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
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

Topic Popularity
Topic proportion estimation w.r.t time
**Topic Models**

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

Draws a static set of topics

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

Assumes dynamic drift of distributions
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 **$K = 50$**
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
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 $z_i$ for each word $w_i$ in a document and draws a $\theta_d$ for $d$.
- The $w_i \in W$ is drawn from a distribution of words associated to the assigned topic $z_i=k$ which is denoted by $\phi_k$.
- We use $X = \{x_1, \ldots, x_D\}$ as training dataset for LDA.
- If we apply a pooling method (for short text dataset), $x_d$ consists of multiple text instances, and number of instances is $T_d$.
Time Series Topic Estimation by LDA (Inference)
Let’s call it TSLDA

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

$$N_{tk}^t = \sum_{d: x_d \in X_t} \theta_{dk} T_d$$

- $\theta_d$ is computed using Dirichlet distribution

$$p(\theta_d | x_d) = \sum_{z} p(\theta_d | z) p(z | x_d)$$
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

\[
JSD(\tilde{\phi}_k||\tilde{\phi}_j) = \frac{1}{2}D_{KL}(\tilde{\phi}_k||T_M) + \frac{1}{2}D_{KL}(\tilde{\phi}_j||T_M)
\]

Where

\[
T_M = \frac{1}{2}(\tilde{\phi}_k + \tilde{\phi}_j)
\]

\(\tilde{\phi}_k\) is \(k\)th LDA topic

\(\tilde{\phi}_j\) 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

\[ \text{JSD}(\phi_{(\text{LDA 55})} \| \phi_{(\text{DTM 0})}) \leq 0.7 \quad \text{(RT)} \]
\[ \text{JSD}(\phi_{(\text{LDA 19})} \| \phi_{(\text{DTM 0})}) \leq 0.7 \quad \text{(RT)} \]

So,

LDA 19 & 55 are Fragmented Topics of DTM 0
Experiment
Datasets and $K$

NeurIPS

- 1987
- 1988
- 1989
- 1990
- 1991
- 1992
- 1993
- 1994
- 1995
- 1996
- 1997
- 1998
- 1999
- 2000
- 2001
- 2002
- 2003
- 2004
- 2005
- 2006
- 2007
- 2008
- 2009
- 2010
- 2011
- 2012
- 2013
- 2014
- 2015
- 2016
- 2017

News

- 01/2016
- 02/2016
- 03/2016
- 12/2016
- 01/2017
- 02/2017
- 10/2017
- 11/2017
- 12/2017

Twitter

- 23/01/2011
- 24/01/2011
- 25/01/2011
- 06/02/2011
- 07/02/2011
- 08/02/2011

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 $V_s$ $\rightarrow$ More drifting*)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Topics</th>
<th>$K(V_s &gt; 70)$</th>
<th>$K(V_s &gt; 90)$</th>
<th>$K(V_s &gt; 120)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NeurIPS</td>
<td>30</td>
<td>13</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>58</td>
<td>56</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>Twitter</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>News</td>
<td>30</td>
<td>29</td>
<td>20</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>57</td>
<td>33</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>83</td>
<td>37</td>
<td>2</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Topics</th>
<th>RT</th>
<th>FT</th>
<th>F 2</th>
<th>F 3</th>
<th>F 4 &amp; more</th>
</tr>
</thead>
<tbody>
<tr>
<td>NeurIPS</td>
<td>30</td>
<td>17</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>42</td>
<td>16</td>
<td>11</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>69</td>
<td>28</td>
<td>25</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Twitter</td>
<td>30</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>14</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>News</td>
<td>30</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>42</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**High Fragmentation**
Follow prior docs

**No Fragmentation**
Instantaneous responses of events
Die quickly
Time Dependent Topic Popularity

Using TSLDA, Computed $\theta_{dk}$
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

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*
• **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

<table>
<thead>
<tr>
<th>Use LDA for</th>
<th>Use DTM for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twitter with high K</td>
<td>NIPS data with few number of topics</td>
</tr>
<tr>
<td>News with high K</td>
<td>News with smaller K</td>
</tr>
<tr>
<td>Short duration datasets e.g. Twitter</td>
<td>Long duration docs e.g. NIPS</td>
</tr>
<tr>
<td>Extract topic popularity</td>
<td>Extract topic drifting</td>
</tr>
</tbody>
</table>
Thank You!

Any Questions

s2036048@s.tsukuba.ac.jp