Leveraging Social Context for Modeling Topic Evolution

Janani Kalyanam, Amin Mantrach, Diego Saez-Trumper, Hossein Vahabi, Gert Lanckriet

UC San Diego

Yahoo Labs
Introduction
Introduction
Introduction
Topic Modeling

- NMF-based
- Bayesian (like LDA)
Bird flu outbreak; everything you need to know goo.gl/F1dnfk #birdflu

U.S to review protocols following birdflu outbreak goo.gl/X88iSe #birdflu

U.S poultry devastated by birdflu outbreak goo.gl/1gX8FC #birdflu
Bird flu outbreak; everything you need to know goo.gl/F1dnfk #birdflu

U.S to review protocols following birdflu outbreak goo.gl/X88iSe #birdflu

U.S poultry devastated by birdflu outbreak goo.gl/1gX8FC #birdflu
Bird flu outbreak; everything you need to know goo.gl/F1dnfk #birdflu

U.S to review protocols following birdflu outbreak goo.gl/X88iSe #birdflu

U.S poultry devastated by birdflu outbreak goo.gl/1gX8FC #birdflu

Kim Kardashian: pregnant again! goo.gl/Ir1knd #celebritygossip

Selina Gomez and Justin Bieber: “just friends” goo.gl/M9dlhj #celebritygossip

Lindsay Lohan messed up contract with Oprah goo.gl/Ir1knd #celebritygossip
Topic Modeling

- NMF-based
- Bayesian (like LDA)

Generally focus on content
What’s needed

in addition to textual content, use context and meta data that surrounds the text to discover the latent topics
Our goal

Does user interactions, and temporal evolution help detect better topics?
How do we do this?
How do we approach this?

- Non Negative Matrix Factorization based method.
- Start with the classical NMF objective..
- build on it..
## Notation

The notation includes:

- $X^t$: A matrix with the number of documents as rows and the number of words as columns.
- $U^t$: A matrix with the number of documents as rows and the number of users as columns.
Notation

$X^t$

$U^t$

#-of-documents

#-of-words

#-of-users
How do we approach this?

\[ X^t \approx W^t H^t \]

"Hilary Clinton challenged Joe Biden"

(10%, 90%)
Ingredients of Objective Function

$$\|X^t - W^t H^t\|^2$$

Variables are $W^t$ $H^t$
How do we approach this?

\[ U^t \approx W^t G^t \]
Ingredients of Objective Function

\[ \|X^t - WH^t\|^2 + \|U^t - WG^t\|^2 \]

Variables are \( W^t \ H^t \ G^t \)
Key Assumption

\[ X^t \approx W^t H^t \quad U^t \approx W^t G^t \]

The \( W^t \) matrix is common to both decompositions.
Key Assumption

\[ X^t \approx W^t H^t \]

\[ U^t \approx W^t G^t \]
Evolution Over Time

\[ X^t \approx W^t M^t H^{t-1} \]
Evolution Over Time

\[ X^t \approx W^t M_T^t H^{t-1} \]

\[ M_T^t \quad \text{Evolution matrix} \]
Ingredients of Objective Function

\[ \|X^t - WH^t\|^2 + \|X^t - WM^t H^{-1}\|^2 + \|U^t - W^t G^t\|^2 + \|U^t - W^t M^t C^t G^{-1}\|^2 \]

\( L_T \)
content part

\( L_C \)
community part

Variables are \( W^t, H^t, G^t, M^t_T, M^t_C \)
Loss Function

\[ L = \mu L_T + (1 - \mu) L_C + R \]

\( \mu \) importance parameter

\( R \) regularization
How to evaluate?
How do we do this?

Split into three categories..
How do we do this?

Split into three categories..

- “good topics”, CONTENT STABLE TOPICS
How do we do this?

Split into three categories..

- “good topics”, *CONTENT STABLE TOPICS*

- “difficult topics”, *COMMUNITY STABLE TOPICS*
How do we do this?

Split into three categories..

- “good topics”, CONTENT STABLE TOPICS
- “difficult topics”, COMMUNITY STABLE TOPICS
- a mixture of the above two or MIXED STABLE TOPICS
How do we do this?

In each category, evaluate how much does adding the contextual information and temporal information really help..
Data

• Content
  – News articles from CNN, BBC, Al jazeera

• Community
  – All tweets which linked to the articles
    • Collect *username* publishing the tweet
    • Collect the *hashtag* in the tweet
Baseline Approaches

• LTECS: Learning Topic Evolution from Content and Social Media activity

• Link-PLSA-LDA (Nallapati et. al. KDD 2008): lacks temporal element
Baseline Approaches

- Online LDA (AlSumait et. al. ICDM 2008): lacks community element

- Joint Past Present Decomposition (Vaca Ruiz et. Al. WWW 2014): lacks of community

- CMF (Recsys 2014): lacks of temporal element
## Results (Community Stable)

### LTECS

<table>
<thead>
<tr>
<th></th>
<th>$K = 5$</th>
<th>$K = 10$</th>
<th>$K = 15$</th>
<th>$K = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NDCG</strong></td>
<td>0.4081</td>
<td>0.4800</td>
<td>0.5029</td>
<td>0.5129</td>
</tr>
<tr>
<td><strong>MAP</strong></td>
<td>0.2653</td>
<td>0.3637</td>
<td>0.4007</td>
<td>0.4173</td>
</tr>
<tr>
<td>$\boldsymbol{\mu}$</td>
<td>$0.01$</td>
<td>$0.5$</td>
<td>$0.5$</td>
<td>$0.5$</td>
</tr>
</tbody>
</table>

### Baseline Approach; NO CONTEXT

<table>
<thead>
<tr>
<th></th>
<th>$K = 5$</th>
<th>$K = 10$</th>
<th>$K = 15$</th>
<th>$K = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NDCG</strong></td>
<td>0.3699</td>
<td>0.4496</td>
<td>0.4608</td>
<td>0.4138</td>
</tr>
<tr>
<td><strong>MAP</strong></td>
<td>0.2191</td>
<td>0.3596</td>
<td>0.3462</td>
<td>0.3420</td>
</tr>
</tbody>
</table>

### Baseline Approach; NO TEMPORAL MODELING

<table>
<thead>
<tr>
<th></th>
<th>$K = 5$</th>
<th>$K = 10$</th>
<th>$K = 15$</th>
<th>$K = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NDCG</strong></td>
<td>0.3454</td>
<td>0.4338</td>
<td>0.4771</td>
<td>0.4827</td>
</tr>
<tr>
<td><strong>MAP</strong></td>
<td>0.2044</td>
<td>0.3190</td>
<td>0.3757</td>
<td>0.3665</td>
</tr>
</tbody>
</table>
# Results (Content Stable)

## LTECS

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.6888</td>
<td>0.6055</td>
<td>0.6317</td>
<td>0.6623</td>
</tr>
<tr>
<td>MAP</td>
<td>0.5655</td>
<td>0.4784</td>
<td>0.5115</td>
<td>0.5559</td>
</tr>
<tr>
<td>$\mu$</td>
<td>$\mu = 1$</td>
<td>$\mu = 1$</td>
<td>$\mu = 0.75$</td>
<td>$\mu = 0.75$</td>
</tr>
</tbody>
</table>

## Baseline Approach; NO CONTEXT

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.6888</td>
<td>0.6055</td>
<td>0.4885</td>
<td>0.6504</td>
</tr>
<tr>
<td>MAP</td>
<td>0.5655</td>
<td>0.4784</td>
<td>0.3089</td>
<td>0.5411</td>
</tr>
</tbody>
</table>

## Baseline Approach; NO TEMPORAL MODELING

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.5846</td>
<td>0.4919</td>
<td>0.4455</td>
<td>0.4327</td>
</tr>
<tr>
<td>MAP</td>
<td>0.4423</td>
<td>0.3207</td>
<td>0.2556</td>
<td>0.2557</td>
</tr>
</tbody>
</table>
# Results (Mixed Stable)

## LTECS

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.9005</td>
<td>0.8868</td>
<td>0.9249</td>
<td>0.9089</td>
</tr>
<tr>
<td>MAP</td>
<td>0.7783</td>
<td>0.7965</td>
<td>0.8964</td>
<td>0.8845</td>
</tr>
<tr>
<td>(\mu = 0.25)</td>
<td>(\mu = 0.75)</td>
<td>(\mu = 0.25)</td>
<td>(\mu = 0.25)</td>
<td></td>
</tr>
</tbody>
</table>

## Baseline Approach; NO CONTEXT

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.8771</td>
<td>0.8762</td>
<td>0.4251</td>
<td>0.4580</td>
</tr>
<tr>
<td>MAP</td>
<td>0.7762</td>
<td>0.7783</td>
<td>0.3232</td>
<td>0.3644</td>
</tr>
</tbody>
</table>

## Baseline Approach; NO TEMPORAL MODELING

<table>
<thead>
<tr>
<th></th>
<th>K = 5</th>
<th>K = 10</th>
<th>K = 15</th>
<th>K = 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>NDCG</td>
<td>0.6712</td>
<td>0.8768</td>
<td>0.8905</td>
<td>0.8765</td>
</tr>
<tr>
<td>MAP</td>
<td>0.5329</td>
<td>0.8223</td>
<td>0.8499</td>
<td>0.8337</td>
</tr>
</tbody>
</table>
Conclusion

- Using community side information helps with “noisy” topics.
Thank You!