Incorporating Domain Knowledge in Latent Topic Models

Final oral examination

David Andrzejewski
Department of Computer Sciences
University of Wisconsin–Madison

24th August 2010
Documents
Topic modeling overview

Documents

Word Counts

\[
\begin{align*}
\text{Documents Word Counts} &= [5 \ 2 \ ... \ 7] \\
\text{Politics} &= [0 \ 6 \ ... \ 1] \\
\text{Andrzejewski (UW-Madison)} &= [3 \ 8 \ ... \ 2]
\end{align*}
\]
Topic modeling overview

Documents

Word Counts

\[ \begin{align*}
\text{Politics} & = [5 \ 2 \ \ldots \ 7] \\
\text{Algorithms} & = [0 \ 6 \ \ldots \ 1] \\
\text{Sports} & = [3 \ 8 \ \ldots \ 2]
\end{align*} \]

Topics

Politics
Sports
Algorithms
Topic modeling overview

Documents

Word Counts

Politics

Algorithms

Sports

Topics

Which words in each topic?
Topic modeling overview

Documents

Word Counts

\[
\begin{bmatrix}
5 & 2 & \ldots & 7
\end{bmatrix}
\]
\[
\begin{bmatrix}
0 & 6 & \ldots & 1
\end{bmatrix}
\]
\[
\begin{bmatrix}
3 & 8 & \ldots & 2
\end{bmatrix}
\]

Which topics in each doc?

Which words in each topic?

Politics
Sports
Algorithms

Andrzejewski (UW-Madison)
Topic modeling overview

Documents

Word Counts

Politics

Algorithms

Sports

Topics

Which topics in each doc?

Which words in each topic?

LEARNED FROM DATA
Topic modeling overview

Documents

Word Counts

\[ \begin{bmatrix}
5 & 2 & \ldots & 7 \\
3 & 8 & \ldots & 2 \\
0 & 6 & \ldots & 1
\end{bmatrix} \]

Which topics in each doc?

Which words in each topic?

LEARNED FROM DATA + DOMAIN KNOWLEDGE

Topics

Politics
Sports
Algorithms
Dataset: 89,574 New Year’s wishes (NYC Times Square website)

Goal: understand/summarize common wish themes

Example wishes:
- Peace on earth
- Own a brewery
- I hope I get into Univ. of Penn graduate school.
- The safe return of my friends in Iraq
- Find a cure for cancer
- To lose weight and get a boyfriend
- I Hope Barack Obama Wins the Presidency
- To win the lottery!
Exploratory browsing with topic models
Goldberg et al., NAACL HLT 2009

- Dataset: 89,574 New Year’s wishes (NYC Times Square website)
- Goal: understand/summarize common wish themes
- Example wishes:
  - Peace on earth
  - own a brewery
  - I hope I get into Univ. of Penn graduate school.
  - The safe return of my friends in Iraq
  - find a cure for cancer
  - To lose weight and get a boyfriend
  - I Hope Barack Obama Wins the Presidency
  - To win the lottery!
Corpus-wide word frequencies
Why latent topic modeling?

- Author/document profiling
  - Match papers to reviewers (Mimno & McCallum, 2007)
  - Assign developers to bugs (Linstead et al., 2007)
  - Scientific impact/influence (Gerrish & Blei, 2009)

- Networks (Henderson & Eliassi-Rad, 2009)

- Trends (Wang & McCallum, 2006)

- Info retrieval (http://rexa.info)
Why latent topic modeling?

- Author/document profiling
  - Match papers to reviewers (Mimno & McCallum, 2007)
  - Assign developers to bugs (Linstead et al., 2007)
  - Scientific impact/influence (Gerrish & Blei, 2009)
- Networks (Henderson & Eliassi-Rad, 2009)
- Trends (Wang & McCallum, 2006)
- Info retrieval (http://rexa.info)
Why latent topic modeling?

- Author/document profiling
  - Match papers to reviewers (Mimno & McCallum, 2007)
  - Assign developers to bugs (Linstead et al., 2007)
  - Scientific impact/influence (Gerrish & Blei, 2009)

- Networks (Henderson & Eliassi-Rad, 2009)

- Trends (Wang & McCallum, 2006)
  - Info retrieval (http://rexa.info)
Why latent topic modeling?

- Author/document profiling
  - Match papers to reviewers (Mimno & McCallum, 2007)
  - Assign developers to bugs (Linstead et al., 2007)
  - Scientific impact/influence (Gerrish & Blei, 2009)
- Networks (Henderson & Eliassi-Rad, 2009)
- Trends (Wang & McCallum, 2006)
- Info retrieval (http://rexa.info)

[Bayesian topic modeling interface]

- View all topics sorted by citations | topic diversity | H-Index

<table>
<thead>
<tr>
<th>Topic Terms</th>
<th>Words</th>
<th>Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>bayesian</td>
<td>0.0607</td>
<td>0.0576 monte carlo</td>
</tr>
<tr>
<td>model</td>
<td>0.0330</td>
<td>0.0079 monte carlo simulation</td>
</tr>
<tr>
<td>inference</td>
<td>0.0226</td>
<td>0.0055 monte carlo simulations</td>
</tr>
</tbody>
</table>
Dirichlet
[20, 5, 5]
Quick stats review

Dirichlet
[20, 5, 5]

Multinomial
[0.6, 0.15, 0.25]
Quick stats review

Dirichlet
[20, 5, 5]

Multinomial
[0.6, 0.15, 0.25]

Observed counts
[3, 1, 2]

A, A, B, C, A, C

Andrzejewski (UW-Madison)
Incorporating Domain Knowledge
Final defense
Quick stats review

Dirichlet [20, 5, 5]

Multinomial [0.6, 0.15, 0.25]

Observed counts [3, 1, 2]

A, A, B, C, A, C

CONJUGACY!
For each topic $t$
    $\phi_t \sim \text{Dirichlet}(\beta)$

For each document $d$
    $\theta_d \sim \text{Dirichlet}(\alpha)$
    For each word $w$
    Topic $z \sim \text{Multinomial}(\theta_d)$
    Word $w \sim \text{Multinomial}(\phi_z)$
For each topic $t$

$$\phi_t \sim \text{Dirichlet}(\beta)$$

For each document $d$

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

For each word $w$

- Topic $z \sim \text{Multinomial}(\theta_d)$
- Word $w \sim \text{Multinomial}(\phi_z)$
For each topic $t$
\[ \phi_t \sim \text{Dirichlet}(\beta) \]
For each document $d$
\[ \theta_d \sim \text{Dirichlet}(\alpha) \]
For each word $w$
- Topic $z \sim \text{Multinomial}(\theta_d)$
- Word $w \sim \text{Multinomial}(\phi_z)$
For each topic $t$
\[ \phi_t \sim \text{Dirichlet}(\beta) \]

For each document $d$
\[ \theta_d \sim \text{Dirichlet}(\alpha) \]

For each word $w$
- Topic $z \sim \text{Multinomial}(\theta_d)$
- Word $w \sim \text{Multinomial}(\phi_z)$
Latent Dirichlet Allocation (LDA)
Blei et al., JMLR 2003

For each topic $t$
$$\phi_t \sim \text{Dirichlet}(\beta)$$

For each document $d$
$$\theta_d \sim \text{Dirichlet}(\alpha)$$

For each word $w$
Topic $z \sim \text{Multinomial}(\theta_d)$
Word $w \sim \text{Multinomial}(\phi_z)$
Latent Dirichlet Allocation (LDA)
Blei et al., JMLR 2003

For each topic $t$
\[ \phi_t \sim \text{Dirichlet}(\beta) \]

For each document $d$
\[ \theta_d \sim \text{Dirichlet}(\alpha) \]

For each word $w$

- Topic $z \sim \text{Multinomial}(\theta_d)$
- Word $w \sim \text{Multinomial}(\phi_z)$
For each topic $t$

$$\phi_t \sim \text{Dirichlet}(\beta)$$

For each document $d$

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

For each word $w$

Topic $z \sim \text{Multinomial}(\theta_d)$

Word $w \sim \text{Multinomial}(\phi_z)$
For each topic $t$

$$\phi_t \sim \text{Dirichlet}(\beta)$$

For each document $d$

$$\theta_d \sim \text{Dirichlet}(\alpha)$$

For each word $w$

Topic $z \sim \text{Multinomial}(\theta_d)$

Word $w \sim \text{Multinomial}(\phi_z)$
Given observed words $\mathbf{w}$, want the posterior over hidden topics $\mathbf{z}$

Use Bayes’ Rule to calculate $P(\mathbf{z}|\mathbf{w})$? INTRACTABLE

One approach: Collapsed Gibbs Sampling

$$
\begin{align*}
\mathbf{w} &= \text{A B B A C B A A} \\
\mathbf{z} &= \text{0 2 2 1 2 1 0 0}
\end{align*}
$$
Given observed words $w$, want the posterior over hidden topics $z$

Use Bayes’ Rule to calculate $P(z|w)$? INTRACTABLE

One approach: Collapsed Gibbs Sampling
Given observed words $\mathbf{w}$, want the posterior over hidden topics $\mathbf{z}$

Use Bayes’ Rule to calculate $P(\mathbf{z}|\mathbf{w})$? **INTRACTABLE**

One approach: Collapsed Gibbs Sampling

$$\mathbf{w} = A\ B\ B\ A\ C\ B\ A\ A$$
$$\mathbf{z} = 0\ 2\ 2\ 2\ 1\ 2\ 1\ 0\ 0$$
Given observed words $\mathbf{w}$, want the posterior over hidden topics $\mathbf{z}$

Use Bayes’ Rule to calculate $P(\mathbf{z}|\mathbf{w})$? **INTRACTABLE**

One approach: Collapsed Gibbs Sampling

\[
\begin{align*}
\mathbf{w} &= \text{A B B A C B A A} \\
\mathbf{z} &= \text{0 2 2 1 2 1 0 0 0}
\end{align*}
\]
Given observed words $w$, want the posterior over hidden topics $z$.

Use Bayes’ Rule to calculate $P(z|w)$? INTRACTABLE

One approach: Collapsed Gibbs Sampling

$w = \begin{array}{cccccccc} A & B & B & A & C & B & A & A \\ \end{array}$

$z = \begin{array}{cccccccc} 2 & 2 & 2 & 1 & 2 & 1 & 0 & 0 \\ \end{array}$
Given observed words \( w \), want the posterior over hidden topics \( z \).

Use Bayes’ Rule to calculate \( P(z|w) \)? **INTRACTABLE**

One approach: Collapsed Gibbs Sampling

\[
\begin{align*}
  w &= A \ B \ B \ A \ C \ B \ A \ A \\
  z &= 2 \ 1 \ 2 \ 1 \ 2 \ 1 \ 0 \ 0
\end{align*}
\]
Given observed words $w$, want the posterior over hidden topics $z$

Use Bayes’ Rule to calculate $P(z|w)$? **INTRACTABLE**

One approach: Collapsed Gibbs Sampling

\[
\begin{align*}
\mathbf{w} &= \text{A B B A C B A A} \\
\mathbf{z} &= \text{2 1 2 1 2 1 0 0}
\end{align*}
\]
Given observed words \( w \), want the posterior over hidden topics \( z \).

Use Bayes’ Rule to calculate \( P(z|w) \)? **INTRACTABLE**

One approach: Collapsed Gibbs Sampling

\[
\begin{align*}
\text{w} &= \text{A B B A C B A A} \\
\text{z} &= 2 \ 1 \ 2 \ 2 \ 2 \ 1 \ 0 \ 0 \ 0
\end{align*}
\]
Outline

1. Topic modeling
   - Latent Dirichlet Allocation (LDA)
   - Issues
   - Related work

2. Preliminary work

3. New work
   - LogicLDA
   - Biological text mining

4. Conclusion
   - Discussion
   - Future work
Topics may not align with user goals / intuitions

- Topic = \{north south carolina korea korean...\}
  (Newman et al, 2009)

- 20 news: “mac” vs “pc”
Topics may not align with user goals / intuitions

- Topic = {north south carolina korea korean...}
  (Newman et al, 2009)
- 20 news: “mac” vs “pc”
Topics may not align with user goals / intuitions

- Topic = \{north south carolina korea korean...\}  
  (Newman et al, 2009)
- 20 news: “mac” vs “pc”
Topics may not align with user goals / intuitions

- Topic = \{north south carolina korea korean...\} (Newman et al, 2009)
- 20 news: “mac” vs “pc”
Popular demand

Posted question on machine learning website*

“LDA is nice, but unpredictable as it does not always give me the topics I wanted...I am looking for something like LDA, but...i can pick the seed words for each topic...”

* http://metaoptimize.com/qa/
Extending LDA - existing work

- LDA: generative probabilistic model → extensible
  - Associated observations (e.g., images-Blei and Jordan, 2003)
  - Document-topic (e.g., correlated topics-Blei and Lafferty 2006)
  - Topic-word (e.g., bigram-Wallach, 2006)
Extending LDA - existing work

- LDA: generative probabilistic model → extensible
- Associated observations (e.g., images-Blei and Jordan, 2003)
- Document-topic (e.g., correlated topics-Blei and Lafferty 2006)
- Topic-word (e.g., bigram-Wallach, 2006)
Extending LDA - existing work

- LDA: generative probabilistic model → extensible
- Associated observations (e.g., images-Blei and Jordan, 2003)
- Document-topic (e.g., correlated topics-Blei and Lafferty 2006)
- Topic-word (e.g., bigram-Wallach, 2006)
Extending LDA - existing work

- LDA: generative probabilistic model → extensible
- Associated observations (e.g., images-Blei and Jordan, 2003)
- Document-topic (e.g., correlated topics-Blei and Lafferty 2006)
- Topic-word (e.g., bigram-Wallach, 2006)
How this work is different

- Existing work: application-specific, richer structure
- This work: general *direct* domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in same sentence”
How this work is different

- Existing work: application-specific, richer structure
- This work: general *direct* domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in same sentence”
How this work is different

- Existing work: application-specific, richer structure
- This work: general \textit{direct} domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in same sentence”
How this work is different

- Existing work: application-specific, richer structure
- This work: general *direct* domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in same sentence”
How this work is different

- Existing work: application-specific, richer structure
- This work: general *direct* domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in the same sentence”
How this work is different

- Existing work: application-specific, richer structure
- This work: general *direct* domain knowledge
  - “These words do not belong in the same topic”
  - “I want a topic about X”
  - “This topic is incompatible with this document”
  - “These two topics are incompatible and should not co-occur in same sentence”
Specific aims from preliminary examination

Specific Aim 1
Develop latent topic models capable of expressing novel forms of domain knowledge (LogicLDA).

Specific Aim 2
Apply knowledge-augmented latent topic models to real-world problems (biological text mining application).
Specific aims from preliminary examination

Specific Aim 1
Develop latent topic models capable of expressing novel forms of domain knowledge (LogicLDA).

Specific Aim 2
Apply knowledge-augmented latent topic models to real-world problems (biological text mining application).
Overview

Special topics which only appear in certain documents

- Delta LDA
- LDA
Overview

Influence latent topic assignment $z_i$ of individual tokens

LDA → Delta LDA → Topic in set
Must-Link and Cannot-Link preferences over words

- Delta LDA
- Topic in set
- LDA
- Dirichlet Forest
Overview

General domain knowledge

- Delta LDA
- Topic in set
- Logic LDA
- Dirichlet Forest
- LDA
Application case study

LDA

Delta LDA → Topic in set → Logic LDA

Dirichlet Forest → Biological text mining application
Outline

1 Topic modeling
   - Latent Dirichlet Allocation (LDA)
   - Issues
   - Related work

2 Preliminary work

3 New work
   - LogicLDA
   - Biological text mining

4 Conclusion
   - Discussion
   - Future work
Restricted topics with ΔLDA
Andrzejewski et al, ECML 2007

Contribution

Special topics which only appear in certain documents

- Document label determines $\alpha$ prior on document-topic $\theta$
- Statistical debugging: we know some runs fail (crash or bad output)
Restricted topics with $\Delta$LDA
Andrzejewski et al, ECML 2007

Contribution

Special topics which only appear in certain documents

- Document label determines $\alpha$ prior on document-topic $\theta$
- Statistical debugging: we know some runs fail (crash or bad output)

$$\alpha = \begin{bmatrix} \alpha^{(s)} \\ \alpha^{(f)} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$
Restricted topics with $\Delta$LDA
Andrzejewski et al, ECML 2007

Contribution

Special topics which only appear in certain documents

- Document label determines $\alpha$ prior on document-topic $\theta$
- Statistical debugging: we know some runs fail (crash or bad output)

```c
int x = my_func()
if (x > 5) {
    branch_42_true++
    ...
} else {
    branch_42_false++
    ...
}
```
**Influence latent topic assignment $z_i$ of individual tokens**

- “**Apple pie**” vs “**Apple iPod**”
- Use “seed” words \{translation, tRNA, ribosome, anticodon\}

\[
\mathbf{w} = \begin{array}{cccccccc}
\end{array}
\]

\[
\mathbf{z} = \begin{array}{cccccccc}
? & 0 & 0 & ? & ? & 0 & ? & ?
\end{array}
\]
Contribution

Influence latent topic assignment $z_i$ of individual tokens

- “Apple pie” vs “Apple iPod”
- Use “seed” words {translation, tRNA, ribosome, anticodon}
Contribution

Influence latent topic assignment $z_i$ of individual tokens

- “Apple pie” vs “Apple iPod”
- Use “seed” words \{translation, tRNA, ribosome, anticodon\}
**Contribution**

Influence latent topic assignment $z_i$ of individual tokens

- “Apple pie” vs “Apple iPod”
- Use “seed” words \{translation, tRNA, ribosome, anticodon\}

| Topic 0 | translation, ribosomal, trna, rrna, initiation, ribosome, protein ribosomes, is, factor, processing, translational, nucleolar pre-rrna, synthesis, small, 60s, eukaryotic, biogenesis, subunit trnas, subunits, large, nucleolus, factors, 40, synthetase, free modification, rna, depletion, eif-2, initiator, 40s, ef-3 anticodon, maturation 18s, eif2, mature, eif4e, synthetases aminoacylation, snornas, assembly, eif4g, elongation |
Dirichlet Forest Prior on topics
Andrzejewski et al, ICML 2009

Contribution

- Must-Link: tie words together (e.g., synonyms)
- Cannot-Link: keep words apart (e.g., antonyms)
Must-Link (college, school)
Inspired by constrained clustering (Basu et al, 2008)

- $\forall t$, we want $P(\text{college}|t) \approx P(\text{school}|t)$
- Cannot be encoded by Dirichlet $\rightarrow$ Dirichlet Tree
Must-Link \((college, school)\)

Inspired by constrained clustering (Basu et al, 2008)

- \(\forall t, \text{ we want } P(college|t) \approx P(school|t)\)
- Cannot be encoded by Dirichlet \(\rightarrow\) Dirichlet Tree
Cannot-Link \((school, cancer)\)

Inspired by constrained clustering (Basu et al., 2008)

- No topic-word multinomial \(\phi_t = P(w|t)\) should have both:
  - High probability \(P(school|t)\)
  - High probability \(P(cancer|t)\)
- Requires **mixture** of Dirichlet Trees (Dirichlet Forest)
Cannot-Link *(school, cancer)*

Inspired by constrained clustering (Basu et al., 2008)

- No topic-word multinomial $\phi_t = P(w|t)$ should have both:
  - High probability $P(school|t)$
  - High probability $P(cancer|t)$
- Requires **mixture** of Dirichlet Trees (Dirichlet Forest)
1 Topic modeling
   - Latent Dirichlet Allocation (LDA)
   - Issues
   - Related work

2 Preliminary work

3 New work
   - LogicLDA
   - Biological text mining

4 Conclusion
   - Discussion
   - Future work
New work: LogicLDA
(submitted to NIPS 2010)

Generalizes preliminary work

- Special topics which only appear in certain documents ($\Delta$LDA)
- Influence latent topic $z_i$ of tokens (topic-in-set)
- Must-Link and Cannot-Link between words (Dirichlet Forest)

New types of domain knowledge

- Arbitrary side information
- Relational constraints among $z_i$
- Generalize other LDA variants
New work: LogicLDA
(submitted to NIPS 2010)

Generalizes preliminary work

- Special topics which only appear in certain documents ($\Delta LDA$)
- Influence latent topic $z_i$ of tokens (topic-in-set)
- Must-Link and Cannot-Link between words (Dirichlet Forest)

New types of domain knowledge

- Arbitrary side information
- Relational constraints among $z_i$
- Generalize other LDA variants
General domain knowledge

- Express domain knowledge as first-order logic (FOL)
- Weighted knowledge base (KB) of rules
- Learned topics influenced by both
  - Word-document statistics (as in LDA)
  - Domain knowledge rules
General domain knowledge

- Express domain knowledge as first-order logic (FOL)
- Weighted knowledge base (KB) of rules
- Learned topics influenced by both
  - Word-document statistics (as in LDA)
  - Domain knowledge rules
General domain knowledge

- Express domain knowledge as first-order logic (FOL)
- Weighted knowledge base (KB) of rules
- Learned topics influenced by both
  - Word-document statistics (as in LDA)
  - Domain knowledge rules
Express domain knowledge as first-order logic (FOL)
Weighted knowledge base (KB) of rules
Learned topics influenced by both
- Word-document statistics (as in LDA)
- Domain knowledge rules
General domain knowledge

- Express domain knowledge as first-order logic (FOL)
- Weighted knowledge base (KB) of rules
- Learned topics influenced by both
  - Word-document statistics (as in LDA)
  - Domain knowledge rules
## Converting LDA variables to logic

<table>
<thead>
<tr>
<th>Value</th>
<th>Logical Predicate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(z_i = t)</td>
<td>(Z(i, t))</td>
<td>Latent topic</td>
</tr>
<tr>
<td>(w_i = v)</td>
<td>(W(i, v))</td>
<td>Observed word</td>
</tr>
<tr>
<td>(d_i = j)</td>
<td>(D(i, j))</td>
<td>Observed document</td>
</tr>
</tbody>
</table>
## Converting LDA variables to logic

<table>
<thead>
<tr>
<th>Value</th>
<th>Logical Predicate</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z_i = t )</td>
<td>( z(i, t) )</td>
<td>Latent topic</td>
</tr>
<tr>
<td>( w_i = v )</td>
<td>( \mathcal{W}(i, v) )</td>
<td>Observed word</td>
</tr>
<tr>
<td>( d_i = j )</td>
<td>( \mathcal{D}(i, j) )</td>
<td>Observed document</td>
</tr>
</tbody>
</table>

### LDA

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{HasLabel}(j, \ell) )</td>
<td>Document label</td>
</tr>
<tr>
<td>( s(i, k) )</td>
<td>Observed sentence</td>
</tr>
</tbody>
</table>
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{(\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L)\}$
- Rule $\psi_k : \forall i \ W(i, \text{endothelium}) \Rightarrow Z(i, 3)$
- Weight $\lambda_k > 0$ (“strength” of rule)

**Example $KB$**

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$W(i, \text{embryo}) \Rightarrow Z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i, j$</td>
<td>$D(i, j) \land \text{HasLabel}(j, +) \Rightarrow \neg Z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify “contradictory” domain knowledge.
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{ (\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L) \}$
- Rule $\psi_k : \forall i \ W(i, \text{endothelium}) \Rightarrow Z(i, 3)$
- Weight $\lambda_k > 0$ ("strength" of rule)

Example $KB$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$W(i, \text{embryo}) \Rightarrow Z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i,j$</td>
<td>$D(i,j) \land \text{HasLabel}(j, +) \Rightarrow \neg Z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify "contradictory" domain knowledge
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{(\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L)\}$
- Rule $\psi_k : \forall i \, \overline{w}(i, \text{endothelium}) \Rightarrow z(i, 3)$
- Weight $\lambda_k > 0$ ("strength" of rule)

Example $KB$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$w(i, \text{embryo}) \Rightarrow z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i, j$</td>
<td>$d(i, j) \land \text{HasLabel},(j, +) \Rightarrow \neg z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify "contradictory" domain knowledge
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{(\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L)\}$
- Rule $\psi_k : \forall i \ W(i, \text{endothelium}) \Rightarrow Z(i, 3)$
- Weight $\lambda_k > 0$ ("strength" of rule)

Example $KB$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$W(i, \text{embryo}) \Rightarrow Z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i, j$</td>
<td>$D(i, j) \land \text{HasLabel}(j, +) \Rightarrow \neg Z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify "contradictory" domain knowledge
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{ (\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L) \}$
- Rule $\psi_k : \forall i \ W(i, \text{endothelium}) \Rightarrow Z(i, 3)$
- Weight $\lambda_k > 0$ ("strength" of rule)

**Example $KB$**

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$W(i, \text{embryo}) \Rightarrow Z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i, j$</td>
<td>$D(i, j) \land \text{HasLabel}(j, +) \Rightarrow \neg Z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify "contradictory" domain knowledge
Domain knowledge in FOL

- CNF Knowledge Base $KB = \{ (\lambda_1, \psi_1), \ldots, (\lambda_L, \psi_L) \}$
- Rule $\psi_k : \forall i \ W(i, \text{endothelium}) \Rightarrow Z(i, 3)$
- Weight $\lambda_k > 0$ (“strength” of rule)

**Example $KB$**

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i$</td>
<td>$W(i, \text{embryo}) \Rightarrow Z(i, 3)$</td>
</tr>
<tr>
<td>500</td>
<td>$i, j$</td>
<td>$D(i, j) \land \text{HasLabel}(j, +) \Rightarrow \neg Z(i, 3)$</td>
</tr>
</tbody>
</table>

Can specify “contradictory” domain knowledge
## Example Cannot-Link rule $\psi_{CL}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i, j, t$</td>
<td>$W(i, \text{neural}) \land W(j, \text{disorder}) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t)$</td>
</tr>
</tbody>
</table>

- $G(\psi_{CL}) = \text{set of ground formulas } g \text{ for every } (i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$

- Each $g \in G(\psi_{CL})$ associated with $\lambda 1_g(z)$ term (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
Example Cannot-Link rule $\psi_{CL}$

\[
\begin{array}{c|c|c}
\lambda_k & \forall & \psi_k \\
\hline
5 & i, j, t & \\neural(i) \land \\disorder(j) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t) \\
\end{array}
\]

- $G(\psi_{CL}) = \text{set of ground formulas } g \text{ for every } (i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$

- Each $g \in G(\psi_{CL})$ associated with $\lambda \mathbb{1}_g(z)$ term
  (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
### Adding logic to LDA via propositionalization

**Example Cannot-Link rule** $\psi_{\text{CL}}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i, j, t$</td>
<td>$W(i, \text{neural}) \land W(j, \text{disorder}) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t)$</td>
</tr>
</tbody>
</table>

- $G(\psi_{\text{CL}})$ = set of ground formulas $g$ for every $(i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$

- Each $g \in G(\psi_{\text{CL}})$ associated with $\lambda g(z)$ term
  (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
Adding logic to LDA via propositionalization

Example Cannot-Link rule $\psi_{CL}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5$</td>
<td>$i, j, t$</td>
<td>$\mathbb{w}(i, \text{neural}) \land \mathbb{w}(j, \text{disorder}) \Rightarrow \neg \mathbb{z}(i, t) \lor \neg \mathbb{z}(j, t)$</td>
</tr>
</tbody>
</table>

- $G(\psi_{CL}) = \text{set of ground formulas } g \text{ for every } (i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$

- Each $g \in G(\psi_{CL})$ associated with $\lambda \mathbb{1}_g(z)$ term
  (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
Adding logic to LDA via propositionalization

Example Cannot-Link rule $\psi_{CL}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i, j, t$</td>
<td>$\neg W(i, \text{neural}) \land \neg W(j, \text{disorder}) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t)$</td>
</tr>
</tbody>
</table>

- $G(\psi_{CL}) =$ set of ground formulas $g$ for every $(i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$
- Each $g \in G(\psi_{CL})$ associated with $\lambda_1 g(z)$ term (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
Adding logic to LDA via propositionalization

Example Cannot-Link rule $\psi_{CL}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 $i, j, t$</td>
<td>$\bar{w}(i, \text{neural}) \wedge \bar{w}(j, \text{disorder})$</td>
<td>$\Rightarrow \neg Z(i, t) \vee \neg Z(j, t)$</td>
</tr>
</tbody>
</table>

- $G(\psi_{CL}) = \text{set of ground formulas } g \text{ for every } (i, j, t)$
  - $i \in \{1, 2, \ldots, N\}$
  - $j \in \{1, 2, \ldots, N\}$
  - $t \in \{1, 2, \ldots, T\}$

- Each $g \in G(\psi_{CL})$ associated with $\lambda \mathbb{1}_g(z)$ term
  (Markov Logic Networks (MLNs), Richardson & Domingos, 2006)
LDA

\[ P \propto \left( \prod_{t=1}^{T} p(\phi_t | \beta) \right) \left( \prod_{j=1}^{D} p(\theta_j | \alpha) \right) \left( \prod_{i=1}^{N} \phi_{z_i}(w_i) \theta_{d_i}(Z_i) \right) \]
LDA

\[ P \propto \left( \prod_{t}^{T} p(\phi_t | \beta) \right) \left( \prod_{j}^{D} p(\theta_j | \alpha) \right) \left( \prod_{i}^{N} \phi_{z_i}(w_i) \theta_{d_i}(Z_i) \right) \]
\[ P \propto \left( \prod_{t} p(\phi_t | \beta) \right) \left( \prod_{j} p(\theta_j | \alpha) \right) \left( \prod_{i} \phi_{z_i}(w_i)\theta_{d_i}(z_i) \right) \times \exp \left[ \sum_{l} \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1} g(z, w, d, o) \right] \]
LogicLDA MAP inference

Find most probable \((z, \phi, \theta)\)

\[
Q(z, \phi, \theta) = \sum_{t}^{T} \log p(\phi_t | \beta) + \sum_{j}^{D} \log p(\theta_j | \alpha) \\
+ \sum_{i}^{N} \log \phi_{z_i}(w_i) \theta_{d_i}(z_i) + \sum_{l}^{L} \sum_{g \in G(\psi_l)} \lambda_{l} \mathbb{1}_{g}(z, w, d, o)
\]

- LDA terms
- Logic terms
LogicLDA MAP inference

Find most probable \((z, \phi, \theta)\)

\[
Q(z, \phi, \theta) = \sum_t \log p(\phi_t|\beta) + \sum_j \log p(\theta_j|\alpha) + \sum_i \log \phi_{zi}(w_i)\theta_{di}(z_i) + \sum_l \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z, w, d, o)
\]

- LDA terms
- Logic terms
LogicLDA MAP inference

Find most probable \((z, \phi, \theta)\)

\[
Q(z, \phi, \theta) = \sum_t T \log p(\phi_t | \beta) + \sum_j D \log p(\theta_j | \alpha) + \sum_i N \log \phi_{zi}(w_i) \theta_{di}(z_i) + \sum_l \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z, w, d, o)
\]

- LDA terms
- Logic terms
Inference techniques

1. Collapsed Gibbs (CGS)
2. LDA followed by MaxWalkSAT (MWS)
3. MaxWalkSAT with LDA (M+L)
4. Mirror descent (Mir)
Inference techniques

1. Collapsed Gibbs (CGS)
2. LDA followed by MaxWalkSAT (MWS)
3. MaxWalkSAT with LDA (M+L)
4. Mirror descent (Mir)
Inference techniques

1. Collapsed Gibbs (CGS)
2. LDA followed by MaxWalkSAT (MWS)
3. MaxWalkSAT with LDA (M+L)
4. Mirror descent (Mir)
Inference techniques

1. Collapsed Gibbs (CGS)
2. LDA followed by MaxWalkSAT (MWS)
3. MaxWalkSAT with LDA (M+L)
4. Mirror descent (Mir)
Inference 1: Collapsed Gibbs Sampling (CGS)

- Combine collapsed Gibbs for LDA and MLN
  - For each $z_i$, must consider all affected ground formulas $g$
  - Select the sample which maximizes the log-posterior $Q(z, \phi, \theta)$

Issues
- Must consider all ground $g$
- Not intended to maximize $Q$
- Slow mixing
Inference 1: Collapsed Gibbs Sampling (CGS)

- Combine collapsed Gibbs for LDA and MLN
- For each $z_i$, must consider all affected ground formulas $g$
  - Select the sample which maximizes the log-posterior $Q(z, \phi, \theta)$

Issues
- Must consider all ground $g$
- Not intended to maximize $Q$
- Slow mixing
Inference 1: Collapsed Gibbs Sampling (CGS)

- Combine collapsed Gibbs for LDA and MLN
- For each $z_i$, must consider all affected ground formulas $g$
- Select the sample which maximizes the log-posterior $Q(z, \phi, \theta)$

**Issues**
- Must consider all ground $g$
- Not intended to maximize $Q$
- Slow mixing
Inference 1: Collapsed Gibbs Sampling (CGS)

- Combine collapsed Gibbs for LDA and MLN
- For each $z_i$, must consider all affected ground formulas $g$
- Select the sample which maximizes the log-posterior $Q(z, \phi, \theta)$

**Issues**

- Must consider all ground $g$
- Not intended to maximize $Q$
- Slow mixing
Inference 1: Collapsed Gibbs Sampling (CGS)

- Combine collapsed Gibbs for LDA and MLN
- For each $z_i$, must consider all affected ground formulas $g$
- Select the sample which maximizes the log-posterior $Q(z, \phi, \theta)$

**Issues**

- Must consider all ground $g$
- Not intended to maximize $Q$
- Slow mixing
Inference 2: LDA then MaxWalkSAT

1. \((z, \phi, \theta) \leftarrow\) MAP inference with respect to LDA
2. \(z \leftarrow\) post-process to maximize weight of satisfied \(g\)

MaxWalkSAT (MWS) - Kautz et al, 1997
For each step, sample an unsatisfied clause
- with probability \(p\), satisfy by flipping an atom \(randomly\)
- else satisfy by flipping an atom \(greedily\)
  (with respect to \(global\) satisfied weight)

Issues
- Must consider all ground \(g\)
- May \(decrease\) full LogicLDA objective \(Q\)
  (satisfying \(KB\) may hurt LDA objective)
Inference 2: LDA then MaxWalkSAT

1. \((z, \phi, \theta) \leftarrow \text{MAP inference with respect to LDA}\)
2. \(z \leftarrow \text{post-process to maximize weight of satisfied } g\)

MaxWalkSAT (MWS) - Kautz et al, 1997

- For each step, sample an unsatisfied clause
  - with probability \(p\), satisfy by flipping an atom *randomly*
  - else satisfy by flipping an atom *greedily*
  (with respect to **global** satisfied weight)

Issues

- Must consider all ground \(g\)
- May *decrease* full LogicLDA objective \(Q\)
(satisfying \(KB\) may hurt LDA objective)
Inference 2: LDA then MaxWalkSAT

1. \((z, \phi, \theta) \leftarrow \text{MAP inference with respect to LDA}\)
2. \(z \leftarrow \text{post-process to maximize weight of satisfied } g\)

MaxWalkSAT (MWS) - Kautz et al, 1997

For each step, sample an unsatisfied clause
- with probability \(p\), satisfy by flipping an atom randomly
- else satisfy by flipping an atom greedily
  (with respect to global satisfied weight)

Issues
- Must consider all ground \(g\)
- May decrease full LogicLDA objective \(Q\)
  (satisfying \(KB\) may hurt LDA objective)
Inference 3: MaxWalkSAT with LDA (M+L)

For each step

1. \((\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)\) \quad z \text{ fixed}

2. \(z \leftarrow \arg\max_{z} Q(z, \phi, \theta)\) \quad (\phi, \theta) \text{ fixed}

Maximizing with respect to \(z\)

Also consider LDA objective on greedy step

**Issues**

- Must consider all ground \(g\)
Inference 3: MaxWalkSAT with LDA (M+L)

For each step

1. \((\phi, \theta) \leftarrow \operatorname{argmax}_{\phi, \theta} Q(z, \phi, \theta)\) \(z\) fixed

2. \(z \leftarrow \operatorname{argmax}_z Q(z, \phi, \theta)\) \((\phi, \theta)\) fixed

Maximizing with respect to \(z\)

Also consider LDA objective on greedy step

Issues

- Must consider all ground \(g\)
Inference 3: MaxWalkSAT with LDA (M+L)

For each step

1. \((\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)\) \(\text{z fixed}\)
2. \(z \leftarrow \arg\max_{z} Q(z, \phi, \theta)\) \((\phi, \theta) \text{ fixed}\)

Maximizing with respect to \(z\)

Also consider LDA objective on greedy step

Issues

- Must consider all ground \(g\)
For each step

1. $(\phi, \theta) \leftarrow \text{argmax}_{\phi, \theta} Q(z, \phi, \theta)$ \hspace{1cm} \text{z fixed}
2. $z \leftarrow \text{argmax}_z Q(z, \phi, \theta)$ \hspace{1cm} \text{(\phi, \theta) fixed}

Maximizing with respect to $z$

Also consider LDA objective on \textit{greedy} step

Issues

- Must consider all ground $g$
For each step

1. \((\phi, \theta) \leftarrow \arg \max_{\phi, \theta} Q(z, \phi, \theta)\) \(z\) fixed
2. \(z \leftarrow \arg \max_{z} Q(z, \phi, \theta)\) \((\phi, \theta)\) fixed

Maximizing with respect to \(z\)

Also consider LDA objective on greedy step

Issues

- Must consider all ground \(g\)
Combinatorial explosion in $G(\psi_k)$

**First-Order Cannot-Link rule $\psi_{CL}$**

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5$</td>
<td>$i, j, t$</td>
<td>$W(i, \text{neural}) \land W(j, \text{disorder}) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t)$</td>
</tr>
</tbody>
</table>

- This rule has a grounding $g$ for **EVERY** $(i, j, t)$
- $N = 10^6$, $T = 100 \gg$ “financial bailout” numbers (uh oh)
- Markov Logic Network tricks fail (e.g., “lifted” inference)
Combinatorial explosion in $G(\psi_k)$

First-Order Cannot-Link rule $\psi_{CL}$

\[
\begin{array}{c|c|c|c}
\lambda_k & \forall & \psi_k \\
\hline
5 & i, j, t & W(i, neural) \land W(j, disorder) \Rightarrow \neg Z(i, t) \lor \neg Z(j, t) \\
\end{array}
\]

- This rule has a grounding $g$ for EVERY $(i, j, t)$
- $N = 10^6$, $T = 100 \gg$ “financial bailout” numbers (uh oh)
- Markov Logic Network tricks fail (e.g., “lifted” inference)
Combinatorial explosion in $G(\psi_k)$

First-Order Cannot-Link rule $\psi_{CL}$

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>$i,j,t$</td>
<td>$w(i, \text{neural}) \land w(j, \text{disorder}) \Rightarrow \neg z(i, t) \lor \neg z(j, t)$</td>
</tr>
</tbody>
</table>

- This rule has a grounding $g$ for EVERY $(i, j, t)$
- $N = 10^6$, $T = 100 \gg$ “financial bailout” numbers (uh oh)
- Markov Logic Network tricks fail (e.g., “lifted” inference)
\[ \lambda_k \quad \forall \quad \psi_k \]
\[ \overline{5 \quad i \quad \overline{W(i, apple) \Rightarrow Z(i, 3)}} \]

... apple orange banana apple apple ...

... 17 18 19 20 21 ...
Ignore trivial rule groundings
Shavlik & Natarajan, 2009

\[
\lambda_k \quad \forall \quad \psi_k \\
\hline
5 \quad i \quad W(i, \text{apple}) \Rightarrow Z(i, 3)
\]

HARD

... apple orange banana apple apple ...
... 17 18 19 20 21 ...
Ignore trivial rule groundings
Shavlik & Natarajan, 2009

\[ \lambda_k \ \forall \ \psi_k \]
\[ 5 \ \ i \ \ \omega(i, \text{apple}) \Rightarrow \Xi(i, 3) \]
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \text{argmax}_{\phi, \theta} Q(z, \phi, \theta)\) \quad \text{\texttt{z fixed}}

2. \(z \leftarrow \text{argmax}_{z} Q(z, \phi, \theta)\) \quad (\phi, \theta) \text{ fixed}
   - \(z \setminus z_{KB} \leftarrow \text{argmax} \text{ with respect to } (\phi, \theta)\)
   - \(z_{KB} \leftarrow \text{mirror descent}\)

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed \(z\) to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)\) \hspace{1cm} z fixed

2. \(z \leftarrow \arg\max_{z} Q(z, \phi, \theta)\) \hspace{1cm} (\phi, \theta) fixed
   - \(z \setminus z_{KB} \leftarrow \arg\max\) with respect to \((\phi, \theta)\)
   - \(z_{KB} \leftarrow\) mirror descent

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed \(z\) to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. $(\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)$ \textbf{z fixed}

2. $z \leftarrow \arg\max_{z} Q(z, \phi, \theta)$ \textbf{(}$(\phi, \theta)$ fixed\textbf{)}
   - $z \setminus z_{KB} \leftarrow \arg\max$ with respect to $(\phi, \theta)$
   - $z_{KB} \leftarrow$ mirror descent

Scalable approach to optimize $z_{KB}$

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed $z$ to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. $((\phi, \theta) \leftarrow \text{argmax } Q(z, \phi, \theta)_{\phi,\theta})$ \text{ z fixed}

2. $z \leftarrow \text{argmax } Q(z, \phi, \theta)$ \text{ (\phi, \theta) fixed}
   - $z \setminus z_{KB} \leftarrow \text{argmax with respect to } (\phi, \theta)$
   - $z_{KB} \leftarrow \text{mirror descent}$

Scalable approach to optimize $z_{KB}$

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed $z$ to recover final assignment

Andrzejewski (UW-Madison)  
Incorporating Domain Knowledge

Final defense 36 / 63
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)\) 
   \[z\] fixed

2. \(z \leftarrow \arg\max_{z} Q(z, \phi, \theta)\) 
   \((\phi, \theta)\) fixed
      - \(z \setminus z_{KB} \leftarrow \arg\max\) with respect to \((\phi, \theta)\)
      - \(z_{KB} \leftarrow\) mirror descent

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed \(z\) to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \text{argmax}_{\phi, \theta} Q(z, \phi, \theta)\) \(z\) fixed

2. \(z \leftarrow \text{argmax}_{z} Q(z, \phi, \theta)\) \((\phi, \theta)\) fixed

- \(z \setminus z_{KB} \leftarrow \text{argmax with respect to} (\phi, \theta)\)
- \(z_{KB} \leftarrow \text{mirror descent}\)

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed \(z\) to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \arg\max_{\phi, \theta} Q(z, \phi, \theta)\)  \hspace{1cm} z \text{ fixed}

2. \(z \leftarrow \arg\max_{z} Q(z, \phi, \theta)\)  \hspace{1cm} (\phi, \theta) \text{ fixed}

   - \(z \setminus z_{KB} \leftarrow \arg\max\) with respect to \((\phi, \theta)\)
   - \(z_{KB} \leftarrow \text{mirror descent}\)

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous

2. Optimize relaxed problem with stochastic gradient descent

3. Round relaxed \(z\) to recover final assignment
Inference 4: Mirror descent (Mir)

For each step

1. \((\phi, \theta) \leftarrow \text{argmax}_{\phi, \theta} Q(z, \phi, \theta)\) \(z\) fixed

2. \(z \leftarrow \text{argmax}_{z} Q(z, \phi, \theta)\) \((\phi, \theta)\) fixed
   - \(z \setminus z_{KB} \leftarrow \text{argmax}\) with respect to \((\phi, \theta)\)
   - \(z_{KB} \leftarrow \text{mirror descent}\)

Scalable approach to optimize \(z_{KB}\)

1. Relax discrete problem to continuous
2. Optimize relaxed problem with stochastic gradient descent
3. Round relaxed \(z\) to recover final assignment
Represent \( \mathbf{1} \) as a polynomial

\[ g = z(i, 1) \lor \neg z(j, 2), \text{ and } t \in \{1, 2, 3\} \]

1. Take complement \( \neg g \)
2. Remove negations \( (\neg g)_+ \)
3. Numeric \( z_{it} \in \{0, 1\} \)
4. Polynomial \( \mathbb{1}_g(z) \)
5. Relax discrete \( z_{it} \)

\[ \neg z(i, 1) \land z(j, 2) \]

\[ (z(i, 2) \lor z(i, 3)) \land z(j, 2) \]

\[ (z_{i2} + z_{i3})z_{j2} \]

\[ 1 - (z_{i2} + z_{i3})z_{j2} \]

\[ z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1] \]

\[ \mathbb{1}_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i,t) \in (\neg g)_+} z_{it} \right) \]
Represent 1 as a polynomial

\[ g = z(i, 1) \lor \neg z(j, 2), \text{ and } t \in \{1, 2, 3\} \]

1. Take complement \( \neg g \)

2. Remove negations \((\neg g)_+\)

3. Numeric \( z_{it} \in \{0, 1\} \)

4. Polynomial \( \mathbb{1}_g(z) \)

5. Relax discrete \( z_{it} \)

\[ \mathbb{1}_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i,t) \in (\neg g)_+} z_{it} \right) \]

\[ \neg z(i, 1) \land z(j, 2) \]

\[ (z(i, 2) \lor z(i, 3)) \land z(j, 2) \]

\[ (z_{i2} + z_{i3})z_{j2} \]

\[ 1 - (z_{i2} + z_{i3})z_{j2} \]

\[ z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1] \]
Represent 1 as a polynomial

\[ g = \overline{z}(i, 1) \lor \overline{z}(j, 2), \text{ and } t \in \{1, 2, 3\} \]

1. Take complement \( \neg g \)
2. Remove negations \((\neg g)_+\)
3. Numeric \(z_{it} \in \{0, 1\}\)
4. Polynomial \(1_g(z)\)
5. Relax discrete \(z_{it}\)

\[
1_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i,t) \in (\neg g)_+} z_{it} \right)
\]

\[
\neg z(i, 1) \land z(j, 2)
\]

\[
(z(i, 2) \lor z(i, 3)) \land z(j, 2)
\]

\[
(z_{i2} + z_{i3})z_{j2}
\]

\[
1 - (z_{i2} + z_{i3})z_{j2}
\]

\[
z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]
\]
Represent 1 as a polynomial

\[ g = z(i, 1) \lor \neg z(j, 2), \text{ and } t \in \{1, 2, 3\} \]

- Take complement \( \neg g \)
- Remove negations \((\neg g)_+\)
- Numeric \( z_{it} \in \{0, 1\} \)
- Polynomial \( 1_g(z) \)
- Relax discrete \( z_{it} \)

\[ 1_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i,t) \in (\neg g)_+} z_{it} \right) \]
Represent $\mathbf{1}$ as a polynomial

$$g = Z(i, 1) \lor \neg Z(j, 2), \text{ and } t \in \{1, 2, 3\}$$

1. Take complement $\neg g$
2. Remove negations $(\neg g)_+$
3. Numeric $z_{it} \in \{0, 1\}$
4. Polynomial $\mathbb{1}_g(z)$
5. Relax discrete $z_{it}$

$$\mathbb{1}_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{Z(i,t) \in (\neg g)_+} z_{it} \right)$$

$$\neg Z(i, 1) \land Z(j, 2)$$

$$(Z(i, 2) \lor Z(i, 3)) \land Z(j, 2)$$

$$(Z_{i2} + Z_{i3})Z_{j2}$$

$$1 - (Z_{i2} + Z_{i3})Z_{j2}$$

$$Z_{it} \in \{0, 1\} \rightarrow Z_{it} \in [0, 1]$$
Represent 1 as a polynomial

\[ g = \mathbb{Z}(i, 1) \lor \neg \mathbb{Z}(j, 2), \text{ and } t \in \{1, 2, 3\} \]

1. Take complement \( \neg g \)
2. Remove negations \((\neg g)_+\)
3. Numeric \( z_{it} \in \{0, 1\} \)
4. Polynomial \( \mathbb{I}_g(z) \)
5. Relax discrete \( z_{it} \)

\[ \neg \mathbb{Z}(i, 1) \land \mathbb{Z}(j, 2) \]
\[ (\mathbb{Z}(i, 2) \lor \mathbb{Z}(i, 3)) \land \mathbb{Z}(j, 2) \]
\[ (z_{i2} + z_{i3})z_{j2} \]
\[ 1 - (z_{i2} + z_{i3})z_{j2} \]
\[ z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1] \]

\[ \mathbb{I}_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i,t) \in (\neg g)_+} z_{it} \right) \]
Represent 1 as a polynomial

\[ g = z(i, 1) \lor \neg z(j, 2), \text{ and } t \in \{1, 2, 3\} \]

1. Take complement \( \neg g \)
2. Remove negations \((\neg g)_+\)
3. Numeric \(z_{it} \in \{0, 1\}\)
4. Polynomial \(1_g(z)\)
5. Relax discrete \(z_{it}\)

\[ 1_g(z) = 1 - \prod_{g_i \neq \emptyset} \left( \sum_{z(i, t) \in (-g_i)_+} z_{it} \right) \]
Technical contribution: scalable $z_{KB}$ inference

$$\text{argmax}_{z} \sum_{i}^{N} \sum_{t}^{T} z_{it} \log \phi_{z_{i}}(w_{i}) \theta_{d_{i}}(z_{i}) + \sum_{l}^{L} \sum_{g \in G(\psi_{l})} \lambda_{l} \mathbb{1}_{g}(z)$$

1. **Continuous relaxation**
   - $z_{i} = t$
   - Represent indicator function $\mathbb{1}_{g}(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$

2. **Stochastic gradient - sample a term from objective function $Q$**
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. **Entropic Mirror Descent** (Beck & Teboulle, 2003)
   \[
   Z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla_{z_{it'}} f)}
   \]

4. **Recover discrete $z$:** $z_{i} = \text{argmax}_{t} z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_{z} \sum_{i}^{N} \sum_{t}^{T} z_{it} \log \phi_{z_{i}}(w_{i}) \theta_{d_{i}}(z_{i}) + \sum_{l}^{L} \sum_{g \in G(\psi_{l})} \lambda_{l} \mathbb{1}_{g}(z)$$

1. Continuous relaxation
   - $z_{i} = t \rightarrow z_{it} \in \{0, 1\}$
   - Represent indicator function $\mathbb{1}_{g}(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   $$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla_{z_{it'}} f)}$$

4. Recover discrete $z$: $z_{i} = \arg\max_{t} z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_i^{N} \sum_t^{T} z_{it} \log \phi_{Zi}(w_i) \theta_d(z_i) + \sum_l^{L} \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1 Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_g(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$

2 Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3 Entropic Mirror Descent (Beck & Teboulle, 2003)

$$z_{it} \leftarrow \frac{z_{it} \exp (\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp (\eta \nabla_{z_{it'}} f)}$$

4 Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it} \log \phi_{z_i}(w_i) \theta_{d_i}(z_i) + \sum_{l=1}^{L} \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$  
   - Represent indicator function $\mathbb{1}_g(z)$ as polynomial in $z_{it}$  
   - Can calculate $\nabla Q$

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$  
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
$$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla z_{it} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla z_{it'} f)}$$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_i^N \sum_t^T z_{it} \log \phi_i(w_t) \theta_i(z_i) + \sum_l^L \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_g(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   - $z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_t z_{it'} \exp(\eta \nabla_{z_{it'}} f)}$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $\mathbf{z}_{KB}$ inference

$$\arg\max_{\mathbf{z}} \sum_{i}^{N} \sum_{t}^{T} z_{it} \log \phi_{z_i}(\mathbf{w}_i) \theta_{d_i}(z_i) + \sum_{l}^{L} \sum_{g \in G(\psi_l)} \lambda_{l} \mathbb{1}_{g}(\mathbf{z})$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_{g}(\mathbf{z})$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$...but $G(\psi_k)$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   $$z_{it} \leftarrow \frac{z_{it} \exp (\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp (\eta \nabla_{z_{it'}} f)}$$

4. Recover discrete $\mathbf{z}$: $z_i = \arg\max_{t} z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $\mathbf{z}_{KB}$ inference

$$\arg\max_{\mathbf{z}} \sum_{i}^{N} \sum_{t}^{T} z_{it} \log \phi_{z_{i}}(w_{i}) \theta_{d_{i}}(z_{i}) + \sum_{l}^{L} \sum_{g \in G(\psi_{l})} \lambda_{l} \mathbb{1}_{g}(\mathbf{z})$$

1. Continuous relaxation
   - $z_{i} = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_{g}(\mathbf{z})$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$ ... but $G(\psi_{k})$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   $$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla_{z_{it'}} f)}$$

4. Recover discrete $\mathbf{z}$: $z_{i} = \arg\max_{t} z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_i^{N} \sum_t^{T} z_{it} \log \phi_z(w_i) \theta_d(z_i) + \sum_l^L \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_g(z)$ as *polynomial* in $z_{it}$
   - Can calculate $\nabla Q$...but $G(\psi_k)$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)

   $$Z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla_{z_{it'}} f)}$$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_i \sum_t z_{it} \log \phi_{z_i}(w_i) \theta_d(z_i) + \sum_l \sum_{g \in G(\psi_l)} \lambda_l 1_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $1_g(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$...but $G(\psi_k)$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)

   $$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla_{z_{it}} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla_{z_{it'}} f)}$$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_{i}^{N} \sum_{t}^{T} z_{it} \log \phi_z(w_i) \theta_d(z_i) + \sum_{l}^{L} \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_g(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$...but $G(\psi_k)$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   $$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla z_{it} f)}{\sum_{t'} z_{it'} \exp(\eta \nabla z_{it'} f)}$$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
Technical contribution: scalable $z_{KB}$ inference

$$\arg\max_z \sum_i^N \sum_t^T z_{it} \log \phi_{z_i}(w_i) \theta_{d_i}(z_i) + \sum_l^L \sum_{g \in G(\psi_l)} \lambda_l \mathbb{1}_g(z)$$

1. Continuous relaxation
   - $z_i = t \rightarrow z_{it} \in \{0, 1\} \rightarrow z_{it} \in [0, 1]$
   - Represent indicator function $\mathbb{1}_g(z)$ as polynomial in $z_{it}$
   - Can calculate $\nabla Q$...but $G(\psi_k)$ may be very large

2. Stochastic gradient - sample a term from objective function $Q$
   - Logic: single ground formula $g$
   - LDA: single corpus index $i$

3. Entropic Mirror Descent (Beck & Teboulle, 2003)
   $$z_{it} \leftarrow \frac{z_{it} \exp(\eta \nabla z_{it} f)}{\sum_t' z_{it'} \exp(\eta \nabla z_{it'} f)}$$

4. Recover discrete $z$: $z_i = \arg\max_t z_{it}$ for $i = 1, \ldots, N$
LogicLDA experiments

- Synthetic and benchmark corpora (e.g., 20news)
- Simple KBs: at most a few rules each
- Questions
  - Can we influence the learned topics?
  - How do the inference techniques perform?
LogicLDA experiments

- Synthetic and benchmark corpora (e.g., 20news)
- Simple KBs: at most a few rules each

Questions
  - Can we influence the learned topics?
  - How do the inference techniques perform?
LogicLDA experiments

- Synthetic and benchmark corpora (e.g., 20news)
- Simple KBs: at most a few rules each
- Questions
  - Can we influence the learned topics?
  - How do the inference techniques perform?
LogicLDA experiments

- Synthetic and benchmark corpora (e.g., 20news)
- Simple KBs: at most a few rules each
- Questions
  - Can we influence the learned topics?
  - How do the inference techniques perform?
LogicLDA experiments

- Synthetic and benchmark corpora (e.g., 20news)
- Simple KBs: at most a few rules each
- Questions
  - Can we influence the learned topics?
  - How do the inference techniques perform?
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

Logic Cannot-Link (movie,film)

\[ \overline{w}(i, \text{movie}) \land \overline{w}(j, \text{film}) \Rightarrow (\neg z(i, t) \lor \neg z(j, t)) \]
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

Logic Cannot-Link (movie,film)

\[
\bar{w}(i, \text{movie}) \land \bar{w}(j, \text{film}) \Rightarrow (\neg \bar{z}(i, t) \lor \neg \bar{z}(j, t))
\]
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

Logic Cannot-Link (movie,film)

\[ w(i, \text{movie}) \land w(j, \text{film}) \Rightarrow (\neg z(i, t) \lor \neg z(j, t)) \]
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

Logic Cannot-Link (movie,film)

\[ \overline{w}(i, \text{movie}) \land \overline{w}(j, \text{film}) \Rightarrow (\neg \overline{z}(i, t) \lor \neg \overline{z}(j, t)) \]
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

Logic Cannot-Link (movie,film)

\[ w(i, \text{movie}) \land w(j, \text{film}) \Rightarrow (\neg z(i, t) \lor \neg z(j, t)) \]
Qualitative results: movie review corpus (Pol)

- Movie review corpus (Pang & Lee, 2005)
  - “movie” overrepresented in negative reviews
  - “film” overrepresented in positive reviews
- Goal: investigate this difference with topic models

**Logic Cannot-Link (movie,film)**

\[
\bar{w}(i, \text{movie}) \land \bar{w}(j, \text{film}) \Rightarrow (\neg z(i, t) \lor \neg z(j, t))
\]
Standard LDA topic
Inference results - maximize log-posterior $Q$
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th></th>
<th>LogicLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mir</td>
<td>8.64</td>
</tr>
<tr>
<td>M+L</td>
<td>9.08</td>
</tr>
<tr>
<td>MWS</td>
<td>8.19</td>
</tr>
<tr>
<td>CGS</td>
<td>9.08</td>
</tr>
<tr>
<td>LDA</td>
<td>3.83</td>
</tr>
<tr>
<td>Alchemy</td>
<td>1.43</td>
</tr>
<tr>
<td>S1</td>
<td>8.64</td>
</tr>
<tr>
<td>S2</td>
<td>3.44</td>
</tr>
<tr>
<td>Mac</td>
<td>3.35</td>
</tr>
<tr>
<td>Comp</td>
<td>2.48</td>
</tr>
<tr>
<td>Con</td>
<td>18.52</td>
</tr>
<tr>
<td>Pol</td>
<td>9.63</td>
</tr>
<tr>
<td>HDG</td>
<td>116.80</td>
</tr>
</tbody>
</table>

Incorporating Domain Knowledge

Final defense 42 / 63
Inference results - maximize log-posterior $Q$
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th></th>
<th>LogicLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mir</td>
</tr>
<tr>
<td>S1</td>
<td>8.64</td>
</tr>
<tr>
<td>S2</td>
<td>3.44</td>
</tr>
<tr>
<td>Mac</td>
<td>3.35</td>
</tr>
<tr>
<td>Comp</td>
<td>2.48</td>
</tr>
<tr>
<td>Con</td>
<td>18.52</td>
</tr>
<tr>
<td>Pol</td>
<td>9.63</td>
</tr>
<tr>
<td>HDG</td>
<td>116.80</td>
</tr>
</tbody>
</table>
Inference results - maximize log-posterior $Q$
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th>LogicLDA</th>
<th>Mir</th>
<th>M+L</th>
<th>MWS</th>
<th>CGS</th>
<th>LDA</th>
<th>Alchemy</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>8.64</td>
<td>9.08</td>
<td>8.19</td>
<td>9.08</td>
<td>3.83</td>
<td>1.43</td>
</tr>
<tr>
<td>S2</td>
<td>3.44</td>
<td>3.49</td>
<td>3.49</td>
<td>3.47</td>
<td>2.17</td>
<td>2.72</td>
</tr>
<tr>
<td>Mac</td>
<td>3.35</td>
<td>3.36</td>
<td>2.49</td>
<td>2.48</td>
<td>2.24</td>
<td>$-5.7^{NC}$</td>
</tr>
<tr>
<td>Comp</td>
<td>2.48</td>
<td>2.48</td>
<td>0.15</td>
<td>2.20</td>
<td>$-0.84$</td>
<td>$NC$</td>
</tr>
<tr>
<td>Con</td>
<td>18.52</td>
<td>19.04</td>
<td>$-3.79$</td>
<td>16.68</td>
<td>$-4.07$</td>
<td>$NC$</td>
</tr>
<tr>
<td>Pol</td>
<td>9.63</td>
<td>$NC$</td>
<td>$NC$</td>
<td>$NC$</td>
<td>9.56</td>
<td>$NC$</td>
</tr>
<tr>
<td>HDG</td>
<td>116.80</td>
<td>$NC$</td>
<td>$NC$</td>
<td>$NC$</td>
<td>64.04</td>
<td>$NC$</td>
</tr>
</tbody>
</table>
Inference results - maximize log-posterior $Q$
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th></th>
<th>LogicLDA</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mir</td>
<td>M+L</td>
<td>MWS</td>
<td>CGS</td>
<td>LDA</td>
<td>Alchemy</td>
</tr>
<tr>
<td>S1</td>
<td>8.64</td>
<td>9.08</td>
<td>8.19</td>
<td>9.08</td>
<td>3.83</td>
<td>1.43</td>
</tr>
<tr>
<td>S2</td>
<td>3.44</td>
<td>3.49</td>
<td>3.49</td>
<td>3.47</td>
<td>2.17</td>
<td>2.72</td>
</tr>
<tr>
<td>Mac</td>
<td>3.35</td>
<td>3.36</td>
<td>2.49</td>
<td>2.48</td>
<td>2.24</td>
<td>$-5.7^{NC}$</td>
</tr>
<tr>
<td>Comp</td>
<td>2.48</td>
<td>2.48</td>
<td>0.15</td>
<td>2.20</td>
<td>$-0.84$</td>
<td>NC</td>
</tr>
<tr>
<td>Con</td>
<td>18.52</td>
<td>19.04</td>
<td>$-3.79$</td>
<td>16.68</td>
<td>$-4.07$</td>
<td>NC</td>
</tr>
<tr>
<td>Pol</td>
<td>9.63</td>
<td>$NC$</td>
<td>$NC$</td>
<td>$NC$</td>
<td>$9.56$</td>
<td>NC</td>
</tr>
<tr>
<td>HDG</td>
<td>116.80</td>
<td>$NC$</td>
<td>$NC$</td>
<td>$NC$</td>
<td>64.04</td>
<td>NC</td>
</tr>
</tbody>
</table>
Inference results - maximize log-posterior $Q$
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th></th>
<th>LogicLDA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mir</td>
</tr>
<tr>
<td>S1</td>
<td>8.64</td>
</tr>
<tr>
<td>S2</td>
<td>3.44</td>
</tr>
<tr>
<td>Mac</td>
<td>3.35</td>
</tr>
<tr>
<td>Comp</td>
<td>2.48</td>
</tr>
<tr>
<td>Con</td>
<td>18.52</td>
</tr>
<tr>
<td>Pol</td>
<td>9.63</td>
</tr>
<tr>
<td>HDG</td>
<td>116.80</td>
</tr>
</tbody>
</table>
LogicLDA can encode existing LDA variants

Example: Hidden Topic Markov Model (HTMM) - Gruber et al, 2007
- Each sentence uses only one topic
- Topic transitions possible between sentences with probability $\epsilon$

FOL encoding of HTMM

$$
\lambda_k \quad \forall \quad \psi_k
\inf \quad i, j, s, t \quad S(i, s) \land S(j, s) \land Z(i, t) \Rightarrow Z(j, t)
$$
LogicLDA can encode existing LDA variants

Example: Hidden Topic Markov Model (HTMM) - Gruber et al, 2007
- Each sentence uses only one topic
- Topic transitions possible between sentences with probability $\epsilon$

FOL encoding of HTMM

<table>
<thead>
<tr>
<th>$\lambda_k$</th>
<th>$\forall$</th>
<th>$\psi_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\infty$</td>
<td>$i, j, s, t$</td>
<td>$s(i, s) \land s(j, s) \land z(i, t) \Rightarrow z(j, t)$</td>
</tr>
<tr>
<td>$-\log \epsilon$</td>
<td>$i, s, t$</td>
<td>$s(i, s) \land \neg s(i + 1, s) \land z(i, t) \Rightarrow z(i + 1, t)$</td>
</tr>
</tbody>
</table>
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$

- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$

- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
    - Dependencies among $z_i$
- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$

- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$

- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$

- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general KBs
LogicLDA summary

Contribution

- FOL domain knowledge
  - Generalize preliminary work
  - Side information
  - Dependencies among $z_i$
- Scalable inference
  - Allows us to find $z$ balancing LDA and logic
  - Necessary for general $KB$s
1. Topic modeling
   - Latent Dirichlet Allocation (LDA)
   - Issues
   - Related work

2. Preliminary work

3. New work
   - LogicLDA
   - Biological text mining

4. Conclusion
   - Discussion
   - Future work
Ron Stewart (Thomson lab) is interested in connections between
- experimentally interesting genes
- human development concepts of interest
Given “seed” terms for each concept
Do discover other related terms

<table>
<thead>
<tr>
<th>Concept</th>
<th>Provided terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>neur dendro(cytes), glia, synapse, neural crest</td>
</tr>
<tr>
<td>Embryo</td>
<td>human embryonic stem cell, inner cell mass, pluripotent</td>
</tr>
<tr>
<td>Blood</td>
<td>hematopoietic, blood, endothel(ium)</td>
</tr>
<tr>
<td>Gastrulation</td>
<td>organizer, gastru(late)</td>
</tr>
<tr>
<td>Cardiac</td>
<td>heart, ventricle, auricle, aorta</td>
</tr>
<tr>
<td>Limb</td>
<td>limb, blastema, zeugopod, autopod, stylopod</td>
</tr>
</tbody>
</table>
**Given** “seed” terms for each concept

**Do** discover other related terms

<table>
<thead>
<tr>
<th>Concept</th>
<th>Provided terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Embryo</td>
<td>neur dendro(cyte), glia, synapse, neural crest</td>
</tr>
<tr>
<td>Embryo Blood</td>
<td>human embryonic stem cell, inner cell mass, pluripotent</td>
</tr>
<tr>
<td>Blood Gastrulation</td>
<td>hematopoietic, blood, endothel(ium)</td>
</tr>
<tr>
<td>Gastrulation</td>
<td>organizer, gastru(late)</td>
</tr>
<tr>
<td>Cardiac Limb</td>
<td>heart, ventricle, auricle, aorta</td>
</tr>
<tr>
<td></td>
<td>limb, blastema, zeugopod, autopod, stylopod</td>
</tr>
</tbody>
</table>
Do standard LDA, then find topics containing seed terms in Top 50

brain system nervous neurons neuronal central development’s neural human gene disease function cortex spinal disorders developing motor cerebral glial peripheral cortical cord disorder astrocytes nerve neurological regions suggest schizophrenia including syndrome neurodegenerative mental involved retardation behavior cerebellum migration behavioral abnormal cerebellar found precursor results amyloid hippocampus sclerosis neurotrophic present
Do standard LDA, then find topics containing seed terms in Top 50

brain system nervous neurons neuronal central development’s neural human gene disease function cortex spinal disorders developing motor cerebral glial peripheral cortical cord disorder astrocytes nerve neurological regions suggest schizophrenia including syndrome neurodegenerative mental involved retardation behavior cerebellum migration behavioral abnormal cerebellar found precursor results amyloid hippocampus sclerosis neurotrophic present
Seed and $n$-gram rules

**Neural** $\rightarrow$ “synapse” $\rightarrow$ Topic 0

$W(i, \text{synapse}) \Rightarrow Z(i, 0)$

**Embryo** $\rightarrow$ “inner cell mass” $\rightarrow$ Topic 1

$W(i, \text{inner}) \land W(i + 1, \text{cell}) \land W(i + 2, \text{mass}) \Rightarrow Z(i, 1)$

$W(i − 1, \text{inner}) \land W(i, \text{cell}) \land W(i + 1, \text{mass}) \Rightarrow Z(i, 1)$

$W(i − 2, \text{inner}) \land W(i − 1, \text{cell}) \land W(i, \text{mass}) \Rightarrow Z(i, 1)$
Seed and $n$-gram rules

**Neural** $\rightarrow$ “synapse” $\rightarrow$ Topic 0

$\overline{w}(i, \text{synapse}) \Rightarrow z(i, 0)$

**Embryo** $\rightarrow$ “inner cell mass” $\rightarrow$ Topic 1

$\overline{w}(i, \text{inner}) \land \overline{w}(i + 1, \text{cell}) \land \overline{w}(i + 2, \text{mass}) \Rightarrow z(i, 1)$

$\overline{w}(i - 1, \text{inner}) \land \overline{w}(i, \text{cell}) \land \overline{w}(i + 1, \text{mass}) \Rightarrow z(i, 1)$

$\overline{w}(i - 2, \text{inner}) \land \overline{w}(i - 1, \text{cell}) \land \overline{w}(i, \text{mass}) \Rightarrow z(i, 1)$
Create new *development* Topic 6
{differentiation, maturation, develops, formation, differentiates}

Development Topic 6 *allows* each seed Topic $t$ in sentence $t \in \{0, \ldots, 5\}$

Sentence($i, i_1, \ldots, i_{S_k}$) ∧ ¬$Z(i_1, 6)$ ∧ … ∧ ¬$Z(i_{S_k}, 6)$ ⇒ ¬$Z(i, 0)$
Create new *development* Topic 6
{differentiation, maturation, develops, formation, differentiates}

Development Topic 6 allows each seed Topic $t$ in sentence $t \in \{0, \ldots, 5\}$

\[
\text{Sentence}(i, i_1, \ldots, i_{S_k}) \land \neg \exists(i_1, 6) \land \ldots \land \neg \exists(i_{S_k}, 6) \implies \neg \exists(i, 0)
\]
Sentence exclusion

- Create new *disease* Topic 7
  \{patient, disease, parasite, chronic, virus, condition, disorder, symptom\}
- Disease Topic 7 prevents each seed Topic \( t \) in sentence
  \( t \in \{0, \ldots, 5\} \)

\[ S(i, s) \land S(j, s) \land Z(i, 7) \Rightarrow \neg Z(j, 0) \]
Experimental setup

- Different KBs
  - LDA - Standard LDA
  - SEED - Seed
  - INCL - Seed + Inclusion
  - EXCL - Seed + Exclusion
  - ALL - Seed + Inclusion + Exclusion

- Labeled Top 50 words for each KB

Would knowing that this word is (statistically) associated with a gene increase your belief that the gene is related to the target concept?
Experimental setup

Different KBs
- LDA - Standard LDA
- SEED - Seed
- INCL - Seed + Inclusion
- EXCL - Seed + Exclusion
- ALL - Seed + Inclusion + Exclusion

Labeled Top 50 words for each KB

Would knowing that this word is (statistically) associated with a gene increase your belief that the gene is related to the target concept?
Experimental setup

- Different KBs
  - LDA - Standard LDA
  - SEED - Seed
  - INCL - Seed + Inclusion
  - EXCL - Seed + Exclusion
  - ALL - Seed + Inclusion + Exclusion

- Labeled Top 50 words for each KB

Would knowing that this word is (statistically) associated with a gene increase your belief that the gene is related to the target concept?
### Accuracy at Top 50 threshold

(means over 10 randomized runs)

<table>
<thead>
<tr>
<th>LogicLDA KBs</th>
<th>ALL</th>
<th>INCL</th>
<th>EXCL</th>
<th>SEED</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>0.59</td>
<td>0.57</td>
<td>0.54</td>
<td>0.54</td>
<td>0.31</td>
</tr>
<tr>
<td>Embryo</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>Blood</td>
<td>0.46</td>
<td>0.47</td>
<td>0.40</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>Gast.</td>
<td>0.18</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0.36</td>
<td>0.37</td>
<td>0.34</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Limb</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>
## Accuracy at Top 50 threshold
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th>LogicLDA KBs</th>
<th>ALL</th>
<th>INCL</th>
<th>EXCL</th>
<th>SEED</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>0.59</td>
<td>0.57</td>
<td>0.54</td>
<td>0.54</td>
<td>0.31</td>
</tr>
<tr>
<td>Embryo</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>Blood</td>
<td>0.46</td>
<td>0.47</td>
<td>0.40</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>Gast.</td>
<td>0.18</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0.36</td>
<td>0.37</td>
<td>0.34</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Limb</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Accuracy at Top 50 threshold  
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th>LogicLDA KBs</th>
<th>ALL</th>
<th>INCL</th>
<th>EXCL</th>
<th>SEED</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>0.59</td>
<td>0.57</td>
<td>0.54</td>
<td>0.54</td>
<td>0.31</td>
</tr>
<tr>
<td>Embryo</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>Blood</td>
<td>0.46</td>
<td>0.47</td>
<td>0.40</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>Gast.</td>
<td>0.18</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0.36</td>
<td>0.37</td>
<td>0.34</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Limb</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Accuracy at Top 50 threshold
(means over 10 randomized runs)

<table>
<thead>
<tr>
<th>LogicLDA KBs</th>
<th>ALL</th>
<th>INCL</th>
<th>EXCL</th>
<th>SEED</th>
<th>LDA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural</td>
<td>0.59</td>
<td>0.57</td>
<td>0.54</td>
<td>0.54</td>
<td>0.31</td>
</tr>
<tr>
<td>Embryo</td>
<td>0.24</td>
<td>0.24</td>
<td>0.23</td>
<td>0.23</td>
<td>0.07</td>
</tr>
<tr>
<td>Blood</td>
<td>0.46</td>
<td>0.47</td>
<td>0.40</td>
<td>0.39</td>
<td>0.13</td>
</tr>
<tr>
<td>Gast.</td>
<td>0.18</td>
<td>0.18</td>
<td>0.16</td>
<td>0.16</td>
<td>0.00</td>
</tr>
<tr>
<td>Cardiac</td>
<td>0.36</td>
<td>0.37</td>
<td>0.34</td>
<td>0.35</td>
<td>0.08</td>
</tr>
<tr>
<td>Limb</td>
<td>0.18</td>
<td>0.18</td>
<td>0.15</td>
<td>0.14</td>
<td>0.09</td>
</tr>
</tbody>
</table>
Precision-Recall Neural

Andrzejewski (UW-Madison)  Incorporating Domain Knowledge  Final defense
Precision-Recall Blood

![Precision-Recall Graph]

- INCL+EXCL
- INCL
- EXCL
- SEED
- LDA

Incorporating Domain Knowledge

Final defense 55 / 63
Precision-Recall Cardiac

![Precision-Recall Graph](image)

- INCL+EXCL
- INCL
- EXCL
- SEED
- LDA
Novel terms discovered

INCL and ALL for *neural*

{dendritic, forebrain, hindbrain, microglial, motoneurons, neuroblasts, neurogenesis, retinal}
Biological application

Contribution

- Apply topic modeling to real-world text mining problem
- LogicLDA allows us to outperform standard LDA
  - Exploit sentence information
  - Better quantitative performance
  - Discover novel terms related to target concept
Biological application

Contribution

- Apply topic modeling to real-world text mining problem
- LogicLDA allows us to outperform standard LDA
  - Exploit sentence information
  - Better quantitative performance
  - Discover novel terms related to target concept
Biological application

Contribution

- Apply topic modeling to real-world text mining problem
- LogicLDA allows us to outperform standard LDA
  - Exploit sentence information
  - Better quantitative performance
  - Discover novel terms related to target concept
Biological application

Contribution

- Apply topic modeling to real-world text mining problem
- LogicLDA allows us to outperform standard LDA
  - Exploit sentence information
  - Better quantitative performance
  - Discover novel terms related to target concept
Biological application

**Contribution**

- Apply topic modeling to real-world text mining problem
- LogicLDA allows us to outperform standard LDA
  - Exploit sentence information
  - Better quantitative performance
  - Discover novel terms related to target concept
1. Topic modeling
   - Latent Dirichlet Allocation (LDA)
   - Issues
   - Related work

2. Preliminary work

3. New work
   - LogicLDA
   - Biological text mining

4. Conclusion
   - Discussion
   - Future work
Latent topic models + domain knowledge

Code available http://pages.cs.wisc.edu/~andrzej/software.html

**ΔLDA (ECML 2007)**
- Restricted topics
- Statistical debugging

**Topic-in-set (SSLNLP 2009)**
- Topic assignments $z_i$
- "Seed" words

**Dirichlet Forest prior (ICML 2009)**
- Must-Link and Cannot-Link over topics $\phi$
- Composite operations (split, merge, isolate)

**LogicLDA (submitted to NIPS 2010)**
- General FOL domain knowledge
- Scalable inference methods
- Biological text mining application
# Latent topic models + domain knowledge


<table>
<thead>
<tr>
<th><strong>ΔLDA (ECML 2007)</strong></th>
<th><strong>Topic-in-set (SSLNLP 2009)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted topics</td>
<td>Topic assignments $z_i$</td>
</tr>
<tr>
<td>Statistical debugging</td>
<td>“Seed” words</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Dirichlet Forest prior (ICML 2009)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Must-Link and Cannot-Link over topics $\phi$</td>
</tr>
<tr>
<td>Composite operations (split, merge, isolate)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>LogicLDA (submitted to NIPS 2010)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>General FOL domain knowledge</td>
</tr>
<tr>
<td>Scalable inference methods</td>
</tr>
<tr>
<td>Biological text mining application</td>
</tr>
</tbody>
</table>
Latent topic models + domain knowledge

Code available http://pages.cs.wisc.edu/~andrzej/software.html

**ΔLDA (ECML 2007)**
- Restricted topics
- Statistical debugging

**Topic-in-set (SSLNLP 2009)**
- Topic assignments $z_i$
- “Seed” words

**Dirichlet Forest prior (ICML 2009)**
- Must-Link and Cannot-Link over topics $\phi$
- Composite operations (**split**, **merge**, **isolate**)

**LogicLDA (submitted to NIPS 2010)**
- General FOL domain knowledge
- Scalable inference methods
- Biological text mining application
Latent topic models + domain knowledge

Code available http://pages.cs.wisc.edu/~andrzeje/software.html

**ΔLDA (ECML 2007)**
- Restricted topics
- Statistical debugging

**Topic-in-set (SSLNLP 2009)**
- Topic assignments $z_i$
- “Seed” words

**Dirichlet Forest prior (ICML 2009)**
- Must-Link and Cannot-Link over topics $\phi$
- Composite operations (split, merge, isolate)

**LogicLDA (submitted to NIPS 2010)**
- General FOL domain knowledge
- Scalable inference methods
- Biological text mining application
Other work

- True vs computer-generated Mondrians (SPIE 2010)

- Identify wishes in text (NAACL HLT 2009)
- Incorporate “diversity” in rankings (NAACL HLT 2007)
- Find text passages relevant to biomedical queries (TREC 2006)
Other work

- True vs computer-generated Mondrians (SPIE 2010)

- Identify wishes in text (NAACL HLT 2009)
- Incorporate “diversity” in rankings (NAACL HLT 2007)
- Find text passages relevant to biomedical queries (TREC 2006)
Other work

- True vs computer-generated Mondrians (SPIE 2010)

- Identify wishes in text (NAACL HLT 2009)
- Incorporate “diversity” in rankings (NAACL HLT 2007)
- Find text passages relevant to biomedical queries (TREC 2006)
Other work

- True vs computer-generated Mondrians (SPIE 2010)
- Identify wishes in text (NAACL HLT 2009)
- Incorporate “diversity” in rankings (NAACL HLT 2007)
- Find text passages relevant to biomedical queries (TREC 2006)
Potential future directions

- **Active learning**: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics (*split* candidates)
  - Develop better “interpretability” objective function
    (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify *how* data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: `citation(d, d')`
Potential future directions

- Active learning: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics (*split* candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)
- Multimodal knowledge
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify *how* data types interact
- Domain knowledge for secondary tasks
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics
- Statistical relational learning (SRL)
  - Example: Citation($d, d'$)
Potential future directions

- Active learning: solicit user feedback which will be most “useful”
  - Identify “weak” topics (split candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)

- Multimodal knowledge
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- Domain knowledge for secondary tasks
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- Statistical relational learning (SRL)
  - Example: citation($d, d'$)
Potential future directions

- **Active learning:** *solicit* user feedback which will be most “useful”
  - Identify “weak” topics *(split* candidates)
  - Develop better “interpretability” objective function
    (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify *how* data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: *citation*(\(d, d'\))
Potential future directions

- **Active learning**: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics *(split* candidates)
  - Develop better “interpretability” objective function *(Chang et al., NIPS 2009)*

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: \( \text{citation}(d, d') \)
Potential future directions

- **Active learning**: solicit user feedback which will be most “useful”
  - Identify “weak” topics (**split** candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: \textit{citation}(d, d')
Potential future directions

- Active learning: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics (*split* candidates)
  - Develop better “interpretability” objective function
    (Chang et al., NIPS 2009)

- Multimodal knowledge
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify *how* data types interact

- Domain knowledge for secondary tasks
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- Statistical relational learning (SRL)
  - Example: *citation*(d, d')
Potential future directions

- **Active learning**: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics (*split* candidates)
  - Develop better “interpretability” objective function
    (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: $\text{citation}(d, d')$
Potential future directions

- Active learning: *solicit* user feedback which will be most “useful”
  - Identify “weak” topics (split candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)
- Multimodal knowledge
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact
- Domain knowledge for secondary tasks
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics
- Statistical relational learning (SRL)
  - Example: \texttt{citation}(d,d')
Potential future directions

- **Active learning:** solicit user feedback which will be most “useful”
  - Identify “weak” topics (split candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: \text{citation}(d, d')
Potential future directions

- **Active learning:** solicit user feedback which will be most “useful”
  - Identify “weak” topics (split candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)

- **Multimodal knowledge**
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- **Domain knowledge for secondary tasks**
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- **Statistical relational learning (SRL)**
  - Example: Citation($d, d'$)
Potential future directions

- Active learning: solicit user feedback which will be most “useful”
  - Identify “weak” topics (split candidates)
  - Develop better “interpretability” objective function (Chang et al., NIPS 2009)

- Multimodal knowledge
  - Text + Image (captions), Text + Experimental Data (biology)
  - Interaction currently “hard-coded” into graphical model
  - Allow user to specify how data types interact

- Domain knowledge for secondary tasks
  - Often topics are to be used as input to another algorithm
  - How to best use domain knowledge to learn suitable topics

- Statistical relational learning (SRL)
  - Example: Citation($d, d'$)
Acknowledgements

- Prelim and thesis committee members
  - Mark Craven (co-advisor)
  - Xiaojin Zhu (co-advisor)
  - Jude Shavlik
  - Ben Liblit
  - Michael Newton
  - Ben Recht

- UW–Madison AI community: Zhu and Craven groups, [Mlgroup], AIRG,…

- Biological collaborators: Brandi Gancarz and Ron Stewart

- Talk feedback: Ameet, Andreas, Bryan, Debbie, and Yimin

- Funding
  - Computation and Informatics in Biology and Medicine training program (CIBM, NIH/NLM T15 LM07359)
  - Wisconsin Alumni Research Foundation (WARF)
  - NIH grant R01 LM07050
Statistical debugging with latent topic models

Predicates → Vocabulary
Predicate counts → Word counts
Program run → Bag-of-words document
Debugging → Latent topic analysis
Bug patterns → Topics

```c
int x = my_func();
if (x > 5) {
  branch_42_true++;
  ...
} else {
  branch_42_false++;
  ...
}
```

Event Counts

<table>
<thead>
<tr>
<th>Event</th>
<th>Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>82</td>
</tr>
</tbody>
</table>
Statistical debugging with latent topic models

Predicate counts → Vocabulary
Program run → Word counts
Debugging → Bag-of-words document
Bug patterns → Latent topic analysis

```
int x = my_func();
if (x > 5) {
    branch_42_true++
    ...
} else {
    branch_42_false++
    ...
}
```

Event Counts

45 0
19 0
0 82
Statistical debugging with latent topic models

- Predicates → Vocabulary
- Predicate counts → Word counts
- Program run → Bag-of-words document
- Debugging → Latent topic analysis
- Bug patterns → Topics

```c
int x = my_func();
if (x > 5) {
    branch_42_true++;
    ...
} else {
    branch_42_false++
    ...
}
```

Event Counts

```
45 0
19 0
...
0 82
```
Statistical debugging with latent topic models

Predicates → Vocabulary
Predicate counts → Word counts
Program run → Bag-of-words document
Debugging → Latent topic analysis
Bug patterns → Topics

```c
int x = my_func();
if (x > 5) {
    branch_42_true++;
    ...
} else {
    branch_42_false++;
    ...
}
```

Event Counts

<table>
<thead>
<tr>
<th></th>
<th>45</th>
<th>19</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Statistical debugging with latent topic models

- Predicates → Vocabulary
- Predicate counts → Word counts
- Program run → Bag-of-words document
- Debugging → Latent topic analysis
- Bug patterns → Topics

```c
int x = my_func();
if (x > 5) {
    branch_42_true++;
    ...
} else {
    branch_42_false++;
    ...
}
```

Event Counts:

```
<table>
<thead>
<tr>
<th>Event</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>
```

Andrzejewski (UW-Madison)
Incorporating Domain Knowledge
Final defense 64 / 63
Experimental results

Rand index vs true clustering (1 = total agreement)

<table>
<thead>
<tr>
<th></th>
<th>ΔLDA</th>
<th>[1]</th>
<th>[2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>exif</td>
<td>1.00</td>
<td>0.88</td>
<td>1.00</td>
</tr>
<tr>
<td>grep</td>
<td>0.97</td>
<td>0.71</td>
<td>0.77</td>
</tr>
<tr>
<td>gzip</td>
<td>0.89</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>moss</td>
<td>0.93</td>
<td>0.93</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Baselines

1. Statistical Debugging: Simultaneous Isolation of Multiple Bugs (Zheng et al., ICML 2006)
2. Scalable Statistical Bug Isolation (Liblit et al., PLDI 2005)
Why latent topic modeling?

- **Graphical model** → **composability/extensibility**
  - Embed other side information in ΔLDA (e.g., static analysis)
  - Use ΔLDA probabilities in other procedures

- **Interpretable topics** \((gz\text{ip})\)
  - Usage topic 1: \texttt{longest\_match()} function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: \texttt{deflate\_fast()} (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model → composability/extensibility
  - Embed other side information in ΔLDA (e.g., static analysis)
  - Use ΔLDA probabilities in other procedures

- *Interpretable* topics (*gzip*)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model → composability/extensibility
  - Embed other side information in $\Delta$LDA (e.g., static analysis)
  - Use $\Delta$LDA probabilities in other procedures
- Interpretable topics (gzip)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model → composability/extensibility
  - Embed other side information in ΔLDA (e.g., static analysis)
  - Use ΔLDA probabilities in other procedures
- *Interpretative* topics (*gzip*)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a "dry" program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model $\rightarrow$ composability/extensibility
  - Embed other side information in ΔLDA (e.g., static analysis)
  - Use ΔLDA probabilities in other procedures
- *Interpretable* topics (*gzip*)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model → composability/extensibility
  - Embed other side information in ΔLDA (e.g., static analysis)
  - Use ΔLDA probabilities in other procedures
- *Interpretable* topics (gzip)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
Why latent topic modeling?

- Graphical model $\rightarrow$ composability/extensibility
  - Embed other side information in $\Delta$LDA (e.g., static analysis)
  - Use $\Delta$LDA probabilities in other procedures
- *Interpretable* topics (*gzip*)
  - Usage topic 1: `longest_match()` function (many redundant strings)
  - Usage topic 2: command-line parsing, exit code (a “dry” program run)
  - Usage topic 3: `deflate_fast()` (fast algorithm command-line flag)
△LDA toy example

Usage Topics

Buggy Topics
ΔLDA toy example

Usage Topics

Buggy Topics

"Successful" Documents

"Failing" Documents
ΔLDA toy example
ΔLDA toy example

Usage Topics

"Successful" Documents

"Failing" Documents

Recovered Topics

LDA

DeltaLDA

Incorporating Domain Knowledge

Final defense 67 / 63
## Debugging Dataset

<table>
<thead>
<tr>
<th>Program</th>
<th>Lines of Code</th>
<th>Bugs</th>
<th>Successful</th>
<th>Failing</th>
</tr>
</thead>
<tbody>
<tr>
<td>exif</td>
<td>10,611</td>
<td>2</td>
<td>352</td>
<td>30</td>
</tr>
<tr>
<td>grep</td>
<td>15,721</td>
<td>2</td>
<td>609</td>
<td>200</td>
</tr>
<tr>
<td>gzip</td>
<td>8,960</td>
<td>2</td>
<td>29</td>
<td>186</td>
</tr>
<tr>
<td>moss</td>
<td>35,223</td>
<td>8</td>
<td>1727</td>
<td>1228</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Program</th>
<th>Word Types</th>
<th>Usage</th>
<th>Bug</th>
</tr>
</thead>
<tbody>
<tr>
<td>exif</td>
<td>20</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>grep</td>
<td>2,071</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>gzip</td>
<td>3,929</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>moss</td>
<td>1,982</td>
<td>14</td>
<td>8</td>
</tr>
</tbody>
</table>
CBI Results

**grep**

- 56 bug1 runs
- 144 bug2 runs

**moss**

- 254 bug1 runs
- 106 bug3 runs
- 147 bug4 runs
- 329 bug5 runs
- 206 bug8 runs
- 186 other runs

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 69 / 63
z-labels via undirected LDA

\[ D \]

\[ \beta \rightarrow \phi \]

\[ T \]

\[ N_d \]

\[ w \rightarrow z \rightarrow \theta \rightarrow \alpha \]
z-labels via undirected LDA
z-labels via undirected LDA

\[
\begin{align*}
\alpha & \quad \theta \\
\beta & \quad \phi \\
& \quad T \\
& \quad z \\
& \quad w \\
& \quad Nd \\
& \quad D
\end{align*}
\]
z-labels via undirected LDA
z-labels via undirected LDA

Constraints

$\alpha \quad \theta \quad z \quad w \quad N_d$

$D$

$\beta \quad \phi \quad T$
Translation

- Given: “translation” seed words
- Do: use $z$-labels to build “translation” topic
- Lots of relevant/seed terms
- Overall captures “translation” concept
Given: “translation” seed words
Do: use z-labels to build “translation” topic

- Lots of relevant/seed terms
- Overall captures “translation” concept
Given: “translation” seed words
Do: use z-labels to build “translation” topic
Lots of relevant/seed terms
Overall captures “translation” concept
Given: “translation” seed words
Do: use z-labels to build “translation” topic
Lots of relevant/seed terms
Overall captures “translation” concept

| Topic 0 | translation, ribosomal, trna, rrna, initiation, ribosome, protein ribosomes, is, factor, processing, translational, nucleolar pre-rrna, synthesis, small, 60s, eukaryotic, biogenesis, subunit trnas, subunits, large, nucleolus, factors, 40, synthetase, free modification, rna, depletion, eif-2, initiator, 40s, ef-3 anticodon, maturation 18s, eif2, mature, eif4e, synthetases aminoacylation, snornas, assembly, eif4g, elongation |
Concept Expansion: Yeast Corpus

- Corpus: approx 9,000 yeast-related abstracts
- Goal: use “seed” words to build a topic around a concept
- Target biological concept: translation
- Seed words: translation, tRNA, anticodon, ribosome
- $C^{(i)} = \{\text{Topic 0}\}$, $\eta = 1$ (hard constraint)
- $T = 100$, $\alpha = 0.5$, $\beta = 0.1$
Dirichlet Forest example

| Topic 13 | go school cancer into well free cure college ... graduate ... law ... surgery recovery ... |

- Topic 13 mixes *college* and *illness* wish topics
- Prefer to split [go school into college] and [cancer free cure well]
- Goal: topics separate these words,
Dirichlet Forest example

<table>
<thead>
<tr>
<th>Topic 13</th>
<th>go school cancer into well free cure college</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>. . graduate . . law . . surgery recovery . .</td>
</tr>
</tbody>
</table>

- Topic 13 mixes *college* and *illness* wish topics
- Prefer to **split** [go school into college] and [cancer free cure well]
- Goal: topics separate these words,
Dirichlet Forest example

Topic 13 | go school cancer into well free cure college
          | ... graduate ... law ... surgery recovery ...

Topic 13(a) | job go school great into good college
              | ... business graduate finish grades away law accepted ...
Topic 13(b) | mom husband cancer hope free son well
              | ... full recovery surgery pray heaven pain aids ...

- Topic 13 mixes *college* and *illness* wish topics
- Prefer to **split** [go school into college] and [cancer free cure well]
- Goal: topics separate these words, as well as related words
Dirichlet Forest example

<table>
<thead>
<tr>
<th>Topic 13</th>
<th>go school cancer into well free cure college... graduate... law... surgery recovery...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 13(a)</td>
<td>job go school great into good college... business graduate finish grades away law accepted...</td>
</tr>
<tr>
<td>Topic 13(b)</td>
<td>mom husband cancer hope free son well... full recovery surgery pray heaven pain aids...</td>
</tr>
</tbody>
</table>

- Topic 13 mixes *college* and *illness* wish topics
- Prefer to **split** *[go school into college]* and *[cancer free cure well]*
- Goal: topics separate these words, **as well as related words**
Control variance of *subsets* of variables

- Sample Dirichlet($\gamma$) at parent, distribute mass to children
- Mass reaching leaves are final multinomial parameters $\phi$
- $\Delta(s) = 0$ for all internal node $s \rightarrow$ standard Dirichlet
  (for our trees, true when $\eta = 1$)
- *Conjugate* to multinomial $\rightarrow$ integrate out (“collapse”) $\phi$

\[
\gamma \left\{ \begin{array}{c}
\eta \beta \\ \eta \beta \\
A \\
B \\
C \\
\end{array} \right. \quad \begin{array}{c}
2\beta \\
\beta \\
\end{array} \\
\]

($\beta = 1, \eta = 50$)
Dirichlet Tree ("dice factory 2.0")
Dennis III 1991, Minka 1999

- Control variance of *subsets* of variables
  - Sample Dirichlet($\gamma$) at parent, distribute mass to children
  - Mass reaching leaves are final multinomial parameters $\phi$
  - $\Delta(s) = 0$ for all internal node $s \to$ standard Dirichlet
    (for our trees, true when $\eta = 1$)
  - *Conjugate* to multinomial $\to$ integrate out ("collapse") $\phi$

\[
\begin{align*}
\gamma & \{ 2\beta, \beta \\
\eta\beta & \{ A, B \\
\eta\beta & \{ C \\\n(\beta = 1, \eta = 50) & 0.91, 0.09, 0.09
\end{align*}
\]
Dirichlet Tree ("dice factory 2.0")
Dennis III 1991, Minka 1999

- Control variance of *subsets* of variables
  - Sample Dirichlet($\gamma$) at parent, distribute mass to children
  - Mass reaching leaves are final multinomial parameters $\phi$
  - $\Delta(s) = 0$ for all internal node $s \rightarrow$ standard Dirichlet
    (for our trees, true when $\eta = 1$)
  - **Conjugate** to multinomial $\rightarrow$ integrate out ("collapse") $\phi$

- Incorporating Domain Knowledge

- Andrzejewski (UW-Madison)

```
γ
{ 2β
  ηβ
    A
    B
  β
    ηβ
    C

(β = 1, η = 50)
```

```
0.91 0.09
0.58 0.42
φ=0.53 0.38 0.09
```
Dirichlet Tree ("dice factory 2.0")
Dennis III 1991, Minka 1999

- Control variance of subsets of variables
  - Sample Dirichlet($\gamma$) at parent, distribute mass to children
  - Mass reaching leaves are final multinomial parameters $\phi$
  - $\Delta(s) = 0$ for all internal node $s \rightarrow$ standard Dirichlet
    (for our trees, true when $\eta = 1$)
  - Conjugate to multinomial $\rightarrow$ integrate out ("collapse") $\phi$

\[
\rho(\phi|\gamma) = \left( \prod_k \phi^{(k)} \gamma^{(k)} - 1 \right) \left( \prod_s \frac{\Gamma\left( \sum_k C(s) \gamma^{(k)} \right)}{\prod_k \Gamma\left( \gamma^{(k)} \right)} \left( \sum_k \phi^{(k)} \right)^{\Delta(s)} \right)
\]
Dirichlet Tree ("dice factory 2.0")
Dennis III 1991, Minka 1999

- Control variance of subsets of variables
  - Sample Dirichlet($\gamma$) at parent, distribute mass to children
  - Mass reaching leaves are final multinomial parameters $\phi$
  - $\Delta(s) = 0$ for all internal node $s$ → standard Dirichlet
    (for our trees, true when $\eta = 1$)
  - Conjugate to multinomial → integrate out ("collapse") $\phi$

\[
p(\phi|\gamma) = \left( \prod_{k}^{L} \phi^{(k)} \gamma^{(k)} - 1 \right) \left( \prod_{s}^{l} \frac{\Gamma \left( \sum_{k}^{C(s)} \gamma^{(k)} \right)}{\prod_{k}^{C(s)} \Gamma \left( \gamma^{(k)} \right)} \left( \sum_{k}^{L(s)} \phi^{(k)} \right)^{\Delta(s)} \right)
\]
Dirichlet Tree ("dice factory 2.0")
Dennis III 1991, Minka 1999

- Control variance of *subsets* of variables
  - Sample Dirichlet($\gamma$) at parent, distribute mass to children
  - Mass reaching leaves are final multinomial parameters $\phi$
  - $\Delta(s) = 0$ for all internal node $s \rightarrow$ standard Dirichlet
    (for our trees, true when $\eta = 1$)
  - **Conjugate** to multinomial $\rightarrow$ integrate out ("collapse") $\phi$

\[
\rho(w | \gamma) = \\
\prod_{s} \left( \frac{\Gamma \left( \sum_{k} C(s) \gamma(k) \right)}{\Gamma \left( \sum_{k} C(s) \left( \gamma(k) + n(k) \right) \right)} \prod_{k} \frac{C(s) \Gamma \left( \gamma(k) + n(k) \right)}{\Gamma \left( \gamma(k) \right)} \right)
\]
Must-Link \((\text{school, college})\) via Dirichlet Tree

- Place \((\text{school, college})\) beneath internal node
- Large edge weights beneath this node (large \(\eta\))

\[
\begin{align*}
\eta \beta \\
\beta \\
\beta \\
\eta \beta \\
\eta \beta
\end{align*}
\]

\(\text{college school lottery}\)

\((\beta = 1, \eta = 50)\)
Must-Link \((school, college)\) via Dirichlet Tree

- Place \((school, college)\) beneath internal node
- Large edge weights beneath this node (large \(\eta\))

\[
\begin{align*}
\text{lottery} & \quad \text{school} \\
\beta & \quad 2\beta \\
\eta\beta & \quad \eta\beta \\
\text{college} & \quad \text{school} \quad \text{lottery}
\end{align*}
\]

\((\beta = 1, \eta = 50)\)
Must-Link \((school, college)\) via Dirichlet Tree

- Place \((school, college)\) beneath internal node
- Large edge weights beneath this node (large \(\eta\))

\[
\begin{align*}
&\eta \beta &\eta \beta \\
&\text{college} &\text{school} &\text{lottery}
\end{align*}
\]

\[
\begin{align*}
\beta &
\end{align*}
\]

\[
\begin{align*}
2\beta &
\end{align*}
\]

\[
\begin{align*}
\text{lottery} &\text{school} &\text{college}
\end{align*}
\]

\((\beta = 1, \eta = 50)\)
Dirichlet Forest LDA

Incorporating Domain Knowledge
Dirichlet Forest LDA

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 76 / 63
Dirichlet Forest LDA

\[ \theta \]

\[ N_d \]

\[ w \]

\[ z \]

\[ D \]

\[ \alpha \]

\[ T \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]

\[ \beta \]

\[ \eta \]

\[ \phi \]

\[ \alpha \]
Dirichlet Forest LDA

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 76 / 63
Dirichlet Forest LDA

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 76 / 63
Dirichlet Forest with biological domain knowledge

- Given: concepts from the Gene Ontology (GO)
  - Biological **processes**: transcription, translation, replication
  - Process **phases**: initiation, elongation, termination
- Do: learn meaningful process+phase “composite” topics
<table>
<thead>
<tr>
<th></th>
<th>LDA</th>
<th>DF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>transcription</td>
<td>• • •</td>
<td>• •</td>
</tr>
<tr>
<td>transcriptional</td>
<td>• • •</td>
<td>• •</td>
</tr>
<tr>
<td>template</td>
<td></td>
<td>• •</td>
</tr>
<tr>
<td>translation</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>translational</td>
<td></td>
<td>• •</td>
</tr>
<tr>
<td>tRNA</td>
<td></td>
<td>• •</td>
</tr>
<tr>
<td>replication</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>cycle</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>division</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>initiation</td>
<td>• • • •</td>
<td>• • •</td>
</tr>
<tr>
<td>start</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>assembly</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>elongation</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>termination</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>disassembly</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>release</td>
<td>• •</td>
<td>• •</td>
</tr>
<tr>
<td>stop</td>
<td>•</td>
<td>• •</td>
</tr>
</tbody>
</table>
Maximal cliques

- Maximal cliques of complement graph $\leftrightarrow$ independent sets
- Worst-case: $3^n/3$ (Moon & Moser 1965)
- We are only concerned with connected graphs, but still $O(3^{n/3})$ (Griggs et al 1988)
- Find cliques with Bron-Kerbosch (branch-and-bound)
Maximal cliques

- Maximal cliques of complement graph $\leftrightarrow$ independent sets
- Worst-case: $3^n$ (Moon & Moser 1965)
- We are only concerned with connected graphs, but still $O(3^{n/3})$ (Griggs et al. 1988)
- Find cliques with Bron-Kerbosch (branch-and-bound)
Maximal cliques

- Maximal cliques of complement graph ↔ independent sets
- Worst-case: $3^n$ (Moon & Moser 1965)
- We are only concerned with *connected* graphs, but still $O(3^n)$ (Griggs et al 1988)
- Find cliques with Bron-Kerbosch (branch-and-bound)
Sampling a Tree from the Forest

Vocabulary: [A, B, C, D, E, F, G]
Must-Links: (A, B)
Cannot-Links: (A, D), (C, D), (E, F)

Cannot-Link-graph:

Andrzejewski (UW-Madison)
Sampling a Tree from the Forest

Connected components

- ABCD
- EF
- G

- A
- B
- C
- D
- E
- F
- G

AB

D

C

E

F

G
Sampling a Tree from the Forest

Subgraph complements

ABCD EF G
A B C D E F G
AB
C
D
E
F
G
Sampling a Tree from the Forest

Maximal cliques

ABCD  EF  G

A  B  C  D  E  F  G

AB

CD

EF

G
Sampling a Tree from the Forest

Sample $q^{(1)}$ for first connected component

$\eta \quad \eta$

$\eta$

$AB \quad C \quad D$

$AB \quad C \quad D$

$A \quad B \quad C \quad D \quad E \quad F \quad G$

$AB \quad C \quad D$

$E$

$F$

$G$
Sampling a Tree from the Forest

$q^{(1)} = 1$ (choose $ABC$)

$\eta \eta \eta \eta = 1$

Andrzejewski (UW-Madison)
Sampling a Tree from the Forest

Insert chosen Cannot-Link subtree

$q^{(1)} = 1$

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 80 / 63
Sampling a Tree from the Forest

Put \((A, B)\) under Must-Link subtree

\[ q^{(1)} = 1 \]

\[ \eta \]

\[ \eta \]

\[ \eta \]

\[ \eta \]

\[ A \quad B \quad C \quad D \quad E \quad F \quad G \]

\[ \text{AB} \]

\[ \text{C} \]

\[ \text{D} \]

\[ \text{E} \]

\[ \text{F} \]

\[ \text{G} \]
Sampling a Tree from the Forest

Sample $q^{(2)}$ for second connected component

$q^{(1)}=1$

$q^{(2)}=\eta$

A B C D E F G

AB C D E F G

Andrzejewski (UW-Madison)
Incorporating Domain Knowledge
Final defense 80 / 63
Sampling a Tree from the Forest

$q^{(2)} = 2$ (choose $F$)

$q^{(1)} = 1$

$\eta \quad \eta$

A B C D

G

AB

|\eta\rangle

|\eta\rangle

$\eta \quad \eta$

E F

G

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense 80 / 63
Sampling a Tree from the Forest

Insert chosen Cannot-Link subtree

\[ q^{(1)} = 1 \]

\[ q^{(2)} = 2 \]

\[ \eta \]

A B C D E F G

\[ \eta \]

Andrzejewski (UW-Madison)

Incorporating Domain Knowledge

Final defense
| Topic | Top words sorted by $\phi = p(\text{word}|\text{topic})$ |
|-------|--------------------------------------------------|
| 0     | love i you me and will forever that with hope    |
| 1     | and health for happiness family good my friends |
| 2     | year new happy a this have and everyone years    |
| 3     | that is it you we be t are as not s will can     |
| 4     | my to get job a for school husband s that into   |
| 5     | to more of be and no money stop live people      |
| 6     | to our the home for of from end safe all come    |
| 7     | to my be i find want with love life meet man     |
| 8     | a and healthy my for happy to be have baby       |
| 9     | a 2008 in for better be to great job president   |
| 10    | i wish that would for could will my lose can     |
| 11    | peace and for love all on world earth happiness  |
| 12    | may god in all your the you s of bless 2008      |
| 13    | the in to of world best win 2008 go lottery      |
| 14    | me a com this please at you call 4 if 2 www     |
| Topic | Top words sorted by $\phi = p(\text{word} | \text{topic})$ |
|-------|--------------------------------------------------------------|
| 0     | love i you me and will forever that with hope                |
| 1     | and health for happiness family good my friends              |
| 2     | year new happy a this have and everyone years                |
| 3     | that is it you we be t are as not s will can                 |
| 4     | my to get job a for school husband s that into               |
| 5     | to more of be and no money stop live people                  |
| 6     | to our the home for of from end safe all come                |
| 7     | to my be i find want with love life meet man                  |
| 8     | a and healthy my for happy to be have baby                   |
| 9     | a 2008 in for better be to great job president               |
| 10    | i wish that would for could will my lose can                  |
| 11    | peace and for love all on world earth happiness              |
| 12    | may god in all your the you s of bless 2008                  |
| 13    | the in to of world best win 2008 go lottery                  |
| 14    | me a com this please at you call 4 if 2 www                  |
| Topic | Top words sorted by $\phi = p(\text{word}|\text{topic})$ |
|-------|-----------------------------------------------------|
| 0     | love forever marry happy together mom back           |
| 1     | health happiness good family friends prosperity      |
| 2     | life best live happy long great time ever wonderful  |
| 3     | out not up do as so what work don was like            |
| 4     | go school cancer into well free cure college         |
| 5     | no people stop less day every each take children     |
| 6     | home safe end troops iraq bring war husband house    |
| 7     | love peace true happiness hope joy everyone dreams   |
| 8     | happy healthy family baby safe prosperous everyone   |
| 9     | better job hope president paul great ron than person |
| 10    | make money lose weight meet finally by lots hope married |
| 12    | god bless jesus loved know everyone love who loves    |
| 13    | peace world earth win lottery around save            |
| 14    | com call if 4 2 www u visit 1 3 email yahoo          |

**Isolate** and to for a the year in new all my 2008

i to wish my for and a be that the in
| Topic | Top words sorted by $\phi = p(\text{word}|\text{topic})$ |
|-------|----------------------------------------------------------|
| 0     | love forever marry happy together mom back               |
| 1     | health happiness good family friends prosperity          |
| 2     | life best live happy long great time ever wonderful       |
| 3     | out not up do as so what work don was like                |
| MIXED | go school cancer into well free cure college              |
| 5     | no people stop less day every each take children         |
| 6     | home safe end troops iraq bring war husband house        |
| 7     | love peace true happiness hope joy everyone dreams       |
| 8     | happy healthy family baby safe prosperous everyone        |
| 9     | better job hope president paul great ron than person      |
| 10    | make money lose weight meet finally by lots hope married  |
| Isolate | and to for a the year in new all my 2008                  |
| 12    | god bless jesus loved know everyone love who loves        |
| 13    | peace world earth win lottery around save                 |
| 14    | com call if 4 2 www u visit 1 3 email yahoo              |
| Isolate | i to wish my for and a be that the in                    |
split([cancer free cure well],[go school into college])

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>love forever happy together marry fall</td>
<td>love forever happy together marry fall</td>
</tr>
<tr>
<td>1</td>
<td>health happiness family good friends</td>
<td>health happiness family good friends</td>
</tr>
<tr>
<td>2</td>
<td>life happy best live love long time</td>
<td>life happy best live love long time</td>
</tr>
<tr>
<td>3</td>
<td>as not do so what like much don was</td>
<td>as not do so what like much don was</td>
</tr>
<tr>
<td>4</td>
<td>out make money house up work grow able</td>
<td>out make money house up work grow able</td>
</tr>
<tr>
<td>5</td>
<td>people no stop less day every each take</td>
<td>people no stop less day every each take</td>
</tr>
<tr>
<td>6</td>
<td>home safe end troops iraq bring war husband</td>
<td>home safe end troops iraq bring war husband</td>
</tr>
<tr>
<td>7</td>
<td>love peace happiness true everyone joy</td>
<td>love peace happiness true everyone joy</td>
</tr>
<tr>
<td>8</td>
<td>happy healthy family baby safe prosperous</td>
<td>happy healthy family baby safe prosperous</td>
</tr>
<tr>
<td>9</td>
<td>better president hope paul ron than person</td>
<td>better president hope paul ron than person</td>
</tr>
<tr>
<td>10</td>
<td>lose meet man hope boyfriend weight finally</td>
<td>lose meet man hope boyfriend weight finally</td>
</tr>
</tbody>
</table>

Isolate and to for a the year in new all my 2008

12 | god bless jesus loved everyone know loves                        | god bless jesus loved everyone know loves                        |
13 | peace world earth win lottery around save                        | peace world earth win lottery around save                        |
14 | com call if 4 www 2 u visit 1 email yahoo 3                       | com call if 4 www 2 u visit 1 email yahoo 3                       |

Isolate i to wish my for and a be that the in me get

Split job go school great into good college

Split mom husband cancer hope free son well

Andrzejewski (UW-Madison)  Incorporating Domain Knowledge  Final defense 83 / 63
merge([love ... marry...],[meet ... married...])

| Topic     | Top words sorted by $\phi = p(\text{word}|\text{topic})$ |
|-----------|----------------------------------------------------------|
| Merge     | love lose weight together forever marry meet             |
| success   | health happiness family good friends prosperity          |
| life      | life happy best live time long wishes ever years          |
| -         | as do not what someone so like don much he               |
| money     | out make money up house work able pay own lots            |
| people    | no people stop less day every each other another          |
| iraq      | home safe end troops iraq bring war return               |
| joy       | love true peace happiness dreams joy everyone            |
| family    | happy healthy family baby safe prosperous                |
| vote      | better hope president paul ron than person bush          |
| Isolate   | and to for a the year in new all my                     |
| god       | god bless jesus everyone loved know heart christ         |
| peace     | peace world earth win lottery around save                |
| spam      | com call if u 4 www 2 3 visit 1                         |
| Split     | i to wish my for and a be that the                      |
| Split     | job go great school into good college hope move          |
|           | mom hope cancer free husband son well dad cure           |
Encoding LDA variants

Concept-Topic Model (Chemudugunta et. al., ISWC 2008)
- Special concept topics only emit certain words
- Topic $t$ special words: $w_{c1}, w_{c2}, \ldots, w_{cn}$

$$\forall i \ S(i, s) \land Z(i, t) \Rightarrow (\overline{W(i, w_{c1})} \lor \overline{W(i, w_{c2})} \lor \ldots \lor \overline{W(i, w_{cn})})$$

Hidden Markov Topic Model (Gruber et. al., AISTATS 2007)
- Same topic used for entire sentence
- Topic transitions allowed at sentence boundary

$$\forall i, j, s, t \ S(i, s) \land S(j, s) \land Z(i, t) \Rightarrow Z(j, t)$$
$$\forall i, s \ S(i, s) \land \overline{S(i+1, s)} \land Z(i, t) \Rightarrow Z(j, t')$$
Encoding LDA variants

Concept-Topic Model (Chemudugunta et. al., ISWC 2008)
- Special *concept* topics only emit certain words
- Topic $t$ special words: $w_{c1}, w_{c2}, \ldots, w_{cn}$

$$\forall i \ Z(i, t) \Rightarrow (\overline{w}(i, w_{c1}) \lor \overline{w}(i, w_{c2}) \lor \ldots \lor \overline{w}(i, w_{cn}))$$

Hidden Markov Topic Model (Gruber et. al., AISTATS 2007)
- Same topic used for entire sentence
- Topic transitions allowed at sentence boundary

$$\forall i, j, s, t \ S(i, s) \land S(j, s) \land Z(i, t) \Rightarrow Z(j, t)$$

$$\forall i, s \ S(i, s) \land \neg S(i + 1, s) \land Z(i, t) \Rightarrow Z(j, t')$$
Decomposing the \( z \)-step

- Remove *trivially true* ground formulas (Shavlik and Natarajan, AAAI 2009)
  - Example
    - \( \overline{w(i, \text{taxes})} \Rightarrow \overline{z(i, 3)} \)
    - *true* irrespective of \( z \) for all \( i \) where \( w_i \neq \text{taxes} \)
  - Let \( z_{KB} \) be all \( z_i \) involved in \( \geq 1 \) *non-trivial* grounding
  - \( z_i \notin z_{KB} \) can simply be set by \( \arg \max_t \phi_t(w_i)\theta_{d_i}(t) \)
Decomposing the \( z \)-step

- Remove *trivially true* ground formulas (Shavlik and Natarajan, AAAI 2009)
- Example
  - \( \overline{w}(i, \text{taxes}) \Rightarrow z(i, 3) \)
  - *true* irrespective of \( z \) for all \( i \) where \( w_i \neq \text{taxes} \)
- Let \( z_{KB} \) be all \( z_i \) involved in \( \geq 1 \) *non-trivial* grounding
- \( z_i \notin z_{KB} \) can simply be set by \( \arg\max_t \phi_t(w_i)\theta_{d_i}(t) \)
Decomposing the $z$-step

- Remove *trivially true* ground formulas
  
  (Shavlik and Natarajan, AAAI 2009)

- Example
  
  - $w(i, \text{taxes}) \Rightarrow z(i, 3)$
  
  - *true* irrespective of $z$ for all $i$ where $w_i \neq \text{taxes}$

- Let $z_{KB}$ be all $z_i$ involved in $\geq 1$ *non-trivial* grounding

- $z_i \notin z_{KB}$ can simply be set by $\text{argmax}_{t} \phi_t(w_i) \theta_{d_i}(t)$
Decomposing the \( z \)-step

- Remove *trivially true* ground formulas (Shavlik and Natarajan, AAAI 2009)

**Example**

- \( w(i, \text{taxes}) \Rightarrow z(i, 3) \)
- *true* irrespective of \( z \) for all \( i \) where \( w_i \neq \text{taxes} \)

- Let \( z_{KB} \) be all \( z_i \) involved in \( \geq 1 \) *non-trivial* grounding

- \( z_i \notin z_{KB} \) can simply be set by argmax \( \phi_t(w_i)\theta_d_i(t) \)
Decomposing the z-step

- Remove *trivially true* ground formulas (Shavlik and Natarajan, AAAI 2009)
- Example
  - \( w(i, \text{taxes}) \Rightarrow z(i, 3) \)
  - *true* irrespective of \( z \) for all \( i \) where \( w_i \neq \text{taxes} \)
- Let \( z_{KB} \) be all \( z_i \) involved in \( \geq 1 \) *non-trivial* grounding
  - \( z_i \notin z_{KB} \) can simply be set by \( \text{argmax}_t \phi_t(w_i)\theta_{d_i}(t) \)
Decomposing the z-step

- Remove *trivially true* ground formulas (Shavlik and Natarajan, AAAI 2009)

Example

- \( w(i, \text{taxes}) \Rightarrow z(i, 3) \)
- *true* irrespective of \( z \) for all \( i \) where \( w_i \neq \text{taxes} \)

Let \( z_{KB} \) be all \( z_i \) involved in \( \geq 1 \) *non-trivial* grounding

- \( z_i \notin z_{KB} \) can simply be set by argmax \( \phi_t(w_i)\theta_{d_i}(t) \)
Collapsed Gibbs Sampling

\[ P(z_i = v | z_{-i}, w) \propto \left( \frac{n^{(d)}_{-i,v} + \alpha}{\sum_u n^{(d)}_{-i,u} + \alpha} \right) \left( \frac{n^{(w_i)}_{-i,v} + \beta}{\sum_w n^{(w'_i)}_{-i,v} + \beta} \right) \times \exp \left( \sum_{\ell} \eta_\ell f_\ell(z_i = v, z_{-i}, w) \right) \]

- MLN research suggests this will not work well
- We do not really care about the full posterior anyways
Collapsed Gibbs Sampling

\[ P(z_i = v | z_{-i}, w) \propto \left( \frac{n^{(d)}_{-i,v} + \alpha}{\sum_u (n^{(d)}_{-i,u} + \alpha)} \right) \left( \frac{n^{(w)}_{-i,v} + \beta}{\sum_{w'} (n^{(w')}_{-i,v} + \beta)} \right) \times \exp \left( \sum_{\ell} \eta_{\ell} f_{\ell}(z_i = v, z_{-i}, w) \right) \]

- MLN research suggests this will not work well
- We do not really care about the full posterior anyways
Using syntax to recover related classes of words

- **cells** $\rightarrow$ special topic 0 (T-cells, lymphocytes)
- **cell states** $\rightarrow$ special topic 1 (activated, treated, transfected)
- $\forall i, j \text{ Dependency}(\text{AdjectivalModifier}, i, j) \Rightarrow (z(i, 0) \Leftrightarrow z(j, 1))$

- Activated lymphocytes pass into the blood stream...
- Transformation of neuraminidase treated lymphocytes...
- ...shortening of process length in OA treated neurons.

Andrzejewski (UW-Madison)  
Incorporating Domain Knowledge  
Final defense 88 / 63
cells → special topic 0 (T-cells, lymphocytes)

cell states → special topic 1 (activated, treated, transfected)

∀i, j Dependency(AdjectivalModifier, i, j) ⇒ (z(i, 0) ⇔ z(j, 1))

Activated lymphocytes pass into the blood stream...

Transformation of neuraminidase treated lymphocytes...

...shortening of process length in OA treated neurons.
Using syntax to recover related classes of words

- **cells** $\rightarrow$ special topic 0 (T-cells, lymphocytes)
- **cell states** $\rightarrow$ special topic 1 (activated, treated, transfected)
- $\forall i, j \text{ Dependency}(\text{AdjectivalModifier}, i, j) \Rightarrow (z(i, 0) \Leftrightarrow z(j, 1))$

- Activated lymphocytes pass into the blood stream...
- Transformation of neuraminidase treated lymphocytes...
- ...shortening of process length in OA treated neurons.
Using syntax to recover related classes of words

- **cells** $\rightarrow$ special topic 0 (T-cells, lymphocytes)
- **cell states** $\rightarrow$ special topic 1 (activated, treated, transfected)
- $\forall i, j \text{ Dependency(AdjectivalModifier, } i, j) \Rightarrow (z(i, 0) \Leftrightarrow z(j, 1))$

- Activated lymphocytes pass into the blood stream...
- Transformation of neuraminidase treated lymphocytes...
- ...shortening of process length in OA treated neurons.
# LogicLDA datasets

| Dataset | $N$   | $W$ | $D$ | $T$ | $|\bigcup_k G(\psi_k)|$ |
|---------|-------|-----|-----|-----|------------------|
| S1      | 16    | 3   | 4   | 2   | 64               |
| S2      | 32    | 4   | 4   | 3   | 192              |
| Mac     | 153986| 3652| 2000| 20  | 4388860          |
| Comp    | 482634| 8285| 5000| 20  | 6295             |
| Con     | 422229| 6156| 2740| 25  | 99847            |
| Pol     | 733072| 13196| 2000| 20  | 12049600000     |
| HDG     | 2903640| 13817| 21352| 100 | 47236814        |