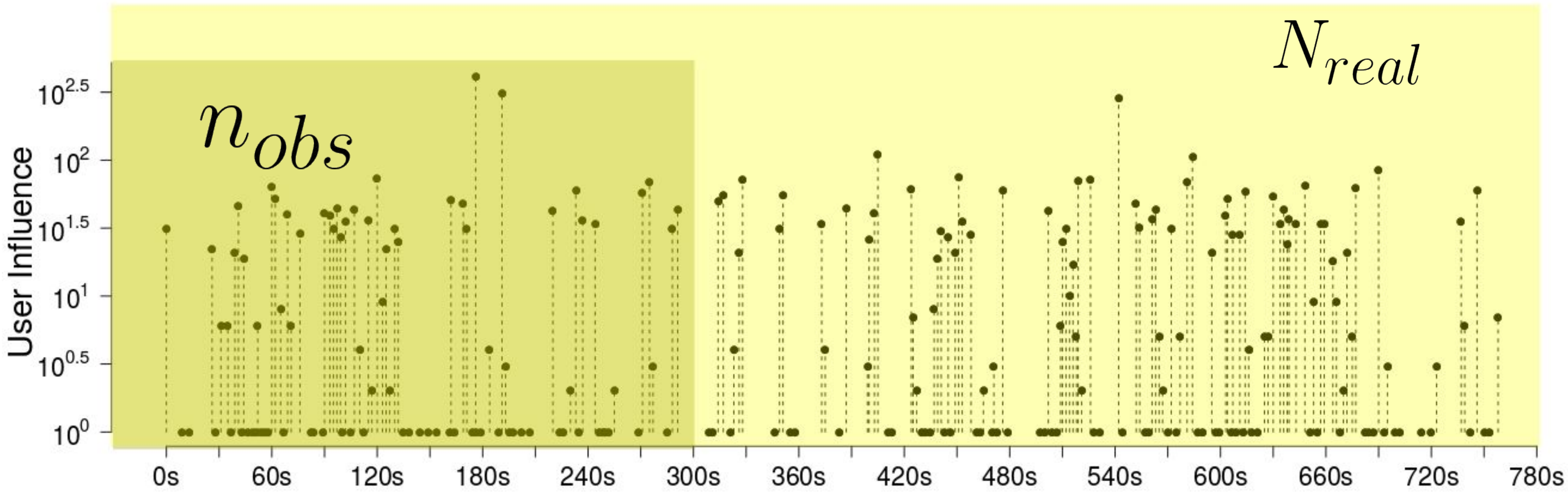
- April’15 to July’15 english news
- 49.7 million tweets
- 110K cascades > 50

Feature-Driven

Generative

NEWS Available Now

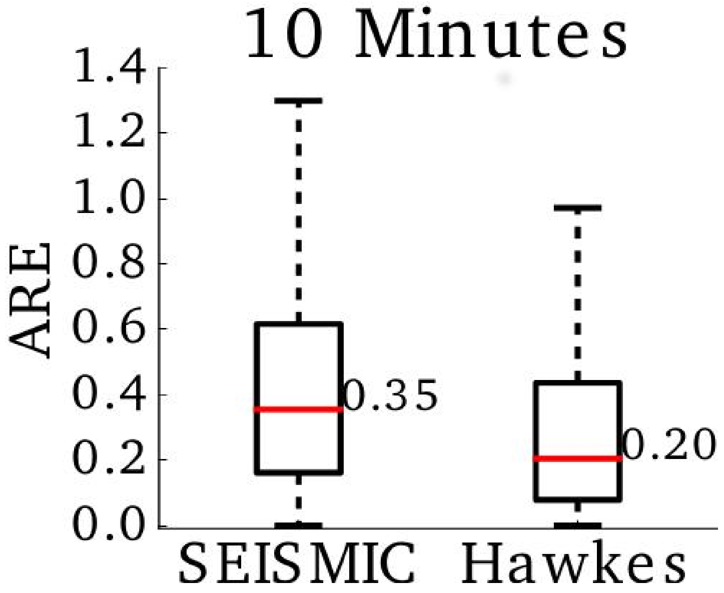
Evaluation Setup

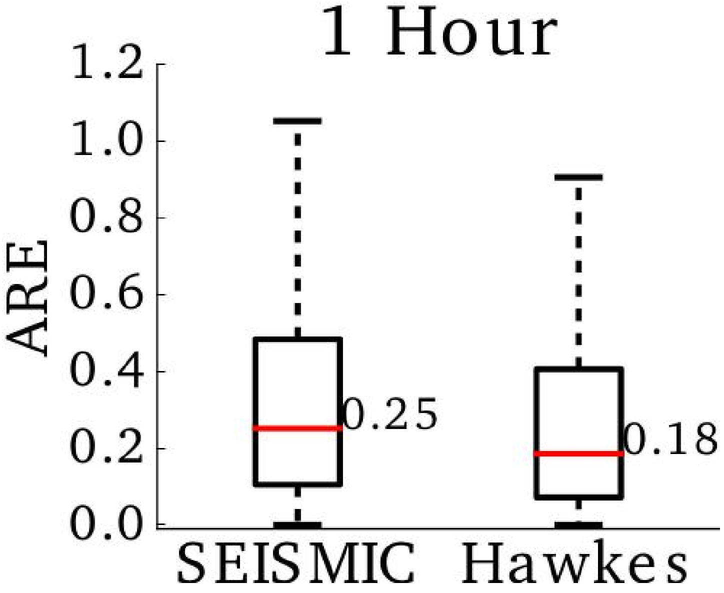


- 10-fold cross-validation for training
- Absolute Relative Error(ARE): $\dfrac{\left|\hat{N}_{\infty} - N_{real}\right|}{N_{real}}$

Results on *Tweet-1Mo*

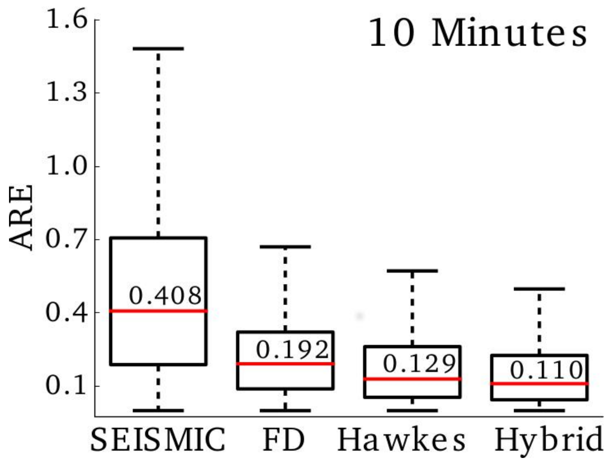
Mean(std.dev) ARE		10 Mins	1 Hour
	SEISMIC	0.70(15.58)	0.51(10.81)
	HAWKES	0.33(0.41)	0.30(0.38)

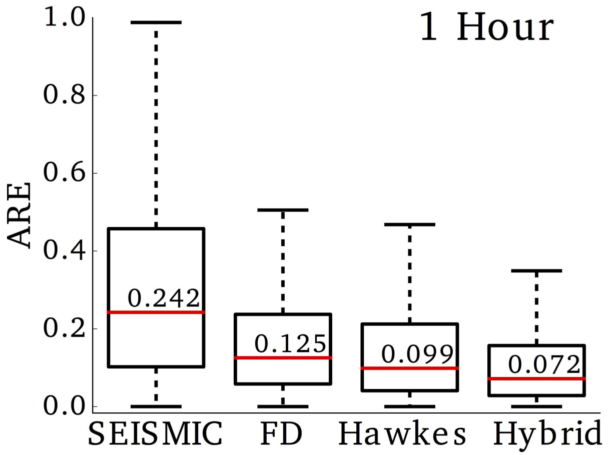




Results on *NEWS*

		10 Mins	1 Hour
Mean(std.dev) ARE	SEISMIC	0.71(4.89)	0.32(0.40)
	Feature-Driven	0.22(0.17)	0.17(0.14)
	HAWKES	0.22(0.80)	0.17(0.36)
	Hybrid	0.15(0.14)	0.11(0.12)

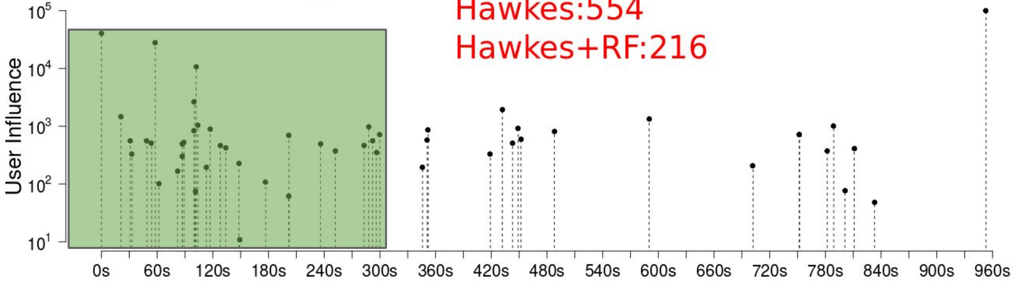


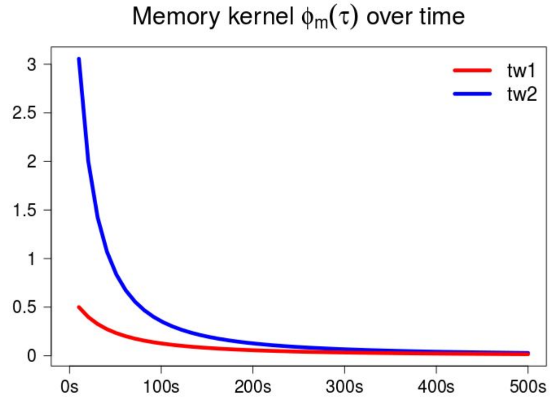


Example Prediction

tw1

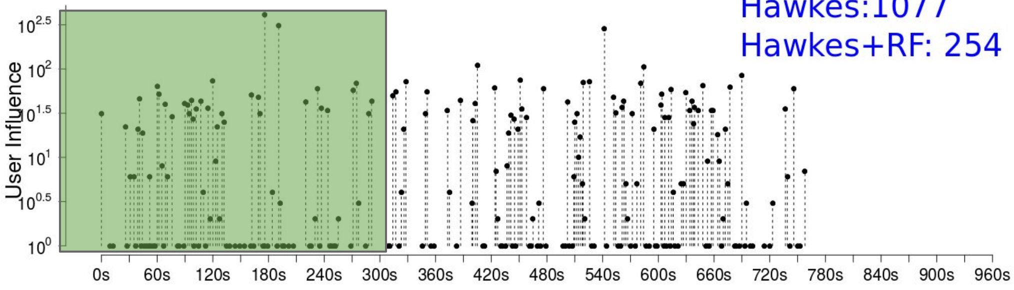
Size:219
Hawkes:554
Hawkes+RF:216

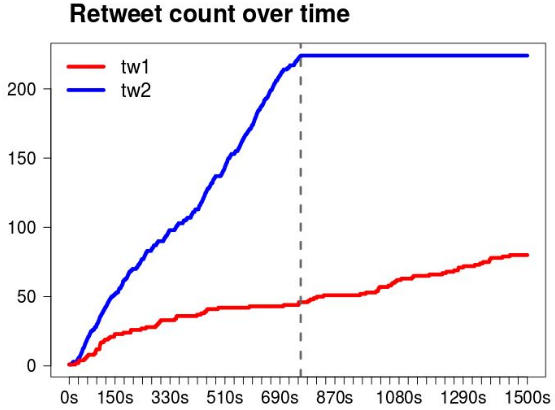




tw2

Size:224
Hawkes:1077
Hawkes+RF: 254





Conclusion

- Established common benchmark to compare feature driven and generative approaches

Conclusion

- Established common benchmark to compare feature driven and generative approaches
- Generative explanatory model with a predictive layer outperforms current state of the art

Conclusion

- Established common benchmark to compare feature driven and generative approaches
- Generative explanatory model with a predictive layer outperforms current state of the art
- Small set of features with only message content and basic user features generalizes over problem space

Conclusion

- Established common benchmark to compare feature driven and generative approaches
- Generative explanatory model with a predictive layer outperforms current state of the art
- Small set of features with only message content and basic user features generalizes over problem space
- Future: interplay between related cascades and RNN based popularity models

Data+Code:https://github.com/s-mishra/featuredriven-hawkes