Topic Models with Logical Constraints on Words

Hayato Kobayashi, Hiromi Wakaki, Tomohiro Yamasaki, and Masaru Suzuki
Corporate Research and Development Center,
Toshiba Corporation, Japan
Topic modeling = Word clustering

- Method to extract latent topics on a corpus
  - Each topic is a distribution on words
Topic modeling = Word clustering

- Method to extract latent topics on a corpus
  - Each topic is a distribution on words

Corpus about Bulgaria

- yogurt
- milk
- food
- fruit
- bacteria
- fat
- cream

...
Topic modeling = Word clustering

• Method to extract latent topics on a corpus
  • Each topic is a distribution on words

Corpus about Bulgaria

... yogurt milk food fruit bacteria fat cream ...
... rose oil organic essential valley pure kazanlak ...
...
Topic modeling = Word clustering

• Method to extract latent topics on a corpus
  • Each topic is a distribution on words
Want to split into “fire dance” and “sexy dance”
Existing work [Andrzejewski+ ICML2009]

• Constraints on words for topic modeling
  • Must-Link(A,B) : A and B appear in the same topic
  • Cannot-Link(A,B) : A and B don’t appear in the same topic

Want to split into “fire dance” and “sexy dance”
Problem of the existing work

- Constraints often don’t align with user’s intention

You might get “blaze” topic instead of “fire dance” topic

Want to split into “fire dance” and “sexy dance”

Cannot-Link(fire, sexy)
This work

- **Logical** constraints on words for topic modeling
  - Conjunctions ($\land$), disjunctions ($\lor$), negations ($\neg$)

Want to split into “fire dance” and “sexy dance”

\[
\text{Cannot-Link}(\text{fire, sexy}) \\
\land (\text{Must-Link}(\text{dance, fire}) \\
\lor \text{Must-Link}(\text{dance, sexy}))
\]
Outline of the rest of this talk

• LDA [Blei+ JMLR2003]
  • One of topic modeling method
• LDA-DF [Andrzejewski+ ICML2009]
  • Must-Link and Cannot-Link
• This work
  • Logical expressions of Must-Links and Cannot-Links
  • Experiment
• Conclusion
Latent Dirichlet Allocation (LDA) [Blei+ JMLR2003]

• Famous Topic modeling method

(i) Assume a generative model of documents
   • Each topic is a distribution on words
   • Each document is a distribution on topics
     • Taken from Dirichlet distributions to generate discrete distributions

(ii) Infer parameters for the two distributions
     inverting the generative model
Generative process of documents in LDA

- Each topic is a distribution on words
- Each document is a distribution on topics
Generative process of documents in LDA

- Each topic is a distribution on words
- Each document is a distribution on topics
Generative process of documents in LDA

• Each topic is a distribution on words
• Each document is a distribution on topics
Generative process of documents in LDA

• Each topic is a distribution on words
• Each document is a distribution on topics
Parameter inference in LDA

• Infer word and topic distributions from a corpus inverting the generative process
LDA-DF [Andrzejewski+ ICML2009]

- Semi-supervised extension of LDA
  - Only conjunction of Must-Links and Cannot-Links
    - Must-Link(A,B) : A and B appear in the same topic
    - Cannot-Link(A,B) : A and B don’t appear in the same topic

- Extending the generative process
  - Each topic is a constrained distribution on words
    - Taken from a Dirichlet tree distribution, which is a generalization of a Dirichlet distribution
  - Each document is a distribution on topics
    - Taken from a Dirichlet distribution
Generative process of LDA-DF

• Always generates a distribution, where yogurt and rose do not appear in the same topic.
Algorithm to generate distributions in LDA-DF

1. Map links to a graph
2. Contract Must-Links
3. Extract the maximal independent sets (MIS)
4. Generate a distribution based on each MIS
Algorithm to generate distributions in LDA-DF

1. Map links to a graph
   • Any conjunction of links can be mapped to a graph

\[
\text{Cannot-Link}(A,B) \land \text{Cannot-Link}(E,G) \\
\land \text{Must-Link}(B,E) \land \text{Must-Link}(C,D)
\]
Algorithm to generate distributions in LDA-DF

2. Contract Must-Links
   • Regard two words on each Must-Link as one word
Algorithm to generate distributions in LDA-DF

3. Extract the maximal independent sets (MIS)
   • MIS = Maximal set of nodes without edges
Algorithm to generate distributions in LDA-DF

4. Generate a distribution based on each MIS
   - Equalize the frequencies of contracted words
   - Zero the frequencies of words not in the MIS

[Diagram showing the algorithm with examples of words CD, F, BE, and CL, ML, CL]
This work

• Algorithm to generate logically constrained distributions on LDA-DF
  • We can not apply the existing algorithm

\[
( \neg \text{Cannot-Link}(A,B) \\
\lor \text{Must-Link}(A,C))
\land \text{Cannot-Link}(B,C)
\]

Words → Nodes
Links → Edges

This constraint cannot be mapped to a graph
Negations

• Delete negations (¬) in a preprocessing stage
  • Weak negation: ¬ Must-Link(A,B) = no constraint
    (A and B need not appear in the same topic)
  • Strong negation: ¬ Must-Link(A,B) = Cannot-Link(A,B)
    (A and B must not appear in the same topic)

\[
(¬ Cannot-Link(A,B) \lor \text{Must-Link}(A,C)) \land Cannot-Link(B,C) \quad \Rightarrow \quad (\text{Must-Link}(A,B) \lor \text{Must-Link}(A,C)) \land Cannot-Link(B,C)
\]

Focus only on conjunctions and disjunctions
Key observation for logical expressions

- Any constrained distribution is represented by a conjunctive expression by two primitives
  - `EqualPrim(A, B)`: makes $p(A) \approx p(B)$
  - `ZeroPrim(A)`: makes $p(A) \approx 0$

$$
\text{EqualPrim}(B, E) \land \text{EqualPrim}(C, D) \\
\land \text{ZeroPrim}(A) \land \text{ZeroPrim}(G)
$$
Substitution of links with primitives

• Must-Link(A,B) = EqualPrim(A,B)

• Cannot-Links(A,B) = ZeroPrim(A) ∨ ZeroPrim(B)

These two distributions satisfy Cannot-Link(A,B)
Proposed algorithm for logical expressions

1. Substitute links with primitives
2. Calculate the minimum disjunctive normal form (DNF) of the primitives
3. Generate distributions for each conjunction of the DNF
Proposed algorithm for logical expressions

1. Substitute links with primitives

\[(\text{Must-Link}(A,B) \lor \text{Must-Link}(A,C)) \land \text{Cannot-Link}(B,C)\]

Primitives

\[
\begin{align*}
\text{EqualPrim}(A,B) & \land \text{EqualPrim}(A,C) \\
\text{ZeroPrim}(B) & \land \text{ZeroPrim}(C)
\end{align*}
\]
Proposed algorithm for logical expressions

2. Calculate the minimum disjunctive normal form (DNF) of the primitives
   • DNF = Disjunction of conjunctions of primitives
Proposed algorithm for logical expressions

3. Generate distributions for each conjunction of the DNF

Combine each conjunction of primitives
Correctness of our method

• [Theorem] Our method and the existing method are asymptotically equivalent w.r.t. conjunctive expressions of links

\[ CL(A,B) \land CL(A,C) \]

Distributions by primitives are the same as distributions by a graph
Customization of new links

• Isolate-Link (ISL)
  • $X_1,\ldots,X_n$ do not appear (nearly)
    (Remove unnecessary words and stop words)
    $$\text{ISL}(X_1,\ldots,X_n) = \bigwedge_{i=1}^{n} \text{ZeroPrim}(X_i)$$

• Imply-Link (IL)
  • $B$ appears if $A$ appears in a topic ($A \rightarrow B$)
    (Use when $B$ has multiple meanings)
    $$\text{IL}(A, B) = \text{EqualPrim}(A, B) \lor \text{ZeroPrim}(A)$$

• Extended Imply-Link (XIL)
  • $Y$ appears if $X_1,\ldots,X_n$ appear in a topic ($X_1,\ldots,X_n \rightarrow Y$)
    $$\text{XIL}(X_1,\ldots,X_n,Y) = \bigwedge_{i=1}^{n} \text{EqualPrim}(X_i, Y) \lor \bigvee_{i=1}^{n} \text{ZeroPrim}(X_i)$$
Interactive topic analysis

• Movie review corpus (1000 reviews) [Pang&Lee ACL2004]
  • No constraints

<table>
<thead>
<tr>
<th>Topic</th>
<th>High frequency words</th>
</tr>
</thead>
<tbody>
<tr>
<td>?</td>
<td>have give night film turn performance year mother take out</td>
</tr>
<tr>
<td>?</td>
<td>not life have own first only family tell yet moment even</td>
</tr>
<tr>
<td>?</td>
<td>movie have n’t get good not see know just other time make</td>
</tr>
<tr>
<td>?</td>
<td>have black scene tom death die joe ryan man final private</td>
</tr>
<tr>
<td>?</td>
<td>film have n’t not make out well see just very watch even</td>
</tr>
<tr>
<td>?</td>
<td>have film original new never more evil n’t time power</td>
</tr>
</tbody>
</table>

All topics are unclear
Interactive topic analysis

- Movie review corpus (1000 reviews)
  - Isolate-Link(\textit{have, film, movie, not, n’t})
    - Remove specified words as well as related unnecessary words

<table>
<thead>
<tr>
<th>Topic</th>
<th>High frequency words</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Isolated)</td>
<td>\textit{have film movie not} good make n’t character see more get</td>
</tr>
<tr>
<td>?</td>
<td>\textit{star war trek} planet effect special lucas jedi science</td>
</tr>
<tr>
<td>Comedy</td>
<td>comedy funny laugh school hilarious evil power bulworth</td>
</tr>
<tr>
<td>Disney</td>
<td>disney voice mulan animated song feature tarzan animation</td>
</tr>
<tr>
<td>Family</td>
<td>life love family mother woman father child relationship</td>
</tr>
<tr>
<td>Thriller</td>
<td>truman murder killer death thriller carrey final detective</td>
</tr>
</tbody>
</table>

“Star Wars” and “Star Trek” are merged, although most topics are clear.
Interactive topic analysis

- Movie review corpus (1000 reviews)
  - Isolate-Link(\textit{have, film, movie, not, n’t})
  - Cannot-Link(\textit{jedi, trek})

\[\text{Dared to select “jedi” since “star” and “war” are too common}\]

<table>
<thead>
<tr>
<th>Topic</th>
<th>High frequency words</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Isolated)</td>
<td>\textit{have film movie not make good n’t character see more get}</td>
</tr>
<tr>
<td>Star Wars</td>
<td>\textit{star war lucas effect jedi special matrix menace computer}</td>
</tr>
<tr>
<td>Comedy</td>
<td>funny comedy laugh get hilarious high joke humor bob smith</td>
</tr>
<tr>
<td>Disney</td>
<td>disney truman voice toy show animation animated tarzan</td>
</tr>
<tr>
<td>Family</td>
<td>family father mother boy child son parent wife performance</td>
</tr>
<tr>
<td>Thriller</td>
<td>killer murder case lawyer man david prison performance</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

\[\text{“Star Trek” disappears, although “Star Wars” is obtained}\]
Interactive topic analysis

• Movie review corpus (1000 reviews)
  • Isolate-Link(\texttt{have, film, movie, not, n’t})
    \wedge \text{Cannot-Link(}\texttt{jedi, trek})
    \wedge (\text{Must-Link(}\texttt{star, jedi}) \lor \text{Must-Link(}\texttt{star, trek}))

<table>
<thead>
<tr>
<th>Topic</th>
<th>High frequency words</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Isolated)</td>
<td>have film movie not make good n’t character see more get</td>
</tr>
<tr>
<td>Star Wars</td>
<td>star war toy jedi menace phantom phantom lucas burton planet</td>
</tr>
<tr>
<td>Star Trek</td>
<td>alien effect star science special trek action computer</td>
</tr>
<tr>
<td>Comedy</td>
<td>comedy funny laugh hilarious joke get ben john humor fun</td>
</tr>
<tr>
<td>Disney</td>
<td>disney voice animated mulan animation family tarzan shrek</td>
</tr>
<tr>
<td>Family</td>
<td>life love family man story child woman young mother</td>
</tr>
<tr>
<td>Thriller</td>
<td>scream horror flynt murder killer lawyer death sequel case</td>
</tr>
</tbody>
</table>

We obtained “Star Wars” and “Star Trek” appropriately.
Conclusion

• Simple algorithm for logical constraints on words for topic modeling
  • **Must-Link**(A,B) : A and B appear in the same topic
  • **Cannot-Link**(A,B) : A and B do not appear in the same topic

• Theorem for the correctness of the algorithm

• Customization of new links
  • **Isolate-Link**(X₁, …, Xₙ) : X₁, …, Xₙ disappear
  • **Imply-Link**(A, B) : B appears if A appears in a topic

• Future Work
  • Comparative experiments on real corpora
Thank you for your attention
Appendix: Visualization of Priors

ML = Must-Link, CL = Cannot-Link, IL = Imply-Link
Appendix: Visualization of Priors

(a) $ML(A, B) \land ML(A, C)$
(b) $ML(A, B) \land CL(B, C)$
(c) $ML(A, B) \lor ML(A, C)$
(d) $IL(B, A) \land IL(C, A)$

ML = Must-Link, CL = Cannot-Link, IL = Imply-Link
Appendix: Visualization of Priors

(e) \((ML(A, B) \lor ML(A, C)) \land CL(B, C)\)

(f) \(IL(B, A) \land IL(C, A) \land CL(B, C)\)

(g) \(ML(C, A) \triangleright ML(C, B)\)

ML = Must-Link, CL = Cannot-Link, IL = Imply-Link