TOPIC MODELING WITH STRUCTURED PRIORS FOR TEXT-DRIVEN SCIENCE

MICHAEL J. PAUL
JOHNS HOPKINS UNIVERSITY
TEXT AS DATA
TEXT AS DATA

natural language processing (NLP)
TEXT AS DATA

text-driven science
TEXT AS DATA

- Computational social science
- Computational journalism
- Crisis informatics
- Public health informatics
- Computational epidemiology

Text-driven science
• Disease monitoring in Twitter


Lamb, Paul, Dredze (2013) Separating fact from fear: Tracking flu infections on Twitter. NAACL.


• Air pollution in Chinese social media


• Measuring healthcare quality from online reviews
TEXT AS DATA

MAKING SENSE
Topic modeling
A topic model is a **probabilistic model** of text

- We pretend that our data (text) are the output of a probabilistic process that generates data
I’ve had the flu and fever all week :( staying home from school and watching a lot of tv
I've had the flu and fever all week :( staying home from school and watching a lot of tv.
I've had the flu and fever all week :( staying home from school and watching a lot of tv
I've had the flu and fever all week :( staying home from school and watching a lot of tv.
I've had the flu and fever all week :( staying home from school and watching a lot of tv.
I've had the flu and fever all week :( staying home from school and watching a lot of tv sick sore throat feel fever flu allergies nose eyes allergy allergic sneezing watch watching tv killing movie seen class school read test doing finish …
I've had the flu and fever all week :( staying home from school and watching a lot of tv...
I've had the flu and fever all week :( staying home from school and watching a lot of tv.
I've had the flu and fever all week :( staying home from school and watching a lot of tv sick sore throat feel fever flu allergies nose eyes allergy allergic sneezing watch watching tv killing movie seen class school read test doing finish ...
I've had the flu and fever all week :( staying home from school and watching a lot of tv sick sore throat feel fever flu... allergies nose eyes allergy allergic sneezing... watch watching tv killing movie seen... class school read test doing finish...
I’ve had the flu and fever all week :( staying home from school and watching a lot of tv
Our imaginary process also needs to generate all these distributions.
Our imaginary process also needs to generate all these distributions

- We need a distribution over distributions
  - Called a **prior** distribution
Dirichletlet($\rho \times$)
TOPIC MODELING

PRIORS

Dirichlet( )
TOPIC MODELING

sick
sore
throat
feel
fever
flu
...

allergies
nose
eyes
allergy
allergic
sneezing
...

watch
watching
tv
killing
movie
seen
...

class
school
read
test
doing
finish
...

Dirichlet( )
TOPIC MODELING

sick
sore
throat
feel
fever
flu

allergies
nose
eyes
allergy
allergic
sneezing

watch
watching
tv
killing
movie
seen

class
school
read
test
doing
finish

Dirichlet()
TOPIC MODELING

sick
sore
throat
feel
fever
flu

allergies
nose
eyes
allergy
allergic
sneezing

watch
watching
tv
killing
movie
seen

class
school
read
test
doing
finish

...
I’ve had the flu and fever all week :( staying home from school and watching a lot of tv
Latent Dirichlet Allocation (LDA)
Blei, Ng, Jordan 2003

The topic and word distributions have Dirichlet priors:

\[ \phi_t \sim \text{Dirichlet}(\tilde{\phi}) \]
\[ \theta_m \sim \text{Dirichlet}(\tilde{\theta}) \]

Paul, Dredze (2011) You are what you tweet: Analyzing Twitter for public health. *5th International Conference on Weblogs and Social Media (ICWSM).*
Topics can be organized in ways that are more interpretable to users
Topics in online doctor reviews:

Both have **positive sentiment**

- **best**
- **years**
- **caring**
- **care**
- **patients**
- **patient**
- **recommend**
- **family**

- **time**
- **staff**
- **great**
- **helpful**
- **feel**
- **questions**
- **office**
- **friendly**

- **office**
- **time**
- **appointment**
- **rude**
- **staff**
- **room**
- **didn’t**
- **wait**

Both about **staff/office issues**
# Topics in online doctor reviews:

<table>
<thead>
<tr>
<th></th>
<th>Staff/Office</th>
<th>Personality</th>
<th>Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Positive</strong></td>
<td>time staff great helpful feel questions office friendly</td>
<td>best years caring care patients patient recommend family</td>
<td>surgery first son life surgeon daughter recommend thank</td>
</tr>
<tr>
<td><strong>Negative</strong></td>
<td>office time appointment rude staff room didn’t wait</td>
<td>care medical patients doesn’t help know don’t problem</td>
<td>pain told went said surgery later didn’t months</td>
</tr>
</tbody>
</table>
A multi-dimensional topic model
Word distributions are grouped into different concepts

- e.g. sentiment and aspect

Analyzing online drug forums:

Paul, Dredze (2013) **Summarizing drug experiences with multi-dimensional topic models.** *North American ACL (NAACL).*

Paul, Chisolm, Johnson, Vandrey, Dredze (in preparation) **Who participates in online drug communities? A large-scale analysis of demographic and temporal trends.**
3-dimensional model:
- Drug type
- Route of administration (i.e. method of intake)
- Aspect

<table>
<thead>
<tr>
<th>Drug (22 total)</th>
<th>Route</th>
<th>Aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alcohol</td>
<td>Injection</td>
<td>Chemistry</td>
</tr>
<tr>
<td>Amphetamine</td>
<td>Oral</td>
<td>Culture</td>
</tr>
<tr>
<td>Cannabis</td>
<td>Smoking</td>
<td>Effects</td>
</tr>
<tr>
<td>Cocaine</td>
<td>Snorting</td>
<td>Health</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>Usage</td>
</tr>
<tr>
<td>Salvia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tobacco</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Suppose we want to model: \((\text{Marijuana}, \text{Oral}, \text{Chemistry})\)
<table>
<thead>
<tr>
<th>Marijuana</th>
<th>Oral</th>
<th>Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>weed</td>
<td>capsules</td>
<td>solvent</td>
</tr>
<tr>
<td>cannabis</td>
<td>consumes</td>
<td>extraction</td>
</tr>
<tr>
<td>thc</td>
<td>toast</td>
<td>evaporate</td>
</tr>
<tr>
<td>marijuana</td>
<td>stomach</td>
<td>evaporated</td>
</tr>
<tr>
<td>stoned</td>
<td>chewing</td>
<td>solvents</td>
</tr>
<tr>
<td>bowl</td>
<td>ambien</td>
<td>evaporated</td>
</tr>
<tr>
<td>bud</td>
<td>digestion</td>
<td>solvents</td>
</tr>
<tr>
<td>joint</td>
<td>juice</td>
<td>evaporation</td>
</tr>
<tr>
<td>blunt</td>
<td>absorbed</td>
<td>yield</td>
</tr>
<tr>
<td>herb</td>
<td>ingestion</td>
<td>chloride</td>
</tr>
<tr>
<td>bong</td>
<td>meal</td>
<td>alkaloids</td>
</tr>
<tr>
<td>pot</td>
<td>tiredness</td>
<td>tek</td>
</tr>
<tr>
<td>sativa</td>
<td>chew</td>
<td>compounds</td>
</tr>
<tr>
<td>blaze</td>
<td>juices</td>
<td>evaporating</td>
</tr>
<tr>
<td>indica</td>
<td>gelatin</td>
<td>atom</td>
</tr>
<tr>
<td>smoking</td>
<td>yogurt</td>
<td>aromatic</td>
</tr>
<tr>
<td>blunts</td>
<td>fruit</td>
<td>non-polar</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>purified</td>
</tr>
</tbody>
</table>

Marijuana! Oral! Chemistry!
Marijuana
- weed
- cannabis
- thc
- marijuana
- stoned
- bowl
- bud
- joint
- blunt
- herb
- bong
- pot
- sativa
- blaze
- indica
- smoking
- blunts
- ...

Oral
- capsules
- consumes
- toast
- stomach
- chewing
- ambien
- digestion
- juice
- absorbed
- ingestion
- meal
- tiredness
- chew
- juices
- gelatin
- yogurt
- fruit
- ...

Chemistry
- solvent
- extraction
- evaporate
- evaporated
- solvents
- evaporation
- yield
- chloride
- alkaloids
- tek
- compounds
- evaporating
- atom
- aromatic
- non-polar
- purified
- jar
- ....
FACTORIAL LDA

DRUG DISCUSSIONS

<table>
<thead>
<tr>
<th>Marijuana</th>
<th>Oral</th>
<th>Chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>weed</td>
<td>capsules</td>
<td>solvent extraction</td>
</tr>
<tr>
<td>cannabis</td>
<td>consumes</td>
<td>extraction</td>
</tr>
<tr>
<td>thc</td>
<td>toast</td>
<td>evaporate</td>
</tr>
<tr>
<td>marijuana</td>
<td>stomach</td>
<td>evaporated</td>
</tr>
<tr>
<td>stoned</td>
<td>chewing</td>
<td>solvents</td>
</tr>
<tr>
<td>bowl</td>
<td>ambien</td>
<td>evaporation</td>
</tr>
<tr>
<td>bud</td>
<td>digestion</td>
<td>yield</td>
</tr>
<tr>
<td>joint</td>
<td>juice</td>
<td>chloride</td>
</tr>
<tr>
<td>blunt</td>
<td>absorbed</td>
<td>alkaloids</td>
</tr>
<tr>
<td>herb</td>
<td>ingestion</td>
<td>tek</td>
</tr>
<tr>
<td>bong</td>
<td>meal</td>
<td>compounds</td>
</tr>
<tr>
<td>pot</td>
<td>tiredness</td>
<td>evaporating</td>
</tr>
<tr>
<td>sativa</td>
<td>chew</td>
<td>atom</td>
</tr>
<tr>
<td>blaze</td>
<td>juices</td>
<td>aromatic</td>
</tr>
<tr>
<td>indica</td>
<td>gelatin</td>
<td>non-polar</td>
</tr>
<tr>
<td>smoking</td>
<td>fruit</td>
<td>purified</td>
</tr>
<tr>
<td>blunts</td>
<td></td>
<td>jar</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>

```
\text{exp}( \\
\begin{array}{c}
\text{Marijuana} \\
\text{Oral} \\
\text{Chemistry}
\end{array} ) = \\
\begin{array}{c}
\text{thc} \\
\text{method} \\
\text{extraction}
\end{array}
```

\begin{array}{c}
\text{plant} \\
\text{material} \\
\text{cannabis}
\end{array}

\begin{array}{c}
\text{simpl}
\end{array}

\begin{array}{c}
\text{coffee} \\
\text{oil}
\end{array}

\begin{array}{c}
\text{contains} \\
\text{jar}
\end{array}

\begin{array}{c}
\text{dried} \\
\text{water}
\end{array}

\begin{array}{c}
\text{extract} \\
\text{results}
\end{array}

\begin{array}{c}
\text{...}
\end{array}
Dirichlet
word distribution for the triple:

(Marijuana Oral Chemistry)

oil  water  butter  thc
weed  hash  cannabis
alcohol  make  milk
high  marijuana  add
...  mixture  hours
try  brownies

~ Dirichlet( )

thc  method  extraction
plant  material  cannabis
simple  coffee  oil
contains  jar
dried  process
dry  water
extract  results
...
~ Dirichlet( )
word distribution for the triple:

Marijuana Oral Chemistry

- oil
- water
- butter
- thc
- weed
- hash
- cannabis
- alcohol
- make
- milk
- high
- marijuana
- add
- ...
- mixture
- hours
- try
- brownies

Dirichlet
Prior for triple \( (i,j,k) \):

\[
\tilde{\phi}_{(i,j,k)v} = \exp(\omega_{iv}^{\text{drug}} + \omega_{jv}^{\text{route}} + \omega_{kv}^{\text{aspect}})
\]

\[
\phi_{(i,j,k)} \sim \text{Dirichlet}(\tilde{\phi}_{(i,j,k)})
\]

distribution over words for this triple

In general, prior for tuple \( t \):

\[
\tilde{\phi}_{tv} = \exp(\sum_{k=1}^{K} \omega_{tv}^{(k)})
\]

\[
\phi_{t} \sim \text{Dirichlet}(\tilde{\phi}_{t})
\]
Prior for triple \((i,j,k)\):

\[
\tilde{\varphi}_{(i,j,k)v} = \exp(\omega_{i\nu}^{\text{(drug)}} + \omega_{j\nu}^{\text{(route)}} + \omega_{k\nu}^{\text{(aspect)}})
\]

\[
\varphi_{(i,j,k)} \sim \text{Dirichlet}(\tilde{\varphi}_{(i,j,k)})
\]

distribution over words for this triple

In general, prior for tuple \(t\):

\[
\tilde{\varphi}_{tv} = \exp\left(\sum_{k=1}^{K} \omega_{tv}^{(k)}\right)
\]

\[
\varphi_{t} \sim \text{Dirichlet}(\tilde{\varphi}_{t})
\]

Document priors:

\[
\tilde{\theta}_{mt} = \exp\left(\sum_{k=1}^{K} \alpha_{mtk}^{(k)}\right)
\]

\[
\theta_{m} \sim \text{Dirichlet}(\tilde{\theta}_{m})
\]
**Marijuana**
- weed
- cannabis
- thc
- marijuana
- stoned
- bowl
- bud
- joint
- blunt
- herb
- bong
- pot
- sativa
- blaze
- indica
- smoking
- blunts
- ...

**Oral**
- capsules
- consumes
- toast
- stomach
- chewing
- ambiен
- digestion
- juice
- absorbed
- ingestion
- meal
- tiredness
- chew
- juices
- gelatin
- yogurt
- fruit
- ...

**Chemistry**
- solvent
- extraction
- evaporate
- evaporated
- solvents
- evaporation
- yield
- chloride
- alkaloids
- tek
- compounds
- evaporating
- atom
- aromatic
- non-polar
- purified
- jar
- ...

Where did these vectors come from?
Weights learned from a supervised model are then used to create a **Gaussian prior** over the FLDA weights:

\[ \sim N(\mu, \sigma^2) \]
We can use this model to extract specific information about new drugs

- e.g. dosage, desired effects, negative effects

“What is the dosage when taking mephedrone orally?”
We can use this model to extract specific information about new drugs

- e.g. dosage, desired effects, negative effects

“What is the dosage when taking mephedrone orally?”
We can use this model to extract specific information about new drugs

- e.g. dosage, desired effects, negative effects

“What is the dosage when taking mephedrone orally?”

Mephedrone Oral Usage

If it is [someone who isn’t you]’s first time using Mephedrone [someone who isn’t me] recommends a 100mg oral dose on an empty stomach.
We can use this model to extract specific information about new drugs

- e.g. dosage, desired effects, negative effects

“What is the dosage when taking mephedrone orally?”

If it is [someone who isn’t you]’s first time using Mephedrone [someone who isn’t me] recommends a 100mg oral dose on an empty stomach.

Reference text:

It is recommended by users that Mephedrone be taken on an empty stomach. Doses usually vary between 100mg – 1g.
We can use this model to extract specific information about new drugs.
FACTORIAL LDA

Components:

- weed
cannabis
thc
marijuana
stoned
bowl
bud
joint
blunt
herb

capsules
consumes
toast
stomach
chewing
ambien
digestion
juice
absorbed
ingestion

- solvent
extraction
evaporate
evaporated
solvents
evaporation
yield
chloride
alkaloids
tek

Topics:

- oil
water
butter
thc
weed
hash
hash
alcohol
make
milk
FACTORIAL LDA
This Cartesian product can be huge!

- And not all triples make sense...

FACTORIAL LDA

SPARSITY
Proposed solution: learn a sparsity pattern

\[ \tilde{\theta}_{m_t} = b_{\tilde{t}} \exp(\sum_{k=1}^{K} \alpha_{m_t}^{(k)}) \quad b_{\tilde{t}} \in (0,1); \ b_{\tilde{t}} \sim \text{Beta}(\rho < 1) \]
<table>
<thead>
<tr>
<th>“Topic”</th>
<th>“Approach”</th>
<th>“Focus”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“SPEECH”</td>
<td>“I.R.”</td>
<td>“M.T.”</td>
</tr>
<tr>
<td>speech</td>
<td>document</td>
<td>translation</td>
</tr>
<tr>
<td>spoken</td>
<td>retrieval</td>
<td>machine</td>
</tr>
<tr>
<td>recognition</td>
<td>documents</td>
<td>source</td>
</tr>
<tr>
<td>state</td>
<td>question</td>
<td>mt</td>
</tr>
<tr>
<td>vocabulary</td>
<td>web</td>
<td>parallel</td>
</tr>
<tr>
<td>recognizer</td>
<td>answering</td>
<td>french</td>
</tr>
<tr>
<td>utterances</td>
<td>query</td>
<td>bilingual</td>
</tr>
<tr>
<td>synthesis</td>
<td>answer</td>
<td>transfer</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>“EMPIRICAL”</td>
<td>“THEORETICAL”</td>
</tr>
<tr>
<td>task</td>
<td>theory</td>
<td>word</td>
</tr>
<tr>
<td>tasks</td>
<td>description</td>
<td>algorithm</td>
</tr>
<tr>
<td>performance</td>
<td>formal</td>
<td>method</td>
</tr>
<tr>
<td>improve</td>
<td>forms</td>
<td>accuracy</td>
</tr>
<tr>
<td>accuracy</td>
<td>treatment</td>
<td>best</td>
</tr>
<tr>
<td>learning</td>
<td>linguistics</td>
<td>sentence</td>
</tr>
<tr>
<td>demonstrate</td>
<td>syntax</td>
<td>statistical</td>
</tr>
<tr>
<td>using</td>
<td>ed</td>
<td>natural</td>
</tr>
<tr>
<td></td>
<td></td>
<td>development</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>“Focus”</th>
<th>“SPEECH”</th>
<th>“APPL.”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“METHODS”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b=0.20)</td>
<td></td>
</tr>
<tr>
<td>EMPirical</td>
<td>dialogue</td>
<td>spoken</td>
</tr>
<tr>
<td></td>
<td>spoken</td>
<td>speech</td>
</tr>
<tr>
<td></td>
<td>speech</td>
<td>dialogues</td>
</tr>
<tr>
<td></td>
<td>understanding</td>
<td>task</td>
</tr>
<tr>
<td></td>
<td>recognition</td>
<td>recognition</td>
</tr>
</tbody>
</table>

|                | “METHODS”      |               |
|                | (b=0.99)       |               |
| THEORETICAL    | speech         | words         |
|                | words          | recognition   |
|                | prosodic       | written       |
|                | phonological   | spoken        |

<table>
<thead>
<tr>
<th>“Focus”</th>
<th>“DATA”</th>
<th>“APPL.”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“METHODS”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b=1.00)</td>
<td></td>
</tr>
<tr>
<td>EMPirical</td>
<td>corpus</td>
<td>data</td>
</tr>
<tr>
<td></td>
<td>data</td>
<td>corpus</td>
</tr>
<tr>
<td></td>
<td>training</td>
<td>annotation</td>
</tr>
<tr>
<td></td>
<td>model</td>
<td>annotated</td>
</tr>
<tr>
<td></td>
<td>tagging</td>
<td>test</td>
</tr>
<tr>
<td></td>
<td>annotated</td>
<td>collection</td>
</tr>
<tr>
<td></td>
<td>test</td>
<td>xml</td>
</tr>
</tbody>
</table>

|                | “METHODS”    |               |
|                | (b=0.00)     |               |
| THEORETICAL    | rules        | rule          |
|                | rule         | model         |
|                | model        | shown         |
|                | shown        | models        |
|                | models       | right         |

<table>
<thead>
<tr>
<th>“Focus”</th>
<th>“MODELING”</th>
<th>“APPL.”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“METHODS”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b=0.50)</td>
<td></td>
</tr>
<tr>
<td>EMPirical</td>
<td>models</td>
<td>approach</td>
</tr>
<tr>
<td></td>
<td>model</td>
<td>shown</td>
</tr>
<tr>
<td></td>
<td>approach</td>
<td>error</td>
</tr>
<tr>
<td></td>
<td>shown</td>
<td>statistical</td>
</tr>
</tbody>
</table>

|                | “METHODS”    |               |
|                | (b=0.07)     |               |
| THEORETICAL    | rules        | rule          |
|                | rule         | model         |
|                | model        | shown         |
|                | shown        | models        |
|                | models       | right         |

<table>
<thead>
<tr>
<th>“Focus”</th>
<th>“GRAMMAR”</th>
<th>“APPL.”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>“METHODS”</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b=0.01)</td>
<td></td>
</tr>
<tr>
<td>EMPirical</td>
<td>grammar</td>
<td>parsing</td>
</tr>
<tr>
<td></td>
<td>parsing</td>
<td>syntactic</td>
</tr>
<tr>
<td></td>
<td>syntactic</td>
<td>tree</td>
</tr>
<tr>
<td></td>
<td>tree</td>
<td>parse</td>
</tr>
<tr>
<td></td>
<td>parse</td>
<td>dependency</td>
</tr>
<tr>
<td></td>
<td>dependency</td>
<td>treebank</td>
</tr>
</tbody>
</table>

|                | “METHODS”    |               |
|                | (b=0.57)     |               |
| THEORETICAL    | grammar      | parsing       |
|                | parsing      | grammars      |
|                | grammars     | formalism     |
|                | formalism    | parsing       |
|                | parsing      | based         |
|                | based        | efficient     |
|                | efficient    | unification   |
“Surgery”
surgery
pain
went
dr
surgeon
told
procedure
months
performed
removed
left
fix
said
later
years

“Family”
dr
best
children
years
kids
cares
hes
care
old
daughter
child
husband
family
pediatrician
trust

---

TOPIC STRUCTURES

(SPARSE) DAG

---

66
Structured-prior topic models

A family of topic models in which the Dirichlet priors are functions of underlying components

The priors over distributions are weighted combinations of components:

$$\tilde{\phi}_{tv} = \exp\left(\sum_{c=1}^{C(\phi)} \beta_{tc} \omega_{cv}\right)$$

$$\phi_t \sim \text{Dirichlet}(\tilde{\phi}_t)$$

distribution over words in $i$th topic
The priors over distributions are weighted combinations of components:

$$\tilde{\theta}_{mt} = \exp(\sum_{c=1}^{C(\theta)} \alpha_{mc} \delta_{ct})$$

$$\theta_m \sim \text{Dirichlet}(\tilde{\theta}_m)$$

distribution over topics in $m$th document
The priors over distributions are weighted combinations of components:

\[
\exp(\beta_{11} \times \omega_1 + \beta_{12} \times \omega_2 + \beta_{13} \times \omega_3 + \beta_{14} \times \omega_4 + \beta_{15} \times \omega_5)
\]
The priors over distributions are weighted combinations of components:

\[ \exp(\beta_1 \times \omega_1 + \beta_2 \times \omega_2 + \beta_3 \times \omega_3 + \beta_4 \times \omega_4 + \beta_5 \times \omega_5) \]
The priors over distributions are weighted combinations of components:

We can induce different structures by constraining the values of $\beta$. 
The priors over distributions are weighted combinations of components:
The priors over distributions are weighted combinations of components:

$$\exp(1 \times \omega_1 + 0 \times \omega_2 + 0 \times \omega_3 + 0 \times \omega_4 + 0 \times \omega_5)$$
The priors over distributions are weighted combinations of components:

\[ \exp(0 \times \omega_1 + 0 \times \omega_2 + 1 \times \omega_3 + 0 \times \omega_4 + 0 \times \omega_5) \]
The priors over distributions are weighted combinations of components:

Tree: Each topic’s $\beta$ vector is zero in all but one component
The priors over distributions are weighted combinations of components:

**Factorization:** Like a tree, but a nonzero component in each factor
\( \beta_t \) has one value of 1 in each factor; 0 elsewhere

Factorial LDA:

\[
\tilde{\phi}_{tv} = \exp\left( \sum_{c=1}^{C(\phi)} \beta_{tc} \omega_{cv} \right)
\]

\[
\tilde{\theta}_{mt} = \exp\left( \sum_{c=1}^{C(\theta)} \alpha_{mc} \delta_{ct} \right)
\]

\( \delta_c \) is transpose of \( \beta_t \)
• Initial solution: learn a sparsity pattern
SPRITE

SPECIAL CASE: FLDA
**SPRITE** generalizes many existing topic models:

<table>
<thead>
<tr>
<th>Model</th>
<th>Document priors</th>
<th>Topic priors</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>Single component</td>
<td>Single component</td>
</tr>
<tr>
<td>SCTM</td>
<td>Single component</td>
<td>Sparse binary ( \beta )</td>
</tr>
<tr>
<td>SAGE</td>
<td>Single component</td>
<td>Sparse ( \omega )</td>
</tr>
<tr>
<td>FLDA</td>
<td>Binary ( \delta ) is transpose of ( \beta )</td>
<td>Factored binary ( \beta )</td>
</tr>
<tr>
<td>PAM</td>
<td>( \alpha ) are supertopic weights</td>
<td>Single component</td>
</tr>
<tr>
<td>DMR</td>
<td>( \alpha ) are feature values</td>
<td>Single component</td>
</tr>
</tbody>
</table>
SPRITE

PARAMETER ESTIMATION

Need to estimate values for the parameters:

- Word and topic distributions
  - Collapsed Gibbs sampling
- Component parameters (i.e. $\beta$, $\omega$)
  - Gradient ascent
What if $\beta$ has constraints?

\[
\beta_{tc} \in \{0, 1\}, \forall c \\
\sum_c \beta_{tc} = 1
\]
Relaxing the constraints

\[
\beta_{tc} \in \{0,1\}, \forall c
\]
\[
\sum_c \beta_{tc} = 1
\]
Relaxing the constraints

\[ \beta_{tc} \in \{0,1\}, \forall c \]
\[ \sum_c \beta_{tc} = 1 \]

\[ \bar{\beta}_t \sim \text{Dirichlet}(\rho < 1) \]

\[ \beta_{tc} \in (0,1), \forall c \]
\[ \sum_c \beta_{tc} = 1 \]

Sparsity-encouraging prior distribution
Relaxing the constraints

\[ \beta_{tc} \in \{0,1\}, \forall c \]
\[ \sum_c \beta_{tc} = 1 \]

Tightening the constraints by increasing the prior

\[ \tilde{\beta}_t \sim \text{Dirichlet}(\rho < 1) \]

Sparsity-encouraging prior distribution

\[ \beta_{tc} \in (0,1), \forall c \]
\[ \sum_c \beta_{tc} = 1 \]
Modeling perspective in text:

“What opinions are people tweeting about gun control?”

Benton, Paul, Hancock, Dredze. *A structured model of topic and perspective in social media.* In preparation.
Suppose we want to model how perspective influences topics

- e.g. certain topics are “pro” or “anti” gun control

A single-component SPRITE model:

\[ \tilde{\phi}_{tv} = \exp(r_t \omega_v) \]

The \(t\)th topic’s perspective association

The \(v\)th word’s perspective association
Suppose we want to model how perspective influences topics

- e.g. certain topics are “pro” or “anti” gun control

A single-component SPRITE model:

\[
\tilde{\phi}_{tv} = \exp(r_t \omega_v)
\]

The \(v\)th word’s perspective association

\[
\tilde{\theta}_{mt} = \exp(\alpha_m r_t)
\]

The \(m\)th tweet’s perspective association

The \(t\)th topic’s perspective association
Suppose we want to model how perspective influences topics

- e.g. certain topics are “pro” or “anti” gun control

A single-component SPRITE model:

\[ \tilde{\phi}_{tv} = \exp(r_t \omega_v) \]

The \( v \)th word’s perspective association

\[ \tilde{\theta}_{mt} = \exp(\alpha_m r_t) \]

The \( t \)th topic’s perspective association

Prior is a function of:

- Hashtags (#GunControlNow vs #NoGunControl)
- Survey data (% gun ownership in each state)
The model can help estimate the % gun ownership in US states:

- Mean error: 8.4
- Baselines: 12.7 – 16.4
Suppose we want to model perspective and we want to organize topics in a hierarchy.
Suppose we want to model perspective and we want to organize topics in a hierarchy.

\[
\tilde{\phi}_{tv} = \exp(r_t \omega_{0v} + \sum_{c=1}^{C(\phi)} \beta_{tc} \omega_{cv})
\]

with (soft) constraints that \(\beta_i\) is an indicator vector.
• Organizes topics in a variety of useful ways
  • Can be tailored toward different applications
• Generalizes many topic models
  • While opening up new possibilities
• Allows practitioners to make sense of big text data
  • Can drive new scientific research
There’s no end to exciting questions we can ask of big, open data

We need methods to understand what people are saying on the web and learn meaningful trends

This requires models that can discover patterns automatically, while accommodating user expectations
Thank you with help from:

Committee: Mark Dredze, Jason Eisner, Eric Horvitz, Hanna Wallach
Funding: Microsoft Research, NSF, JHU Dean’s office

Public opinion:
• Adrian Benton
• Braden Hancock

Doctor reviews:
• Byron Wallace
• Urmimala Sarkar
• Thomas Trikalinos

Drug forums:
• Meg Chisolm
• Matthew Johnson
• Ryan Vandrey
THANK YOU

QUESTIONS?