Drifting and Popularity: A Study of Time Series Analysis of Topics

Muhammad Haseeb UR Rehman Khan(Presenter),

Kei Wakabayashi

s2036048@s.tsukuba.ac.jp

University of Tsukuba, Japan

IARIA

筑波大学
University of Tsukuba

About Me

Research in the field of NLP, Machine Learning especially in Topic Modeling

- PhD in Informatics, University of Tsukuba, 2020-2023
- MS in Library and Information Studies, University of Tsukuba, 2018-2020

Topic Modeling

Unsupervised way of summarizing and categorizing document

Time series features of topics

Topic Drifting

Transitions of topics over time

Matrix
Vector
Signal
Spectral

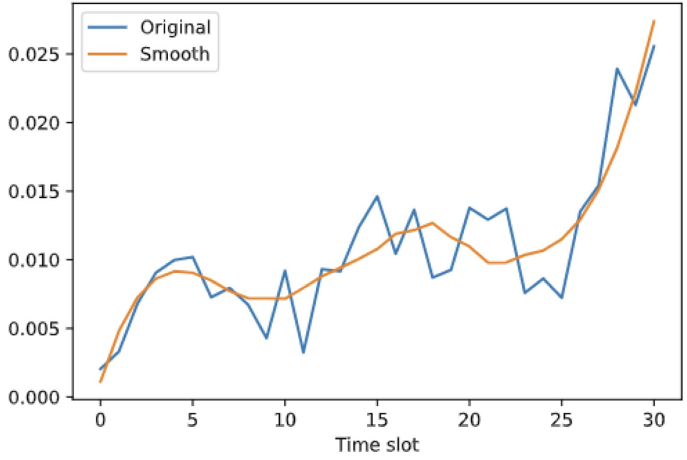
Matrix
Component
Signal
Source

Matrix
Signal
Source
Spectral

Signal
Source
Spectral
Processing

Topic Popularity

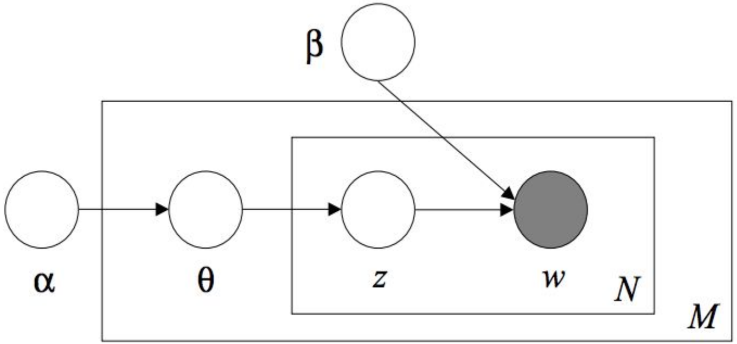
Topic proportion estimation w.r.t time



Topic Models

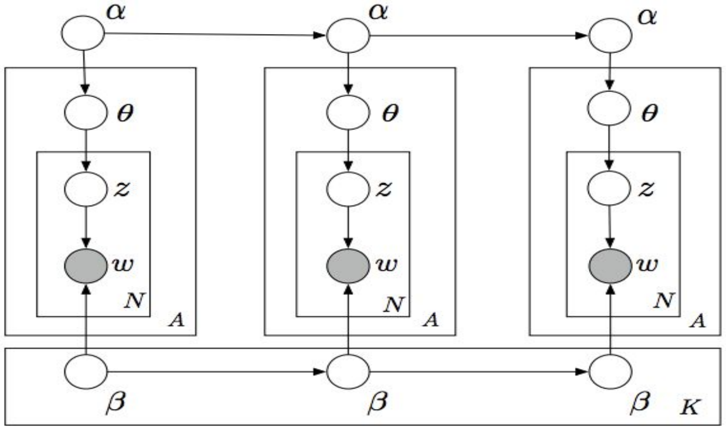
LDA (37,000 citations)
Latent Dirichlet Allocation[1]

Draws a static set of topic



DTM (2,700 citations)
Dynamic Topic Models[2]

Assumes dynamic drift of distributions



Background (Previous Research)

- Koike et al.[3] proposed a method that draws a time series graph to find the bursty topic detection
- Used a subset of **Twitter data** (keywords based queries) and news
- Used **DTM** with ***K = 50***

Background (My Work)

- Khan et al.[4] performed similar type of experiment and made time series graphs of topics for multiple purposes
- Used full TREC **Twitter data** Tweets2011
- Used **LDA** with ***K = 1000***
- We believe that there are many diverse topics in Twitter

Research Question

If time series topic information can be extracted by using LDA then why do we need to use DTM?

Problem Statement:

Can time series topic information be extracted from LDA without using DTM?

Draw empirical Insights by using multiple datasets and configurations

**topic information means topic drifting and topic popularity*

LDA Training and Inference

LDA Model Training

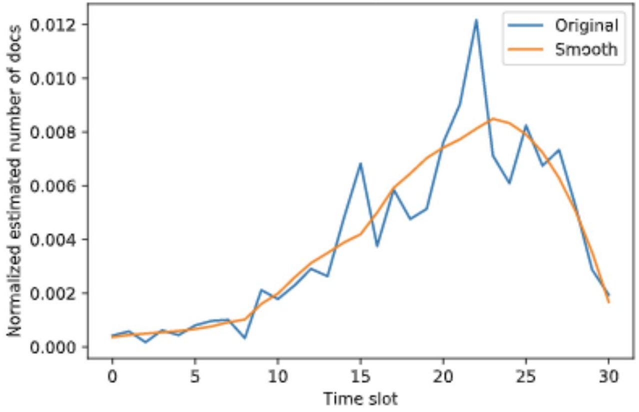
- LDA assumes a $\boldsymbol{z_i}$ for each word $\boldsymbol{w_i}$ in a document and draws a $\boldsymbol{\theta_d}$ for $\boldsymbol{d}$.
- The $\boldsymbol{w_i \in W}$ is drawn from a distribution of words associated to the assigned topic $\boldsymbol{z_i=k}$ which is denoted by $\boldsymbol{\phi_k}$
- We use $\boldsymbol{X = \{x_1,\dots\dots x_D\}}$ as training dataset for LDA
- If we apply a pooling method (for short text dataset), $\boldsymbol{x_d}$ consists of multiple text instances, and number of instances is $\boldsymbol{T_d}$

Time Series Topic Estimation by LDA(Inference)

Let's call it TSLDA

- Create inference dataset by associating $\boldsymbol{x_d}$ with time slice $\boldsymbol{t}$
- If no pooling then $\boldsymbol{T_d = 1}$
- Estimate the topic distribution $\boldsymbol{\theta_d}$ for each document
- Calculate the estimated number of documents for each topic $\boldsymbol{k}$ at each time slice $\boldsymbol{t}$ using:

$$N_k^t = \sum_{d:\mathbf{x}_d \in X_t} \theta_{dk} T_d$$



- $\boldsymbol{\theta_d}$ is computed using Dirichlet distribution

$$p(\theta_d|\mathbf{x}_d) = \sum_{\mathbf{z}} p(\theta_d|\mathbf{z})p(\mathbf{z}|\mathbf{x}_d)$$

Similarity Analysis of Both Topic Models

LDA and DTM Topics Similarity Analysis
By Jensen-Shannon Divergence

- To check the relationship between both set of topics, we used JS divergence similarity measure

$$JSD(\tilde{\phi}_k||\tilde{\phi}_j^t) = \frac{1}{2}D_{KL}(\tilde{\phi}_k||T_M) + \frac{1}{2}D_{KL}(\tilde{\phi}_j^t||T_M)$$

- Where $T_M = \frac{1}{2}(\tilde{\phi}_k + \tilde{\phi}_j^t)$

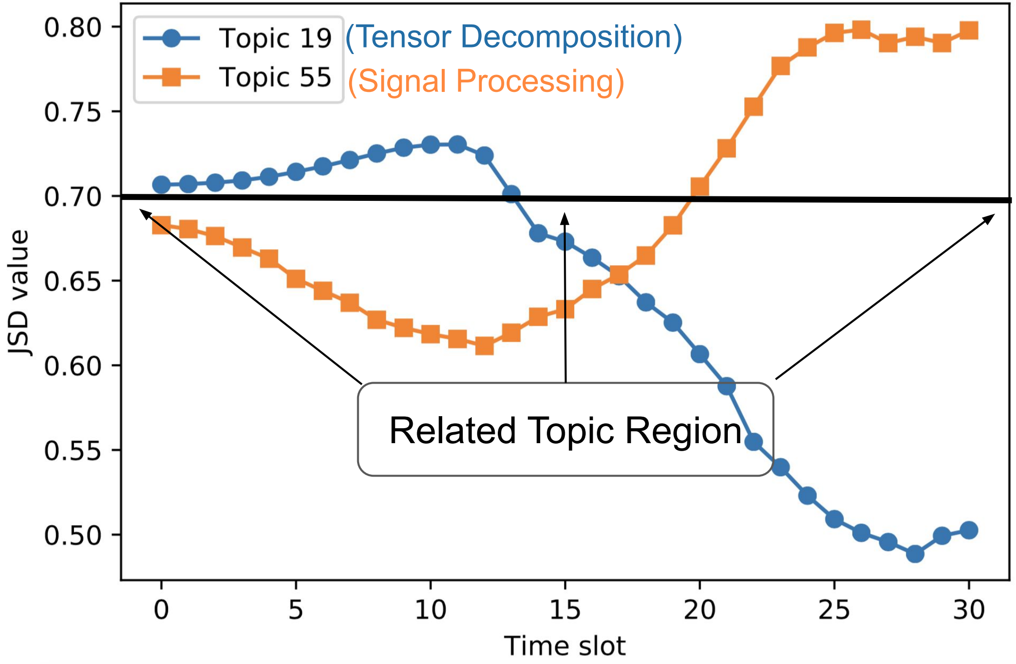
$\tilde{\phi}_k$ is ***k***th LDA topic

$\tilde{\phi}_j^t$ is ***j***th DTM topic at ***t***

- $D_{KL}(\phi_1||\phi_2)$ is the Kullback-Leibler divergence

$$D_{KL}(\phi_1||\phi_2) = \sum_{w \in W} P(w|\phi_1) \log \frac{P(w|\phi_1)}{P(w|\phi_2)}$$

LDA and DTM Topics Similarity Analysis
using Jensen-Shannon Divergence



DTM topic 0 with LDA topic 19 and 55

$\mathrm{JSD}(\phi_{(\mathrm{LDA}\ 55)} \| \phi^{(0-20)}_{(\mathrm{DTM}\ 0)}) \leq 0.7$ (RT)

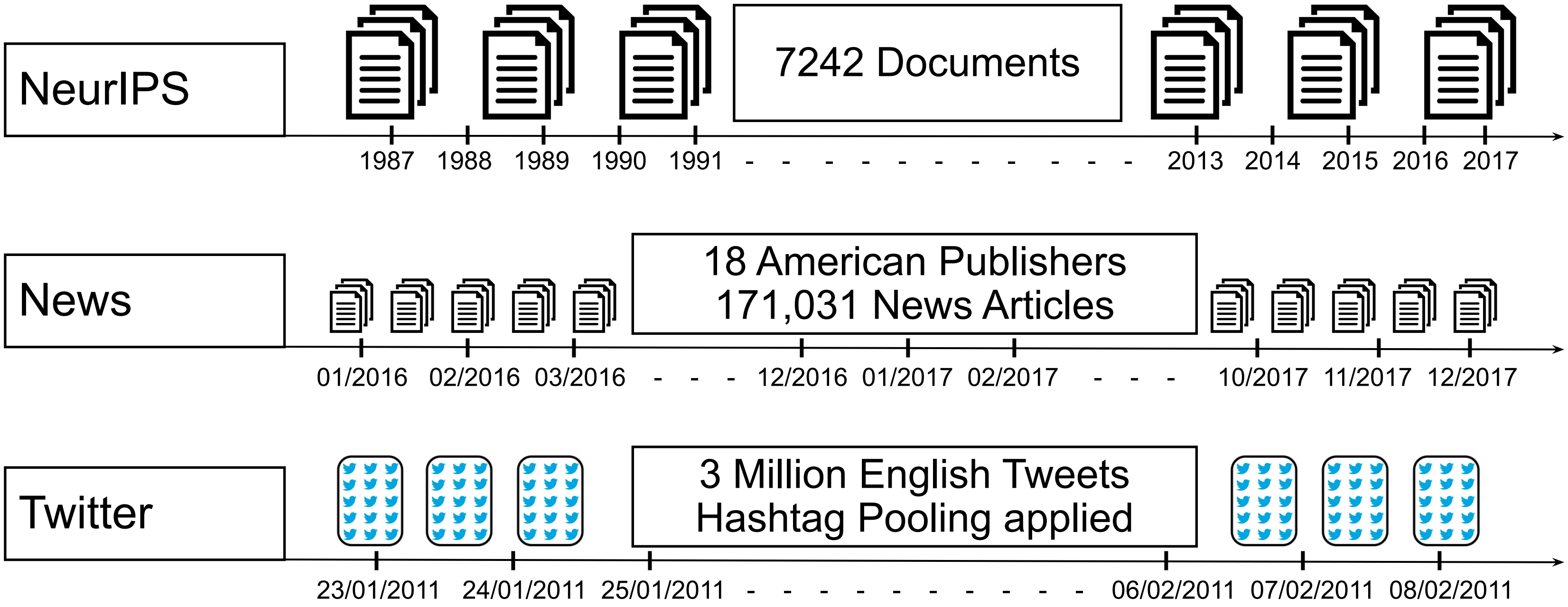
$\mathrm{JSD}(\phi_{(\mathrm{LDA}\ 19)} \| \phi^{(15\text{-}30)}_{(\mathrm{DTM}\ 0)}) \leq 0.7$ (RT)

So,

LDA 19 & 55 are
Fragmented Topics
of
DTM 0

Experiment

Datasets and ***K***



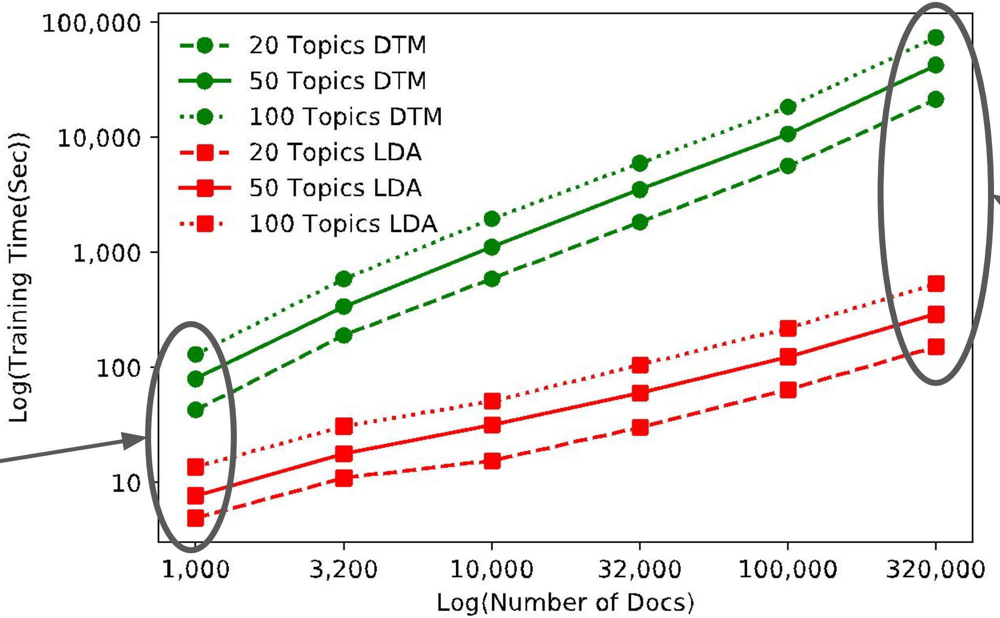
Number of Topics(***K***): 30, 60, 90(Analysis)

Results

- Training time
- Topic drifting
- JS analysis (Fragmentation)
- Time dependent topic popularity

Training Time Cost

DTM and LDA trained on Twitter dataset



DTM 10X in training

DTM 100X in training

Topic Drifting

DTM trained on **dataset** and ***K*** one at a time
V_s is all unique words in ***t*** (**Higher V_s → More drifting**)

Dataset	Topics	$K(V_s > 70)$	$K(V_s > 90)$	$K(V_s > 120)$
NeurIPS	30	13	8	0
	60	58	56	11
	90	90	90	90
Twitter	30	3	1	0
	60	3	0	0
	90	1	0	0
News	30	29	20	5
	60	57	33	1
	90	83	37	2

Strong Drifting

No Drifting

Weak Drifting

Fragmentation by JS Analysis

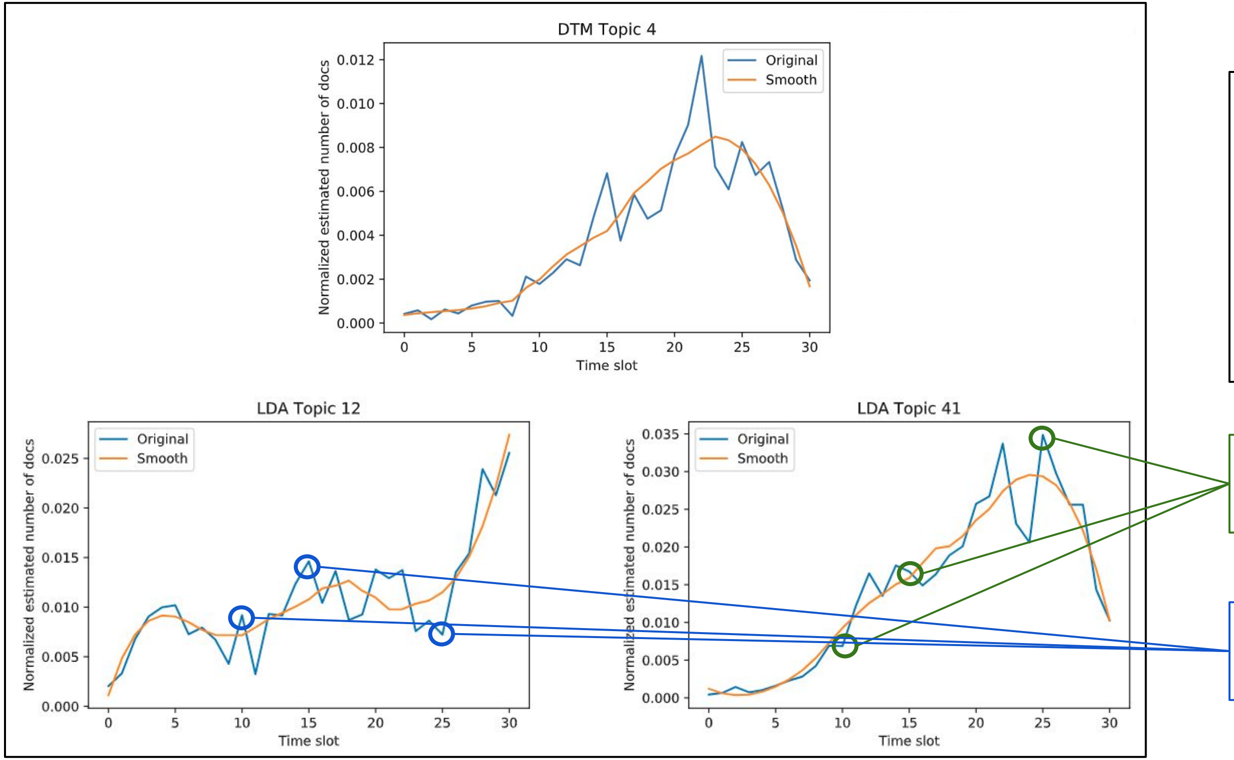
FT values are number of DTM topics being fragmented into two(**F2**), three(**F3**) or more(**F4&more**) LDA topics.

Dataset	Topics	RT	FT	F 2	F 3	F 4 & more
NeurIPS	30	17	4	3	1	0
	60	42	16	11	5	0
	90	69	28	25	2	1
Twitter	30	5	0	0	0	0
	60	11	1	1	0	0
	90	14	3	2	1	0
News	30	8	1	1	0	0
	60	24	2	2	0	0
	90	42	4	4	0	0

High Fragmentation
Follow prior docs

No Fragmentation
Instantaneous responses of events
Die quickly

Time Dependent Topic Popularity



Using TSLDA,
Computed $\boldsymbol{\theta}_{dk}$
Calculated N_k^t
and built graphs

$N_{41}^{10} \parallel N_{41}^{15} \parallel N_{41}^{25}$

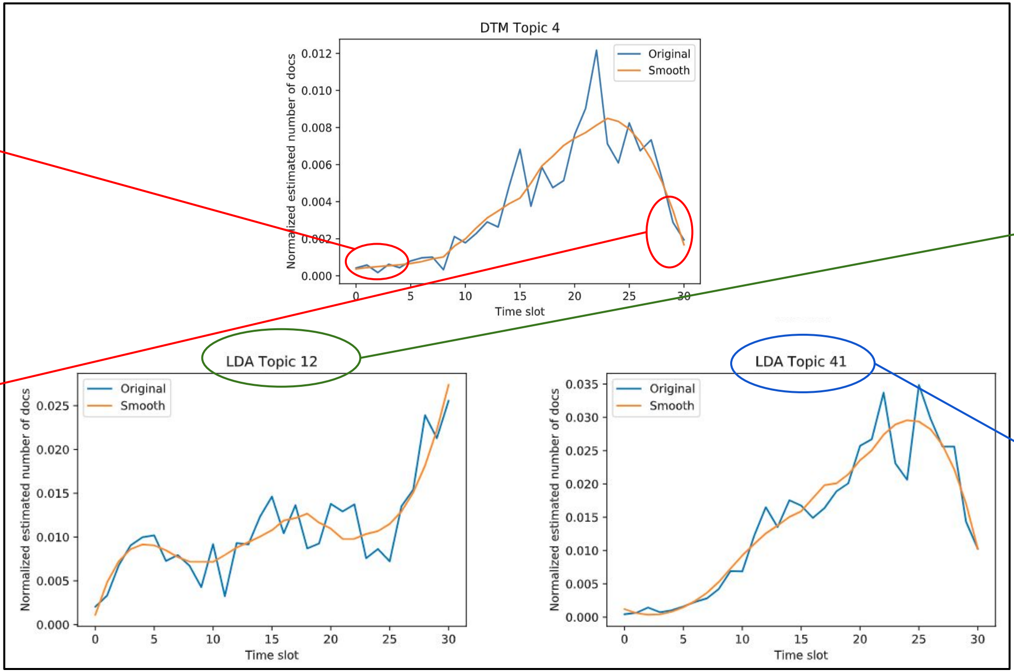
$N_{12}^{10} \parallel N_{12}^{15} \parallel N_{12}^{25}$

Topic Transition Information Embedded with DTM
(Not good for topic popularity analysis)

Information retrieval from documents

Drifted

Document analysis with **LDA**



Information retrieval

Variational topic model **LDA**

Key Points

Topic Drifting

- **No topic drifting in *Twitter*** so only topic popularity for such datasets
- **For *NeurIPS*,** TF increased with an increase of *K,* means **high topic drifting for higher K**
- **For News**, because V_s is low, it is **weak topic drifting**

Topic Popularity

- We can **construct** topic popularity graphs for **both models**
- DTM topics have **topic transition information embedded with topics** so it's **little vague** from *DTM*

Conclusion

- **Training time** for DTM was **100X more** as compared to LDA for large datasets

*Can time series ****topic information*** *be extracted from LDA without DTM?(RQ)*

- **Topic drifting** **is a unique property of DTM**, but some datasets like ***Twitter*** **do not have topic transition information**, so applying DTM to such datasets is **waste of resources**

- Time series **topic popularity** **can be extracted from both models**, but topic popularity extracted **using LDA is precise** as compared to DTM because **DTM has topic transition embedded in the topics**

**topic information means topic drifting and topic popularity*

Suggestions Based on Findings

Use LDA for	**Use DTM for**
Twitter with high K	NIPS data with few number of topics
News with high K	News with smaller K
Short duration datasets e.g. Twitter	Long duration docs e.g. NIPS
Extract topic popularity	Extract topic drifting

Thank You!

Any Questions

s2036048@s.tsukuba.ac.jp