COMP 330: Latent Dirichlet Allocation

Chris Jermaine and Kia Teymourian

Rice University

We've Talked about LSI

Idea:

- Map each doc to a position in a latent space
- Such that position has some semantic meaning
- Search using distance in latent space

"LDA" is a Bayesian variant on this idea

- Except we explicitly give meaning to latent space
- Dim in latent space corresponds to importance of a "topic" in the doc

Basic LDA Idea (From Orig Paper)

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

Generative Process

for t = 1 to k **do:**
  $\phi_t \sim \text{Dirichlet}(\beta)$

- First, generate the topic vectors
- Each is a vector of probs, of len w
- Where w is num of words in corpus
- Each entry tells us importance of word in topic

Generative Process

```
for t = 1 to k do:
    φ_t ~ Dirichlet(β)
for i = 1 to n do:
    θ_i ~ Dirichlet(α)
```

- Next, gen the importance of each topic in each doc
- Each is a vector of probs, of len k
- Each entry tells us importance of topic in doc

Generative Process

```
for t = 1 to k do:
   φ_t ~ Dirichlet(β)
for i = 1 to n do:
   θ_i ~ Dirichlet(α)
   for j = 1 to len (doc i)
      topic z_{i,j} ~ Categorical(θ_i)
      word w_{i,j} ~ Categorical(φ_{z_{i,j}})
```

- Next, gen all of the words
- First choose a topic
- Then use the topic to gen the word

How to Learn the Model?

Gibbs Sampling!!

- First step in deriving the algorithm?
- As always: write down the PDF!!

How to Learn the Model?

- The PDF is simply:

$$\prod_t \text{Dirichlet}(\phi_t|\beta)\times$$

$$\prod_i \text{Dirichlet}(\theta_i|\alpha)\times$$

$$\prod_{i,j} \text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Gibbs Sampler Outline

As usual, basic algorithm is:

```
initialize each φ_t, θ_i, z_{i,j}
for iter = 1 to big do:
  for t = 1 to k do:
    update φ_t
  for i = 1 to n do:
    update θ_i
    for j = 1 to len (doc i) do:
      update z_{i,j}
```

Now Let's Go Thru Updates

Remember: to update a variable x, drop all terms in PDF w/o it

This gives you a function $f(x) \propto f(x| \textbf{everything else})$

Which is what you need for Gibbs sampling

Updating Phi For a Given t

- The PDF is simply:

$$\prod_t \text{Dirichlet}(\phi_t|\beta)\times$$
$$\prod_i \text{Dirichlet}(\theta_i|\alpha)\times$$
$$\prod_{i,j} \text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Updating the Phi For a Given t

- The PDF is simply:

$$\text{Dirichlet}(\phi_t|\beta)\times \prod_{i,j \text{ where } z_{i,j}=t} \text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Updating the Phi For a Given t

- The PDF is simply:

$$\text{Dirichlet}(\phi_t|\beta)\times \prod_{i,j \text{ where } z_{i,j}=t} \text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- Fortunately, Dirichlet it conjuagate prior for Categorical
- From Wikipedia, can have $\phi_t \sim \text{Dirichlet}(\beta + cnt)$ where cnt is a vector with cnt_w the total number of times topic t produced word w

Updating Theta For a Given i

- The PDF is simply:

$$\prod_t \text{Dirichlet}(\phi_t|\beta)\times$$

$$\prod_i \text{Dirichlet}(\theta_i|\alpha)\times$$

$$\prod_{i,j} \text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Updating Theta For a Given i

- The PDF is simply:

$$\text{Dirichlet}(\theta_i|\alpha) \times$$

$$\prod_{i,j} \text{Categorical}(z_{i,j}|\theta_i)$$

Updating Theta For a Given i

- The PDF is simply:

$$\text{Dirichlet}(\theta_i|\alpha)\times$$
$$\prod_{i,j}\text{Categorical}(z_{i,j}|\theta_i)$$

- Dirichlet is still conjuage for Categorical!

- From Wikipedia, can have $\theta_i \sim \text{Dirichlet}(\alpha + cnt)$ where cnt is a vector with cnt_t the total number of words in doc i produced by topic t

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\prod_t \text{Dirichlet}(\phi_t|\beta)\times$$

$$\prod_i \text{Dirichlet}(\theta_i|\alpha)\times$$

$$\prod_{i,j} \text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- From Bayes' rule, we have:

$$Pr[z_{i,j} = t|\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}] = \frac{p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t)}{p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j})}$$

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- From Bayes' rule, we have:

$$\begin{aligned} Pr[z_{i,j} = t|\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}] &= \frac{p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t)}{p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j})} \\ &\propto p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t) \end{aligned}$$

- (since $z_{i,j}$ does not appear in denominator)

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- From Bayes' rule, we have:

$$\begin{aligned} Pr[z_{i,j} = t|\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}] &= \frac{p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t)}{p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j})} \\ &\propto p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t) \\ &= p(z_{i,j} = t)p(w_{i,j}|z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors}) \end{aligned}$$

- (since generation of θ_i, all ϕ vectors independent of $z_{i,j}$)

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- From Bayes' rule, we have:

$$\begin{aligned}
Pr[z_{i,j} = t|\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}] &= \frac{p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t)}{p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j})} \\
&\propto p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t) \\
&= p(z_{i,j} = t)p(w_{i,j}|z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors}) \\
&\propto p(z_{i,j} = t)p(w_{i,j}|z_{i,j} = t)
\end{aligned}$$

- (since no $z_{i,j}$ in these terms)

Updating Topic Assignment For jth Word in Doc i

- The PDF is simply:

$$\text{Categorical}(z_{i,j}|\theta_i)\text{Categorical}(w_{i,j}|\phi_{z_{i,j}})$$

- From Bayes' rule, we have:

$$\begin{aligned}
Pr[z_{i,j} = t|\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}] &= \frac{p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t)}{p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j})} \\
&\propto p(z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors }, w_{i,j}|z_{i,j} = t) \\
&= p(z_{i,j} = t)p(w_{i,j}|z_{i,j} = t)p(\theta_i, \text{all } \phi \text{ vectors}) \\
&\propto p(z_{i,j} = t)p(w_{i,j}|z_{i,j} = t) \\
&= \theta_t \times (w_{i,j}\text{th entry in } \phi_t)
\end{aligned}$$

- So, to update $z_{i,j}$, compute $\theta_t \times (w_{i,j}\text{th entry in } \phi_t)$ for each t
- Then normalize to get k different probabilities
- And use that probability vector to choose the responsible topic

Questions?