TopicFlow Model:  
Modeling topic-specific global influence of hyperlinked documents

Ramesh Nallapati, Dan McFarland and Christopher Manning  
Stanford University  
AISTATS, Fort Lauderdale  
04/12/2011
Introduction:
Modeling global influence

Typical Link based influence models ignore content!
Introduction: Modeling topical global influence!

What we want is topic specific global influence!

- Sports (ESPN)
- Politics (CNN)
- Movies (IMDb)
- Books (Amazon.com)
TopicFlow model

• A new topic model for global topic-specific influence
  – combines network flow models with topic modeling
TopicFlow model:
Independent topic sources
TopicFlow model:
Single source for all topics
TopicFlow model: definitions

- At each document, net incoming flow on each topic is equal to the net outgoing topical flow:
  \[ \sum_{i \in \text{Pa}(d)} f_{i,d}^{(k)} = \sum_{j \in \text{Ch}(d)} f_{d,j}^{(k)} \]

- Normalized incoming topic flows define the topic proportions for word generation:
  \[ \theta_d^{(k)} = \frac{f_{i,d}^{(k)}}{\sum_{K_{k'}=1}^{K} f_{i,d}^{(k')}} \]

- Topic specific influence of a document:
  \[ I(d, k) = (f_{i,d}^{(k)} - f_{s,d}^{(k)}) \theta_d^{(k)} \]
TopicFlow model: What the flow means

• Flow from document A to B is the amount of A’s topical `influence’ that is attributed to B
  – All citations of A share the incoming flow to A among themselves
  – Document A assigns flow to document B proportional to the similarity of their topical content

• Flow from source models missing incoming citations

• Flow into sink models missing outgoing citations
TopicFlow model: Intuition

- More the number of documents that assign topical flow to a document $d \Rightarrow$
  - more is the probability that $d$ discusses the topic
  - higher is the document’s topic-specific influence

- “Wisdom of the crowds” model
  - In reality, text is produced first and incoming citations appear next
  - Works well in practice because the model has the “luxury of hindsight”!
TopicFlow model: How does it model global topical influence?

- Flow balance constraints help spread the influence globally
  - Dynamic interplay between words and flow
### Topic Sensitive PageRank vs. TopicFlow

<table>
<thead>
<tr>
<th>Property</th>
<th>Topic Sensitive PageRank</th>
<th>TopicFlow model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models global topical influence?</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Models broken links in influence?</td>
<td>Yes, using teleportation</td>
<td>Yes, using source/sink formalism</td>
</tr>
<tr>
<td>Requires Labeled Data?</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Learns the edge weights?</td>
<td>No, but allows the user to input arbitrary weights</td>
<td>Learns automatically based on textual similarity</td>
</tr>
<tr>
<td>Models simultaneously dependency between words and links?</td>
<td>No. Assumes topic labels to be given and works only with links</td>
<td>Yes</td>
</tr>
<tr>
<td>Can track flow of topics across the network?</td>
<td>Not explicitly. Only outputs the final PageRank vectors</td>
<td>Yes. The model learns flow values along each edge in the network.</td>
</tr>
</tbody>
</table>
TopicFlow model: Learning and Inference

- **Objective function**

\[
\max_{f, \beta} \sum_{d=1}^{M} \sum_{n=1}^{N_d} \log \left( \sum_{k=1}^{K} \beta_{kw} \theta_d^{(k)}(n) \right) - \frac{1}{2} \lambda \left( \sum_{k} \|f_{s,k}\|^2 + \sum_{d} \sum_{k} \|f_{d,k}\|^2 \right)
\]

- **Flow balance constraints**

\[
\forall d \in V' - \{s\} \quad \sum_{i \in Pa(d)} f_{i,d}^{(k)} = \sum_{j \in Ch(d)} f_{d,j}^{(k)} \quad \text{and}
\]

\[
\sum_{k=1}^{K} \sum_{d=1}^{M} f_{s,d}^{(k)} = 1 \quad \text{If single source}
\]

\[
\sum_{d=1}^{M} f_{s,d}^{(k)} = 1 \quad \text{If multiple topic sources}
\]

- **Observed data log likelihood**
- **Regularization**
TopicFlow model:
Learning and Inference

• Step 1: variational lower bound for log likelihood:

\[
\log P(w) = \sum_{d=1}^{M} \sum_{n=1}^{N_d} \log \left( \sum_{k=1}^{K} \beta_{kw_n} \theta_d^{(k)} \right)
\]

\[
\geq \sum_{d=1}^{M} \sum_{n=1}^{N_d} \sum_{k=1}^{K} \phi_{dnk} \left( \log \beta_{kw_n} + \log \theta_d^{(k)} - \log \phi_{dnk} \right)
\]
TopicFlow model: Learning and Inference

- Step 2: Elimination of equality constraints:
  - Converting flow balance constraints into multinomial constraints:

\[
\forall j \in \text{ch}(d) \quad f_{d,j}^{(k)} = f_{.,d}^{(k)} \psi_{d,j}^{(k)} \quad \text{s.t.} \quad \sum_{j \in \text{ch}(d)} \psi_{d,j}^{(k)} = 1 \quad \text{and} \quad \forall j \psi_{d,j} \geq 0
\]

- Eliminating multinomial constraints using a logistic transformation:

\[
\psi_{s,d}^{(k)} = \frac{\exp(\eta_{s,d}^{(k)})}{\sum_{k'} \sum_{d' \in \text{ch}(s)} \exp(\eta_{s,d'}^{(k')})}
\]

\[
\psi_{d,v}^{(k)} = \begin{cases} 
(1 - \frac{1}{|\text{ch}(d)|}) \frac{\exp(\eta_{d,v}^{(k)})}{\sum_{v' \in \text{ch}(d) - s} \exp(\eta_{d,v'}^{(k)})} & \text{if } v \neq t \\
\frac{1}{|\text{ch}(d)|} & \text{if } v = t
\end{cases}
\]
<table>
<thead>
<tr>
<th>“Machine Translation”</th>
<th>“Syntactic Parsing”</th>
<th>“Supervised learning”</th>
</tr>
</thead>
<tbody>
<tr>
<td>translat</td>
<td>pars</td>
<td>discours</td>
</tr>
<tr>
<td>align</td>
<td>tree</td>
<td>relat</td>
</tr>
<tr>
<td>word</td>
<td>grammar</td>
<td>structur</td>
</tr>
<tr>
<td>model</td>
<td>parser</td>
<td>text</td>
</tr>
<tr>
<td>sentence</td>
<td>node</td>
<td>segment</td>
</tr>
<tr>
<td>english</td>
<td>sentenc</td>
<td>unit</td>
</tr>
<tr>
<td>pair</td>
<td>depend</td>
<td>rhetor</td>
</tr>
<tr>
<td>sourc</td>
<td>rule</td>
<td>marker</td>
</tr>
<tr>
<td>languag</td>
<td>input</td>
<td>linguist</td>
</tr>
<tr>
<td>target</td>
<td>deriv</td>
<td>cue</td>
</tr>
</tbody>
</table>


Aligning sentences in bilingual corpora using lexical information, Chen, 1993

The Candide System for Machine Translation, Berger et al, 1994

Aligning a parallel English-Chinese corpus statistically with lexical criteria, Wu, 1994

K-vec: A new approach for Aligning parallel texts, Fung, 1994

Attention, intentions, and the structure of discourse, 1986 (14.89)

Intricacies of Collins parsing model, 2004 (2.89)


11.97

6.11

6.09

6.59

7.09
Empirical analysis on ACL data

• Comparison with ACL’s most cited papers:
  – 6 out of 10 most cited papers occur in TopicFlow’s top 10 list on at least one topic
  – All 10 papers occur in TopicFlow’s top 52 on at least one topic
  – The top rated papers are ranked as low as 300-2500 on irrelevant topics
  – Examples of most cited papers that are not on TopicFlow’s top 10:
    • Broad dataset or evaluation metrics based papers
      – “Penn TreeBank” paper
      – “Bleu” paper
Citation Recommendation on ACL: results
Case 1: CORA

Document Completion Log-likelihood

Number of topics

Log-likelihood

-3.00E+06
-3.20E+06
-3.40E+06
-3.60E+06
-3.80E+06
-4.00E+06
-4.20E+06

10 20 30 40 60 80 100

LDA

TopicFlow_multsrc

TopicFlow_onessrc

RTM
Case 2: ACL

Document Completion Log-likelihood

- Log-likelihood
- LDA
- TopicFlow_multsrc
- TopicFlow_onescrc
- RTM
Conclusions

• New TopicFlow model is able to capture the notion of topical influence on par with TSP (as measured w.r.t citation recommendation)
  – Unlike TSP, does not require topic labels to be pre-specified
• Better model of text than LDA as well as competitive with RTM
• Future work
  – Model can be applied to a wide range of other textual networked data such as web, blogs, social media, etc.
  – Plan to build a graphical browser that allows users to track the propagation of influence on a specific topic
Backup slides
Topic models for text and links

- Generative Models for Citations [Nallapati et al ICWSM 2008, KDD 2008], [Chang and Blei, AISTATS, 2009]
  - Capture topical correlations between cited and citing documents
  - but do not model influence explicitly
Topic models for text and links

- Citation influence [Dietz et al, ICML 2007]
- HTM [Sun et al, EMNLP 2009]
  - Only capture local influences, not global
Topic models for text and links

• Correlations by venues, corpora, time
  [Wong et al, AISTATS 2009], [Mimno et al, NIPS WS 2008], [Daume III, IJCNLP, 2009]
  – No global influence metric
ACL: Citation Recommendation

• For each test abstract, retrieve a ranked list of most likely citations from the training set
• Evaluate using Mean Average Precision w.r.t. true citations
• Models compared:
  – TFIDF: baseline Lucene search using test abstract as query and full text of training set papers as documents.
  – + LDA: LDA cosine similarity interpolated with TFIDF
  – + RTM: similar to above

\[
\text{score}(d; q) = \frac{\sum_{k=1}^{K} \theta_d^{(k)} \theta_q^{(k)}}{\|\theta_d\| \|\theta_q\|}
\]
TopicFlow model: Update equations

- Positional distribution over topics:
  \[ \phi_{dnk} \propto \beta_{kw_n} \theta_d^{(k)} \]

- Topic distribution over words:
  \[ \beta_{kw_n} \propto \sum_{d=1}^{M} \sum_{n=1}^{N_d} \phi_{dnk} \]

LDA computes expectation of \( \Theta \) over its prior, but we don’t since there is no prior!

Same as LDA!
TopicFlow model: Update equations

- **Document outflow logistic variables:**

\[
\frac{\partial}{\partial \eta_{u,d}^k} \left( \log \{ P(w) \} \right) = \sum_{d' \in \text{Ch}(u)} \phi_{d',nk} \left( \frac{1}{f_{..,d'}} - \frac{1}{\sum_{k'} f_{..,d'}} \right) f_{..,u}^{(k)} \left( 1 - \frac{1}{|\text{Ch}(u)|} \right) \frac{\partial}{\partial \eta_{u,d}^k} \psi_{u,d'}^{(k)}
\]

- **where**

\[
\frac{\partial}{\partial \eta_{u,d}^k} \psi_{u,d'}^{(k)} = \begin{cases} \psi_{u,d}^k \left( 1 - \psi_{u,d}^k \right) & \text{if } d = d' \\ -\psi_{u,d}^k \psi_{u,d'}^k & \text{otherwise} \end{cases}
\]

Positive for my document; negative for others
TopicFlow model: Update equations

- Source outflow logistic variables:

\[
\frac{\partial}{\partial \eta_{s,d}^k} \left( \log \{ P(w) \} \right)
= \begin{cases} 
\sum_{d'} \sum_{n=1}^{N_{d'}} \phi_{d'n k} \left( \frac{1}{f_{.,d'}} - \frac{1}{\sum_{k'} f_{.,d'}} \right) f_{.,s} \frac{\partial}{\partial \eta_{s,d}^k} \psi_{s,d'}^{(k)} & \text{if independent sources} \\
\sum_{d'} \sum_{k'} \sum_{n=1}^{N_{d'}} \phi_{d'n k'} f_{.,s} \left( \frac{1}{f_{.,d'}} \frac{\partial}{\partial \eta_{s,d}^k} \psi_{s,d'}^{(k')} - \frac{1}{\sum_{k'} f_{.,d'}} \right) \sum_{k''} \frac{\partial}{\partial \eta_{s,d}^k} \psi_{s,d'}^{(k'')} & \text{if single source} 
\end{cases}
\]

- Appears \( \mathcal{O}(M^2 K) \) for multiple sources and \( \mathcal{O}(M^2 K^3) \) for single source

- Can use some dynamic programming like tricks to reduce it to \( \mathcal{O}(MK) \) for both!!
**TopicFlow: Algorithm**

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>for each document $d = 1, \cdots, M$:</td>
</tr>
<tr>
<td>2.</td>
<td>for each topic $k = 1, \cdots, K$:</td>
</tr>
<tr>
<td>3.</td>
<td>update $\theta_d^{(k)}$ using Eq. 6</td>
</tr>
<tr>
<td>4.</td>
<td>for each position $n = 1, \cdots, N_d$:</td>
</tr>
<tr>
<td>5.</td>
<td>update $\phi_{dnk}$ using Eq. 14</td>
</tr>
<tr>
<td>6.</td>
<td>for each topic $k = 1, \cdots, K$:</td>
</tr>
<tr>
<td>7.</td>
<td>for each word $w = 1, \cdots, V$:</td>
</tr>
<tr>
<td>8.</td>
<td>update $\beta_{kw}$ using Eq. 13</td>
</tr>
<tr>
<td>9.</td>
<td>Learn outflows of the document on topic $k$ using Eq. 20</td>
</tr>
<tr>
<td>10.</td>
<td>Learn the source outflows using Eq. 22</td>
</tr>
<tr>
<td>11.</td>
<td>Synchronize all flows</td>
</tr>
</tbody>
</table>

Topic distributions over words are not estimated at inference time.
TopicFlow model: Learning and Inference

• Trick 1: variational lower bound for log likelihood:

\[
\log P(w) = \sum_{d=1}^{M} \sum_{n=1}^{Nd} \log \left( \sum_{k=1}^{K} \beta_{kw_n} \theta^{(k)}_d \right)
\]

\[
\geq \sum_{d=1}^{M} \sum_{n=1}^{Nd} \sum_{k=1}^{K} \phi_{dnk} (\log \beta_{kw_n} + \log \theta^{(k)}_d - \log \phi_{dnk})
\]

– Easier to learn using EM style optimization
TopicFlow model: Learning and Inference

• Trick 2: Elimination of equality constraints:
  – Converting flow balance constraints into multinomial constraints:
    \[
    \forall j \in \text{ch}(d) \quad f_{d,j}^{(k)} = f_{.,d}^{(k)} \psi_{d,j}^{(k)} \quad \text{s.t.} \quad \sum_{j \in \text{ch}(d)} \psi_{d,j}^{(k)} = 1 \text{ and } \forall j \psi_{d,j} \geq 0
    \]
  – Eliminating multinomial constraints using a logistic transformation:
    \[
    \psi_{s,d}^{k} = \frac{\exp(\eta_{s,d}^{k})}{\sum_{k'} \sum_{d' \in \text{ch}(s)} \exp(\eta_{s,d'}^{k'})}
    \]
    \[
    \psi_{d,v}^{k} = \begin{cases} 
    \frac{1}{|\text{ch}(d)|} \sum_{v' \in \text{ch}(d) - s} \exp(\eta_{d,v'}^{k}) & \text{if } v \neq t \\
    \frac{1}{|\text{ch}(d)|} & \text{if } v = t 
    \end{cases}
    \]

Single source

Document
Dataset 1: ACL anthology

• Training set:
  – full-content of 9,824 papers published in or before 2005
  – 36,604 hyperlinks

• Test set:
  – abstracts of 1,040 test document published after 2005
  – No hyperlinks within the test set
  – Hyperlinks to the training documents used only for testing on citation recommendation task

• 46,160 unique words after stopping and stemming
Dataset 2: CORA

- Abstracts of research papers in CS
- Removed documents with no links
- Training set:
  - 11,442 documents
  - 24,582 links
- Test set
  - 11,230 documents
  - 23,379 links
- 7,185 unique words after stopping and stemming
ACL: Citation Recommendation

- Models compared:
  - + TopicFlow: expected Influence of the document with respect to the topical distribution of the query
    \[ I(d, k) = (f^{(k)}_{t,d} - f^{(k)}_{s,d})\theta^{(k)}_{d} \]
  - + Topic Sensitive PageRank
    - Same as above, but TopicFlow replaced by TSP
    - For each document, topic labels computed from the argmax of topic distributions from the multi-source TopicFlow model.
    - Tuned the teleportation prior
- For each model, tuned the parameter that interpolates model score with TFIDF score
  - Tuned using grid search on a 30 topic model on a development set
  - Tested on models with different number of topics
Train the model on the one half of each document and measure the model’s likelihood on the remaining half.
  
  Higher likelihood implies better predictive power.

Models used:

- Latent Dirichlet Allocation:
  
  - uses no hyperlink information

- RTM [Chang and Blei, AISTATS 2009]:
  
  - State of the art model for links and text

- TopicFlow model
  
  - One source
  
  - Independent topical sources

All models are made to predict the likelihood of words only.
  
  - Allows fair comparison
Acknowledgments

- Dan Ramage
  - TSP implementation
  - Other important modeling suggestions
  - Reviewing the manuscript
- David Vickrey and Rajat Raina
  - Suggestions on optimization
- Steve Bethard
  - Citation retrieval implementation
  - Wilcoxon test software
Resources

• Implementation available in JavaNLP under Research project as `edu.stanford.nlp.topicmodeling.topicflow`
  – Read `Package.html` for usage instructions

• Demo available at:

• This presentation is available at


• Final version under review at AISTATS 2011.
Introduction

• Candidate solution: Topic Sensitive PageRank
  [Haveliwala, 2003]
  – Variant of PageRank
    • Stationary distribution of a random web surfer
    • Has a teleportation probability to reach any document from any other document to satisfy ergodicity and acyclicity
  – ‘Teleportation’ probability is assigned to only documents on that topic
  – However: requires topic labels to be pre-specified
• Many document corpora are unlabeled
  – Question: can we learn both topics and topical authoritativeness simultaneously?
TopicFlow model: Learning and Inference

• Objective function

$$\max_{f, \beta} \sum_{d=1}^{M} \sum_{n=1}^{N_d} \log\left( \sum_{k=1}^{K} \beta_k \omega_n \theta_d^{(k)} \right)$$

s.t. $$\forall d \in V' - \{s\} \sum_{i \in \text{Pa}(d)} f^{(k)}_{i,d} = \sum_{j \in \text{Ch}(d)} f^{(k)}_{d,j} \text{ and}$$

$$\begin{cases} \sum_{k=1}^{K} \sum_{d=1}^{M} f^{(k)}_{s,d} = 1 & \text{If single source} \\ \sum_{d=1}^{M} f^{(k)}_{s,d} = 1 & \text{If multiple topic sources} \end{cases}$$