Time varying Topics for Modeling Content Diffusion Over Social Network

Pushkar Nagar
Advisor: Prof. Chiranjib Bhattacharyya

Department of Electrical Engineering
Indian Institute of Science.

14^{th} *June* 2017

Outline

Introduction

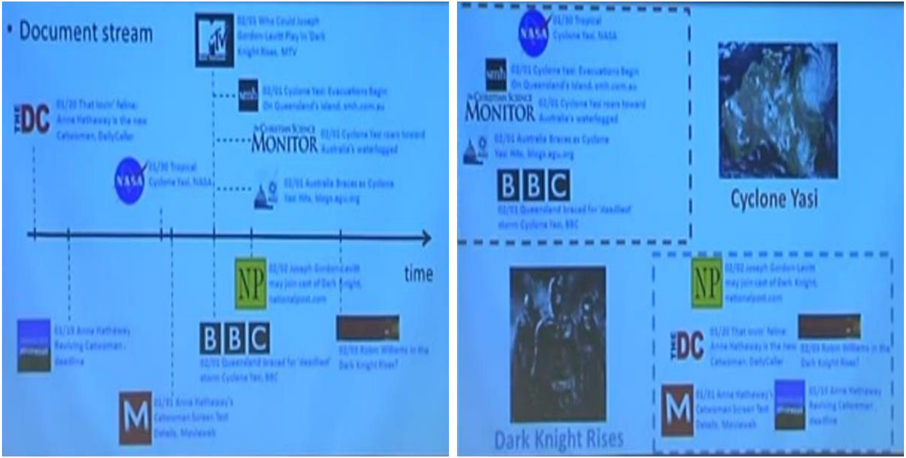


Image Credits[1]

Introduction

- Clustering of large number of documents like blogs, news articles, social site data (twitter streams).
- This type of data has two information: time and content.
- Leverage of both the information, to cluster the documents more accurately.
- Understanding the Temporal Dynamics of arriving of Documents.
- Understanding the Information Diffusion Network.

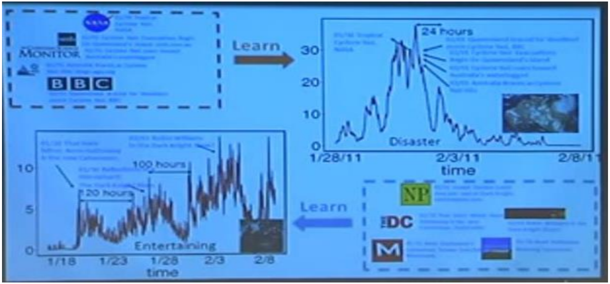


Image Credits[1]

Motivation

- Extracting stories, real life events from continuous time document streams.
- To understand the **Information Diffusion** at broader level(cluster Level across network).
- Learning temporal dynamics and predicting future trends.

Problem Definition

Problems we are addressing are as follows:

- Clustering of documents and finding topics of the cluster.
- Capturing Temporal dynamics of each cluster.
- Studying Information Diffusion at broader level(across network, ex: social network).

Related Work

- Xinran He[4] proposed a joint model for Network Inference and Topic Modeling but they do not consider temporal dynamics of clusters.
- S. Liang[5] proposed two dynamic topic models: short term dependency and long term dependency inference, but it assumes the discretization of time interval.
- D. M. Blei[2] proposed model where the parameters at time t comes from $t - 1$, but it does not capture the changes in temporal dynamics of clusters.

Related Work Contd..

- Nan Du[14] proposed a time series model based on Multi Variate Hawkes Process, but it does not take into account content.
- S. Hosseini[9] proposed a joint model of time and content, in which they consider the influence matrix consisting of network of users. We have used the same dataset namely Event Registry.
- Most of the work has two main drawbacks:
  - Single source of information
  - Time is not considered which is important in influential posting of documents.

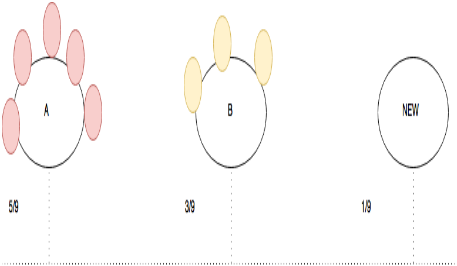
Background

- Dirichlet Process -

$$G \sim DP(\alpha, G_0) \qquad G_0 \text{ is Base Distribution}$$

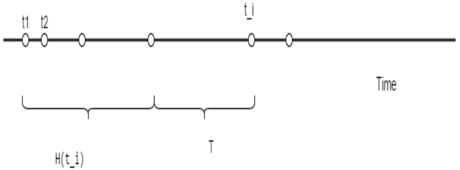
$$\theta_1, \theta_2 \; .... \; \theta_n \sim G \qquad \alpha \text{ is concentration parameter}$$

- The Chinese Restaurant Process



Point Process

- Temporal Point Process -



  Point Process is a Stochastic process modeling temporal point patterns. A temporal point pattern can be explained as a sequence of times of events. Thus Point Process is a class of **counting processes**.

- Point Process is characterized by conditional intensity function defined as :

$$a \tag{1}$$

  where $f^*(t) = f(t|H_t)$ is conditional density function of the time of the next event t_{n+1} given the history of previous events $(..., t_{n-1}, t_n)$

Point Process contd..

- Point process finds its applications in :
  - Sequence of arrivals of requests at a server
  - Sequence of earthquakes
  - Modeling crimes
  - Modeling financial contagions
- The well known Point Processes are :
  - Homogeneous Poisson Process: $\lambda(t) = \lambda_0$
  - Non-Homogeneous Poisson Process: $\lambda(t)$
  - Hawkes Process: Superposition of Homogeneous Poisson and Non-Homogeneous Process

$$\lambda(t) = \lambda_0 + \sum_{t_i \epsilon H_t} \nu(t - t_i) \tag{2}$$

Hawkes Process

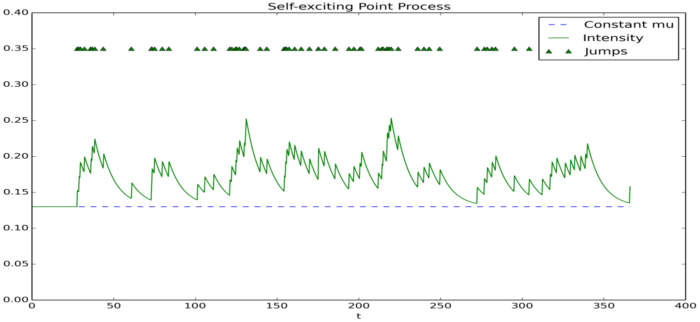
- The Hawkes process is a class of self or mutually exciting point process models (Hawkes, 1971) [4]
- A univariate Hawkes process is defined as

$$\lambda^*(t) = \mu(t) + \alpha \sum_{t_j < t} \gamma(t - t_j; \beta) \tag{3}$$

  where $\mu : \mathbb{R} \longrightarrow \mathbb{R}_+$ is deterministic base intensity, α is a positive parameter and $\gamma(t; \beta)$ is a density function on $(0,\infty)$ depending on parameter β
- If the density is taken to be exponential & μ to be constant , i.e.

$$\lambda^*(t) = \mu + \alpha \sum_{t_j < t} \exp(-(t - t_j)) \tag{4}$$

Simulation of Hawkes Process



A simulation of the Hawkes process with parameter $(\mu, \alpha, \beta) = (0.13, 0.023, 0.11)$. If density is taken to be exponential decaying kernel $= \alpha * \exp^{-\beta * t}$ where $t > 0$. Image Credits[13]

Multi-Dimensional Hawkes Process

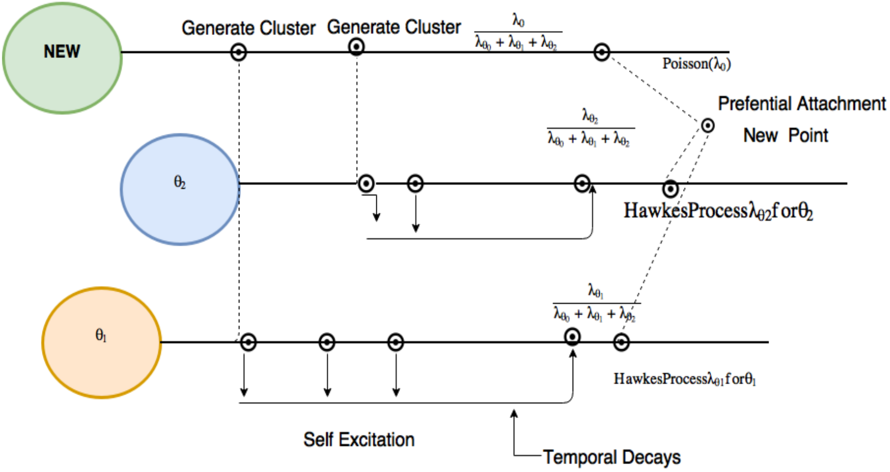
- The multivariate Hawkes process is a multi-dimensional extension to the univariate case (Hawkes, 1971; Liniger, 2009[6]).
- It is characterized by conditional intensity functions $\lambda_i^*(t)$ for each dimension $i \in I$. The intensity function $\lambda^* = [\lambda_1^*, ..., \lambda_D^*]^T$ is defined by

$$\lambda_i^*(t) = \mu_i + \sum_{t_l^j < t} \alpha_{ji} \kappa(t - t_l^j) \tag{5}$$

where

- μ_i is base intensity of dimension i
- $\kappa : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ is a time-decaying triggering kernel
- $\alpha \in \mathbb{R}^+_{D \times D}$ is infectivity matrix characterizing the structure of the network
- $\sum_{t_l^j < t} \alpha_{ji} \kappa(t - t_l^j)$ quantifies the influence of historical events on the instantaneous rate of event at time t in dimension i

Overview of Dirichlet Hawkes Process



Dirichlet Hawkes Process Model

1. Draw t_1 from Poisson(λ_0) and θ_1 from Dir($\theta|\theta_0$).
2. For each word v in document 1 : $w_1^v \sim$ Multi(θ_1)
3. For n $>$ 1:
    1. Draw $t_n >\ t_{n-1}$ from Poisson($\lambda_0+\sum_{i=1}^{n-1}\gamma_{\theta\,i}(t_n,t_i)$), where $\gamma_{\theta\,i}(t_n\ ,\ t_i)$ is kernel for Hawkes Process.
    2. Draw θ_n from Dir($\theta|\theta_0$) with probability $\frac{\lambda_0}{\lambda_0+\sum_{i=1}^{n-1}\gamma_{\theta_i}(t_n,t_i)}$, and draw α_{θ_n} from Dir($\alpha|\alpha_0$)
    3. Reuse previous θ_k for θ_n with probability $\frac{\lambda_{\theta_k}(t_n)}{\lambda_0+\sum_{i=1}^{n-1}\gamma_{\theta_i}(t_n,t_i)}$, where $\lambda_{\theta_k}(t_n) = \sum_{i=1}^{n-1}\gamma_{\theta_i}(t_n,t_i)I[\theta_i = \theta_k]$.
    4. For each word v in document n: $w_n^v \sim \mathit{Multi}(\theta_n)$

contd..

Triggering kernel funtion of Hawkes process is represented as a non-negative combination of K base kernel functions:

$\gamma_{\theta_i}(t_i, t_j) = \sum_{l=1}^{K} \alpha_\theta^l \,.\, \kappa(\tau_l, t_i - t_j)$

where α_θ draws from Dirichlet Distribution, which helps in tracking the different evolving temporal dynamics of clusters , $t_j < t_i$, τ_i represents the typical reference time points.

Inference

1. To calculate Posterior Distribution, Inference technique alternates between two subroutines:
    1. Sample the latent cluster membership by Sequential Monte Carlo Method.
    2. Updating the learned triggering kernel of respective cluster.

Generative Model

- Dirichlet Hawkes Process Model does not capture how information diffusion occurs among the various dependent source of documents.
- Analyzing events over a network created by sources.

Let D(t) $=\{e_i\}_{i=1}^{N(t)}$ denotes the set of events observed until time t, where event e$_i$ is a triplet (t_i, u_i, d_i) which indicates that at time t_i, user u_i shares document d$_i$.

- $\{\phi_k\}_1^{K(t)}$ denotes set of topics over network until time t.
- $\{\psi_{uk}\}_{i=1}^{K_u(t)}$ denotes topics belonging to interest of user u.
- $K_u(t)$ denotes number of topic user u is interested in at time t.
- $N(t)$ represents the number of events until time t.

- Users will form a network such that one user can trigger other users.
- Event triggered by preceding event has same topic index.

Generative Process

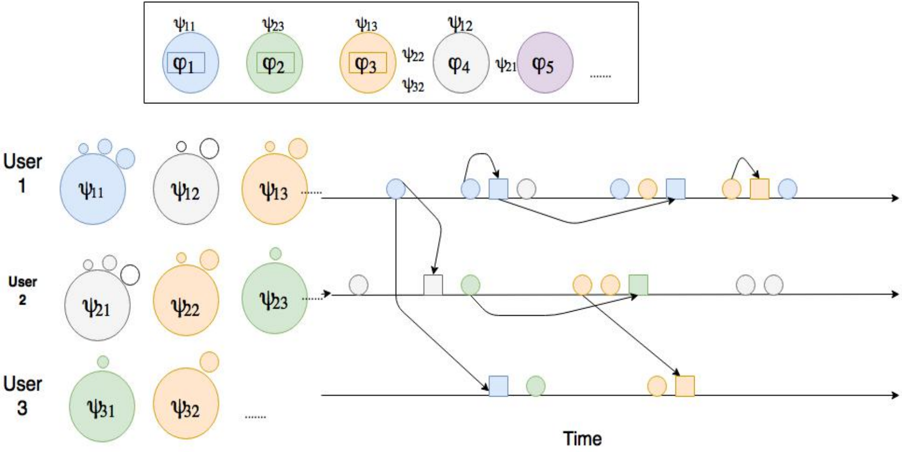


Figure: Top level shows the popularity of topics over network. Interest of each user corresponds to distribution over the topics. Circle show the events generated due to base intensity while square shows the triggered events. The arrows show the triggering relationship.

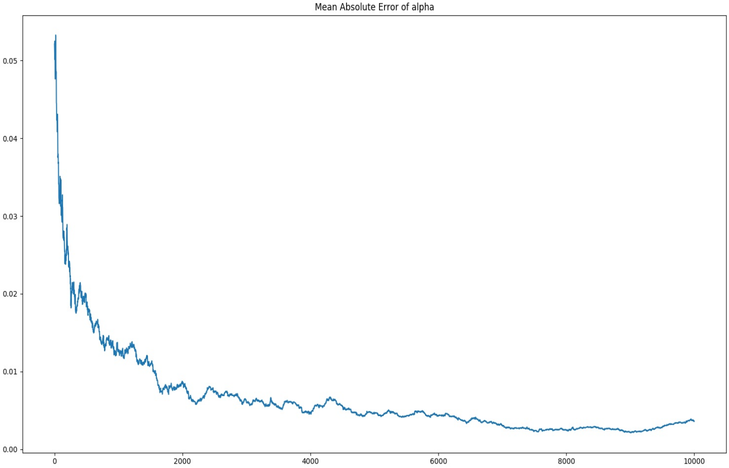
Cont.

1. For all events e_i:
   1. User u publishes documents which follows a Hawkes process with intensity function: $\lambda_u(t) = \mu_u + \sum_{s=1}^{N(t)-1} \lambda_u(t,s)$
   2. $\lambda_u(t,s) = \alpha_{u_s,u}\kappa_{z_s}(t,t_s)$ and $\kappa_{z_s}(t,t_s) = \exp\left(-\beta_k(t-t_s)\right)$
   3. User u draws reused topic ψ_{uj} with probability $\frac{n_{uj}(t)}{n_{u:}(t)+\gamma}$, where $\mathrm{n}_{uk}(t) = \mu_u + \sum_{e \epsilon D_u^0(t)} \exp\left(-\nu(t-t_e)\right) I[\theta_e = \psi_{uk}]$
   4. User u draws new topic $\psi_{u,new}$ with probability $\frac{\gamma}{n_{u:}(t)+\gamma}$.
   5. Draw $\psi_{u,new}$ with probability $\frac{m_k(t)}{n_{u:}(t)+\zeta}$ or new topic with probability $\frac{\zeta}{m_.(t)+\zeta}$ where $\mathrm{m}_k(t) = \sum_{e \epsilon D_u^0(t)} \exp\left(-\nu_k(t-t_e)\right) I[\theta_e = \phi, l_e = 1]$
   6. $\mathrm{w}_{e_i}^v \sim Multi(\phi_k)$

Experiments: Synthetic Data

- Number of users=1000, Number of events = 10^4, Vocabulary size=20.
- event = (time, dimension, document content)
- Mean Absolute Error of matrix α (Influence matrix of users).
- Mean Absolute error(MAE) of base intensity μ_u of all users.

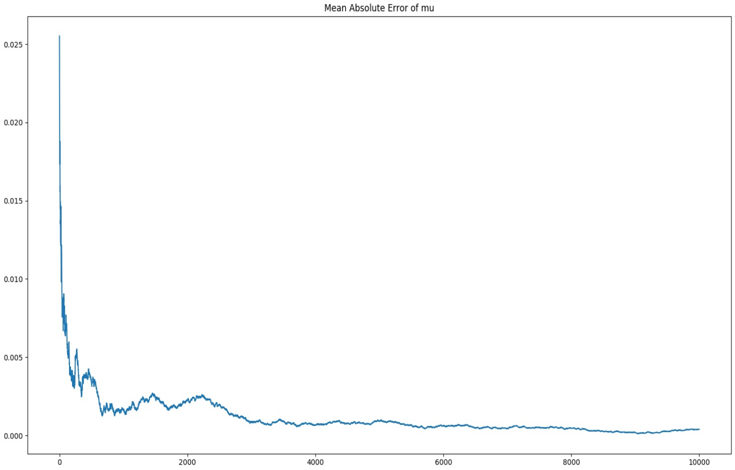
Experiments: Synthetic Data



x-axis: number of events

y-axis: error

Experiments: Synthetic Data



x-axis: number of events

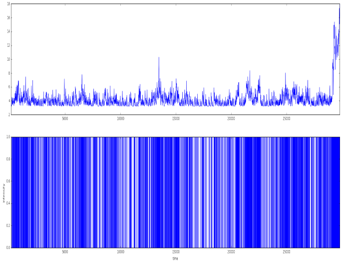
y-axis: error

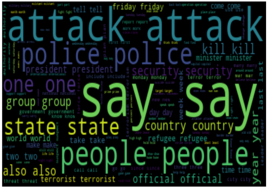
Experiments: Real Dataset

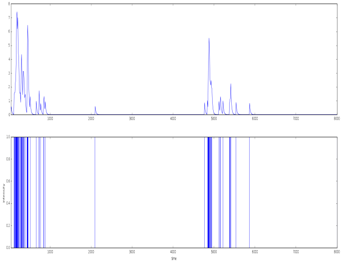
- EventRegistry[1] Dataset: number of events=50000, number of users=100.
- Articles consist of mainly 3 different tags; FIFA, Iran-Sanctions, and Paris-Attack from 2015/11/01 to 2016/01/13.
- Collected data consist of 100 news sites(users).

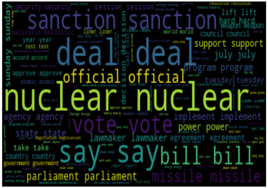
[1] http://eventregistry.org/

Experiments: Real Dataset

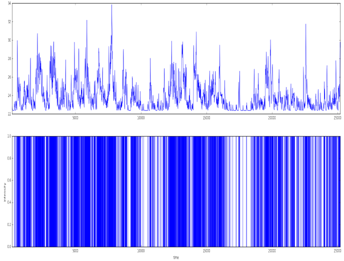


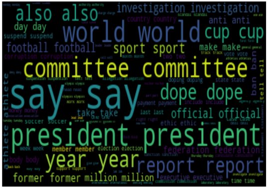






Experiments: Real Dataset





Conclusion and Future Work

- Joint modeling of time and mark of events helps in better understanding the social network.
- Easy adaption of temporal and topical complexity according to the complexity of data.
- Incorporation of Neural Network architectures specifically RNN and LSTM.

References

[1]N .Du, M. Frajtbar, A. Ahmed, A. J. Smola and L. Song. Dirichlet Hakwes processes with applications to clustering continuous document streams KDD, 2015.

[2]D. M. Blei and J. D. Lafferty. Dynamic topic models ICML, 2006.

[3]D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation, Journal of Machine Learning Research, January 2003.

[4]X. He, T. Rekatsinas, J. Foulds, L. Getoor, and Y. Liu. Hawkestopic: A joint model for network inference and topic modeling from text-based cascades ICML, 2015.

contd..

[5]S. Liang, E. Yilmaz, E. Kanoulas. Dynamci Clustering of Streaming Short Documents KDD, 2016.

[6]L.M. Aiello, Georgios, Sensing Trending Topics in Twitter, IEEE 2013

[7]Alan G. Hawkes. Spectra of some self-exciting and mutually exciting point processes.Biometrika, 58(I):8390, 1971.

[8]Mohler, G.O., Short, M.B., Brantingham, P.J., Schoenberg, F.P., and Tita, G. E. Self-exciting point process modeling of crime. Journal of the American Statistical Association, 106(493):100108, 2011.

[9]Hosseini, S.A., Khodadadi, A., Arabzade, S., Rabiee, H.R.: HNP3: A hierarchical non-parametric point process for modeling content diffusion over social media.

contd..

[10] Bacrya, E., Delattreb, S., Hoffmannc, M. and Muzyd,J.F. Modeling microstructure noise with mutually exciting point processes. Quantitative Finance,2012.

[11] Ogata Y. On Lewis' simulation method for point processes. IEEE Transactions on Information Theory, 27(1):2331, 1981.

[12] Liniger, T. Multivariate hawkes processes. ETH Doctoral Dissertation No. 18403, 2009.

[13]T.Iwata, A. Shah, Z. Ghahramani Discovering Latent Influence in Online Social Activites via Shared Cascade Poiosson Processes

[14]https://sites.google.com/site/quantlullaby/self-exciting-point-processes

[15] Nan Du, Le Song, Alexander J. Smola, and Ming Yuan. Learning networks of heterogeneous influence.In NIPS, 2012.

ThankYou!!!