Johns Hopkins
Whiting School of Engineering

Topic Modeling (LDA) pt 2

2/2/26

Recap

- Last class:
  - Topic modeling (LDA)
    - Model formulation
    - Gibbs sampling
    - Practical considerations
- Today
  - Topic model (LDA)
    - Variational Inference
    - Limitations and extensions
    - Example application: Structured topic model and media manipulation

LDA Generative Story

- For each topic k:
  - Draw $\beta_k \sim Dir(\eta)$
- For each document d:
  - Draw $\theta_d \sim Dir(\alpha)$
  - For each word in d:
    - Draw topic assignment $z \sim Multinomial(\theta_d)$
    - Draw $w \sim Multinomial(\beta_z)$

We use the data to estimate these two sets of parameters:

- β, a distribution over vocabulary (1 for each topic)
- θ, a distribution over topics (1 for each document)

Definitions

	Topic 1	Topic 2	…	Topic 30
administration	0.01	**0.12**	…	0.02
advertising	0.02	0.001	…	**0.25**
debt	0.1	0.001	…	0.01
…	…	…	…	…
government	0.01	**0.15**	…	0.01
…	…	…	…	…
spending	**0.12**	0.01	…	0.03
taxes	**0.15**	0.02	…	**0.35**
trillion	**0.19**	0.003	…	0.02

Each "topic" is defined by $\boldsymbol{\beta}$, a multinomial distribution over the entire vocabulary

	Doc 1	Doc 2	…	Doc N
Topic 1	0.10	**0.60**	…	
Topic 3	0.02	0.05	…	
Topic 4	**0.30**	0.1	…	
…	…	…	…	…
Topic 15	**0.20**	0.01	…	**0.40**
…	…	…	…	…
Topic 28	0.01	0.03	…	**0.20**
Topic 29	**0.25**	0.15	…	
Topic 30	0.03	0.01	…	

Each document has associated $\boldsymbol{\theta}$, a multinomial distribution over topics

Review: Bayesian Inference

- Goal: estimate θ, β
- Bayesian approach: we estimate full posterior distribution

$$p(\theta, \beta, z \mid w) = \frac{p(w \mid \theta, \beta, z) p(\theta, \beta, z)}{p(w)}$$

$p(w)$ is intractable!

Gibbs Sampling:

- We generate samples from the posterior distribution
- We estimate θ, β from those samples

Alternative approach: Variational Inference

Variational Inference

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Variational Inference: Key ideas

$$\underbrace{p(\theta, \beta, z}_{z} \,|\, \underbrace{w}_{x})$$

- We create a distribution *q* that approximates *p* but is easier to work with
  - Pick a family of distributions (Q) over the latent variables with its own *variational parameters*
  - Find the setting of the parameters that makes *q* close to the posterior of interest
  - Use *q* with the fitted parameters as a proxy for the posterior

Variational Inference: Compared to Gibbs sampling

- Pros:
  - Deterministic, easy to determine convergence, requires fewer iterations (faster, especially for large data)
  - Doesn't require conjugacy
- Cons:
  - Overall relative accuracy is not known, but Gibbs sampling potentially works better
    - Has guarantees of producing (asymptotically) exact samples from the target density (Robert and Casella, 2004)
    - Anecdotally people have observed Gibbs sampling yields better topics[1]
  - Math is more difficult, Gibbs sampling is often easier to debug

[1]https://maria-antoniak.github.io/2022/07/27/topic-modeling-for-the-people.html

Variational Inference

- We want to approximate $p(z|x)$

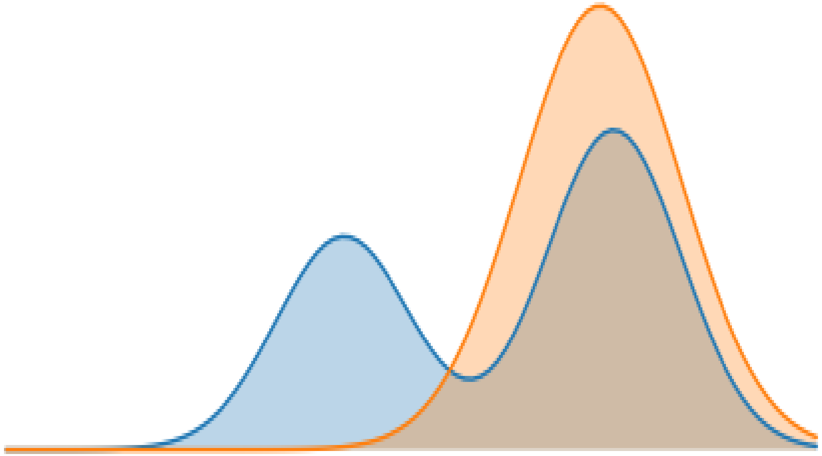
- Define variational distribution $q(z|v)$
  - Find v so that $q(z|v)$ is close to $p(z|x)$

- How do we define "close to"?

Kullback–Leibler (KL) divergence

- $KL\,(q(z)||p(z|x)) = E_q[log\,\frac{q(z)}{p(z|x)}]$

- Characterization
  - q and p are high 👍
  - q is high and p is low 👎
  - q is low 👍



p = blue
q = orange

Image credit: https://dibyaghosh.com/blog/probability/kldivergence.html

Kullback–Leibler (KL) divergence

- How do we minimize $KL\ (q(z)||p(z|x) = E_q[log\frac{q(z)}{p(z|x)}]$?

$$KL\ (q(z)||p(z|x) = E_q[\log(q(z)) - \log(p(z|x))]$$

$$= E_q[\log(q(z))] - E_q[\log(\frac{p(z,x)}{p(x)})]$$

$$= E_q[\log(q(z))] - E_q[\log(p(x,z))] + \log(p(x))$$

$$= -\underbrace{(E_q[\log(p(x,z))] - E_q\ [\log(q(z))])}_{\text{"ELBO" Maximizing this minimizes KL divergence}} + \underbrace{\log(p(x))}_{\text{This is the value we can't estimate}}$$

"ELBO"
Maximizing this minimizes
KL divergence

This is the value
we can't estimate

The evidence lower bound (ELBO)

- $\log(p(x)) = \log \int_z \; p(x,z)$

$$= \log \int_z \; p(x,z) \frac{q(z)}{q(z)}$$

$$= \log(E_q[\frac{p(x,z)}{q(z)}])$$

$$\geq (E_q[log \frac{p(x,z)}{q(z)}])$$

$$\geq \underbrace{E_q[\log(p(x,z))] - E_q[\log(q(z))]}_{\text{“ELBO”}}$$

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Recap

- We want to approximate $p(z|x)$

- Define variational distribution $\mathrm{q}(z|v)$
  - Find v so that $\mathrm{q}(z|v)$ is close to $p(z|x)$
    - i.e. so that $KL\ (q(z|v)||p(z|x))$ is low
    - i.e. so that $E_q[\log(p(x,z))]$ - $E_q[\log(q(z))]$ is high

Mean Field Variational Inference

- We assume that the variational distribution factorizes

$$q(z_1, \ldots z_m) = \prod_{j=1}^{m} q(z_j)$$

- Finally, getting back to LDA, we can define separate a q for θ, β, z

- [Latent variables actually probably are dependent, so this won't contain the true posterior]

Choose q

$$p(\theta, \beta, z \mid w) = \frac{p(w \mid \theta, \beta, z) p(\theta, \beta, z)}{p(w)}$$

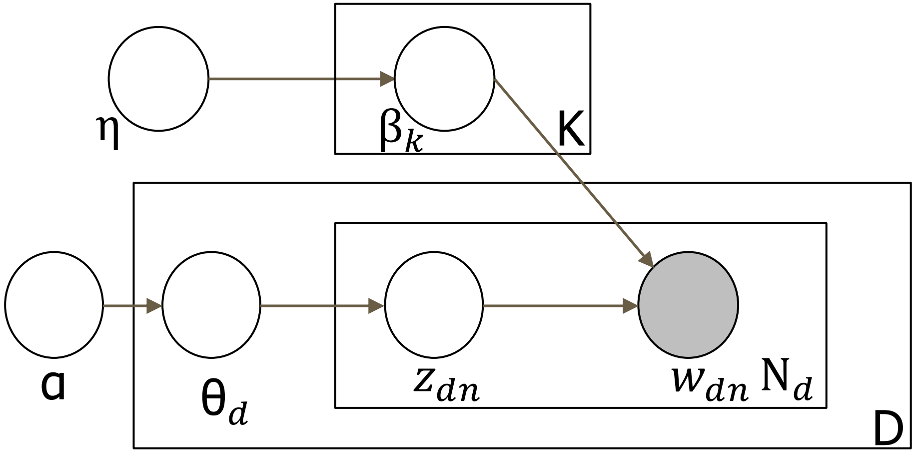
- Choose:

$$q(\theta, \beta, z) = \prod_{i=1}^{K} q(\beta_i | \lambda_i) \prod_{d=1}^{D} q_d(\theta_d, z_d | \gamma_d, \phi_d)$$

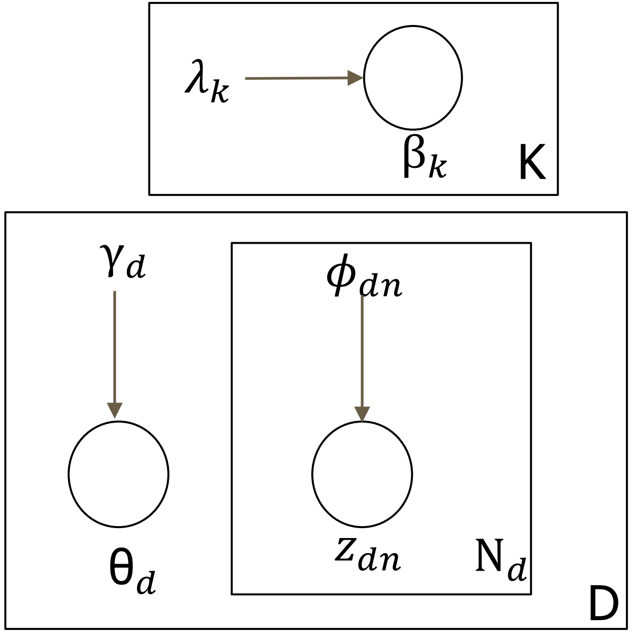
where $q_d(\theta, z) = q(\theta|\gamma) \prod_n^N q(z_n|\phi_n)$

- Assume that:
  - $q(\beta|\lambda)$ is a Dirichlet distribution with variational parameters λ
  - $q(\theta|\gamma)$ is a Dirichlet distribution with variational parameters γ
  - $q(z_n|\phi_n)$ is a multinomial (categorial) distribution with variational parameters ϕ_n

Choose q



η
β_k
K
α
θ_d
z_{dn}
w_{dn}
N_d
D



λ_k
β_k
K
γ_d
ϕ_{dn}
θ_d
z_{dn}
N_d
D

Optimize q

This is like Gibbs sampling!

- Common approach: use *coordinate ascent* to optimize
  - Update the variational parameters one at a time
  - At each update, we chose the value of the parameter that maximizes the ELBO (holding other variational parameters constant)
- With our choice of q, we can compute closed-form updates by taking derivatives of the ELBO and setting them to 0

$$\phi \propto \eta_{w_n} \exp\{E_q[\log(\theta_i)|\gamma]\}$$

$q(z_n|\phi_n)$ Topic assignments for each word

$$\gamma_{di} = \alpha_i + \sum_{n=1}^{N} \phi_{dni}$$

$q(\theta|\gamma)$ topic vector for each document

$$\lambda_{iv} = \eta + \sum_{d=1}^{D} \sum_{n=1}^{N_d} \phi_{dni}\,, where\ w_{dn} = v$$

$q(\beta|\lambda)$, distributions over vocabulary

Full procedure

- Choose q

- For each iteration
  - For each variational parameter
    - Update the parameter to maximize the ELBO
- End at convergence

[Use q to approximate posterior: we can take expectations of q to estimate parameters]

Popular LDA packages

- gensim
  - Python: https://radimrehurek.com/gensim/index.html
  - Variational inference

- Mallet
  - Java: https://mimno.github.io/Mallet/topics.html
  - Python wrapper: https://github.com/maria-antoniak/little-mallet-wrapper
  - Gibbs Sampling

LDA Extensions

Problem 1: Topic Correlations

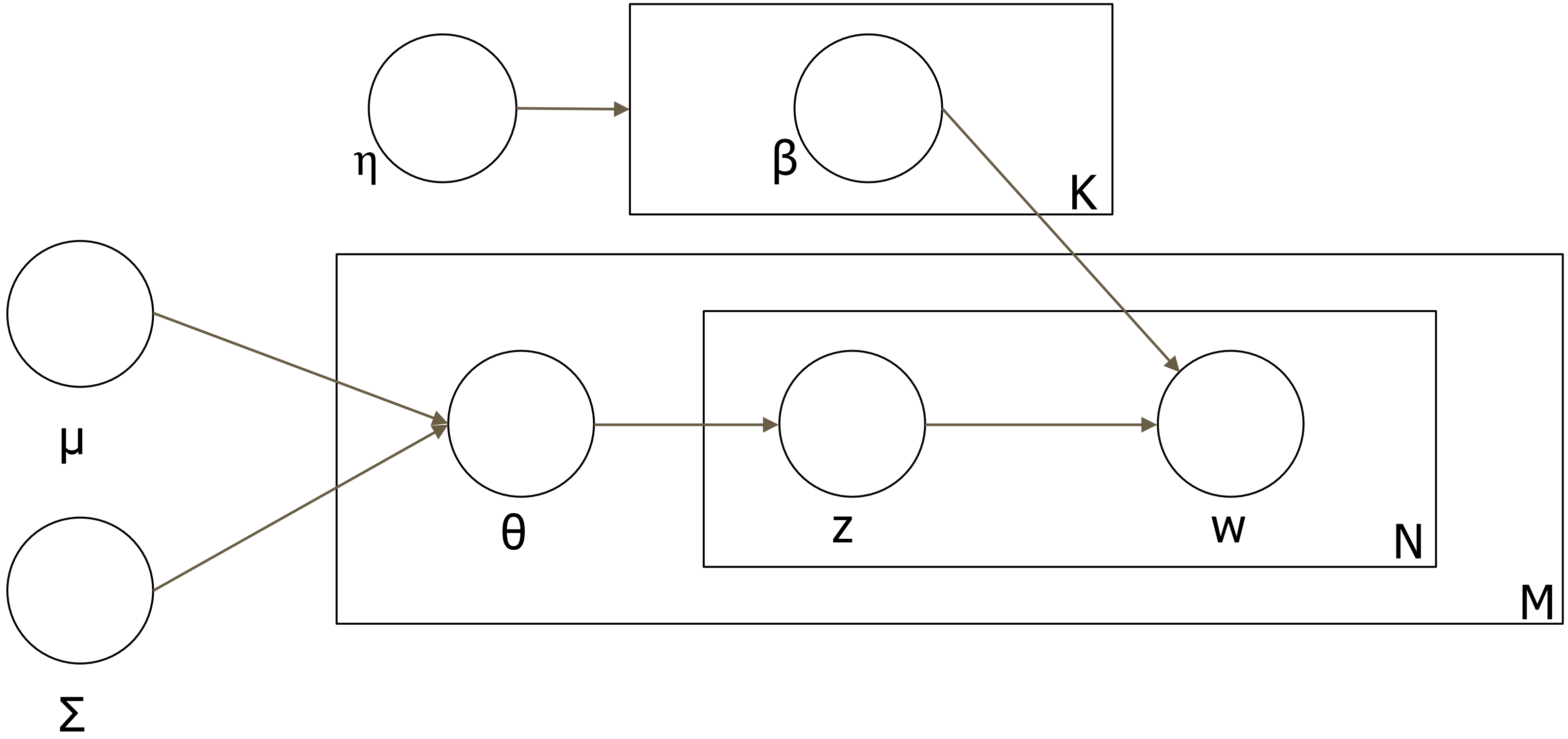
- LDA
  - In a vector drawn from a Dirichlet distribution (θ), elements are nearly independent

- Reality
  - A document about biology is more likely to also be about chemistry than skateboarding

LDA Generative Story

- For each topic k:
  - Draw $\beta_k \sim Dir(\eta)$
- For each document d:
  - Draw $\cancel{\theta_d \sim Dir(\alpha)}$ Draw $g_D \sim N(\mu, \Sigma); \theta_D = f(g_D)$ Σ = Topic covariance matrix
  - For each word in d:
    - Draw topic assignment $z \sim Multinomial(\theta_d)$
    - Draw $w \sim Multinomial(\beta_z)$

- β, a distribution over vocabulary (1 for each topic)
- θ, a distribution over topics (1 for each document)


η
β
K
μ
θ
z
w
N
M
Σ

Example application: Structured topic model and media manipulation

Motivating application: Communications theory of media manipulation

- Communications scholarship on media influence:

- "the media may not be successful much of the time in telling people what to think, but is stunningly successful in telling its readers what to think about" [Cohen, 1963]

- Given a corpus of newspaper articles, we can determine how it may be influencing public opinion by analyzing changes in topic coverage
  - We don't know exactly what topics are in advance: we need to be able to discover them from the corpus

Motivating application: Communications theory of media manipulation

- Agenda setting
  - ***What*** topics are covered
- Framing
  - ***How*** topics are covered
- Priming
  - What effect reporting has on public opinion
  - "Framing works to shape and alter audience members' interpretations and preferences through priming"

Entman's thesis: we can use this framework to understand bias in the media

"agenda setting, framing and priming fit together as *tools of power*"

Entman, Robert M. "Framing bias: Media in the distribution of power." *Journal of communication* 57.1 (2007): 163-173.

Motivating application: Communications theory of media manipulation

- Further refine framing definition: "process of culling a few elements of perceived reality and assembling a narrative that highlights connections among them to promote a particular interpretation" [Entman, 2007]
- Topic Level
  - Abortion is a moral issue
  - Abortion is health issue
  - [This should remind you agenda setting]
- Word Level
  - "Estate tax" vs. "Death tax"

Framing: Additional Background

- Equivalence: different presentations of logically-identical information (Scheufele and Iyengar, 2012)
- *Emphasis:* "qualitatively different yet potentially relevant considerations" (Chong and Druckman, 2007, p.114)

	Issue-Specific	Issue-generic
Equivalence		
Emphasis		

Mendelsohn, Julia, Ceren Budak, and David Jurgens. "Modeling Framing in Immigration Discourse on Social Media." NAACL. 2021.
Card, Dallas, et al. "The media frames corpus: Annotations of frames across issues." ACL. 2015.

Framing: Additional Background

Topic Model?

General taxonomy; classification models

	Issue-Specific	Issue-generic
Equivalence	90% unemployment vs. 10% employment	90% vs. 10%
Emphasis	Immigration: hero/worker vs. threat/job security	Morality, Economy, Security and Defense

Media Frames Corpus – issue-generic policy frames

Mendelsohn, Julia, Ceren Budak, and David Jurgens. "Modeling Framing in Immigration Discourse on Social Media." NAACL. 2021.
Card, Dallas, et al. "The media frames corpus: Annotations of frames across issues." ACL. 2015.

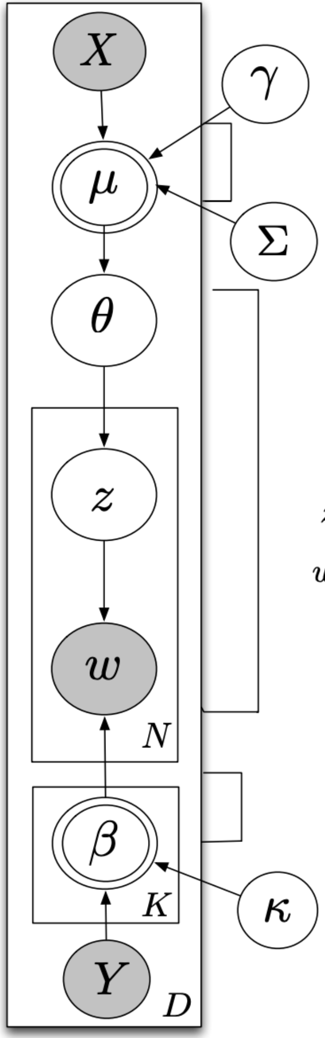
Problem: LDA assumptions conflict with analysis goals

- LDA
  - The topic distributions (θ) are drawn from the same distribution Dir(α) for all documents
- Reality
  - We often use LDA to look at how topics vary across documents
  - Example
    - We run LDA on a corpus of Democratic/Republican speeches.
    - Look at topic prevalence in Republican speeches and Democratic speeches
    - Conclude Republicans talk about taxes more than Democrats
  - But we've assumed that all speeches are drawing topics the same way
  - We need more LDA extensions

Solution: Structured Topic Model

- Topical prevalence: the proportion of document devoted to a given topic
  - X - matrix of covariate information
  - Useful for *agenda setting*

- Topical content: the rate of word use within a given topic
  - Y - matrix of covariate information
  - Useful for *framing*



Topic Prevalence:

$$\mu_{d,k} = X_d \gamma_k$$
$$\gamma_k \sim \mathcal{N}(0, \sigma_k^2)$$
$$\sigma_k^2 \sim \text{Gamma}(s^\gamma, r^\gamma)$$

Language Model:

$$\theta_d \sim \text{LogisticNormal}(\mu_d, \Sigma)$$
$$z_{d,n} \sim \text{Mult}(\theta_d)$$
$$w_{d,n} \sim \text{Mult}(\beta_d^{k=z_{d,n}})$$

Topical Content:

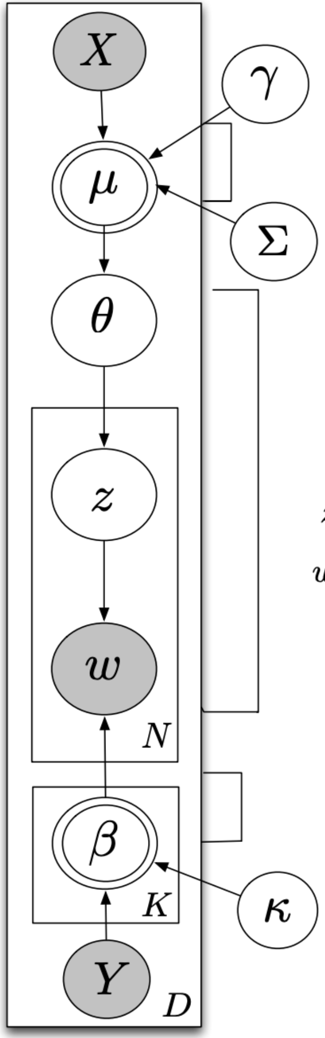
$$\beta_{d,v}^k \propto \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{y,\cdot} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim \text{Laplace}(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim \text{Gamma}(s^\kappa, r^\kappa)$$

Roberts, Margaret E., et al. "The structural topic model and applied social science." *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*. Vol. 4. No. 1. 2013.

Solution: Structured Topic Model

- X could be Democrat/Republican as well as date of speech
  - Captures that Republicans talk more about *taxes* but rate varies by year

- Y could be Democrat/Republican
  - Captures that Democrats focus on social benefits and Republicans focus on government imposition



Topic Prevalence:

$$\mu_{d,k} = X_d \gamma_k$$
$$\gamma_k \sim \mathcal{N}(0, \sigma_k^2)$$
$$\sigma_k^2 \sim \text{Gamma}(s^\gamma, r^\gamma)$$

Language Model:

$$\theta_d \sim \text{LogisticNormal}(\mu_d, \Sigma)$$
$$z_{d,n} \sim \text{Mult}(\theta_d)$$
$$w_{d,n} \sim \text{Mult}(\beta_d^{k=z_{d,n}})$$

Topical Content:

$$\beta^k_{d,v} \propto \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{y,\cdot} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim \text{Laplace}(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim \text{Gamma}(s^\kappa, r^\kappa)$$

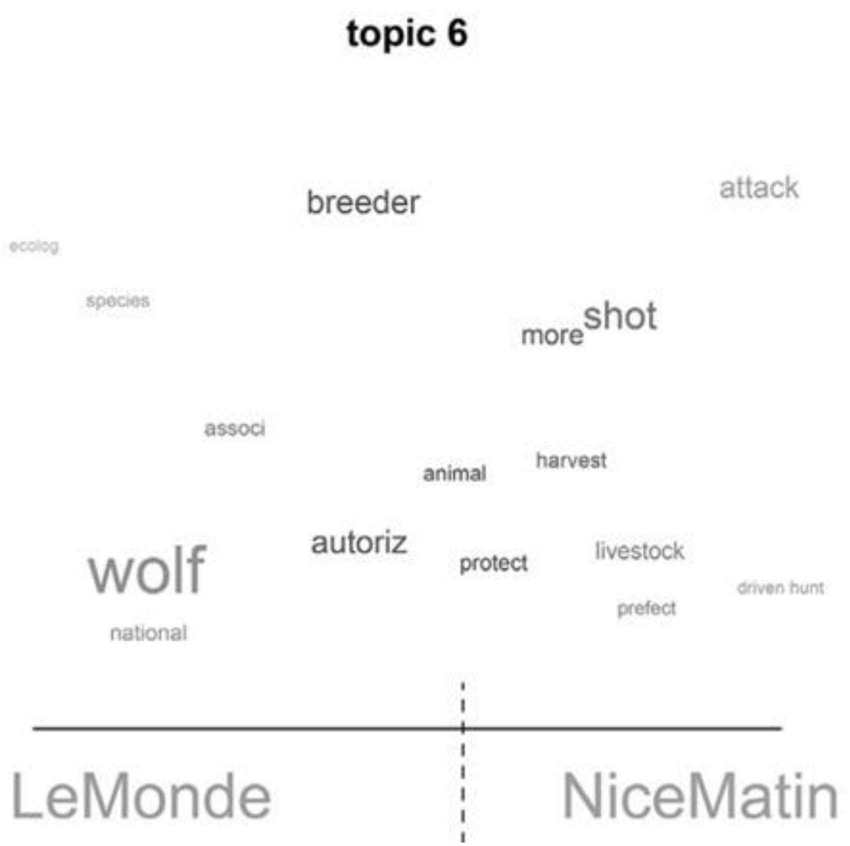
Roberts, Margaret E., et al. "The structural topic model and applied social science." *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*. Vol. 4. No. 1. 2013.

STM Example

21-year corpus on media coverage of grey wolf recovery in France

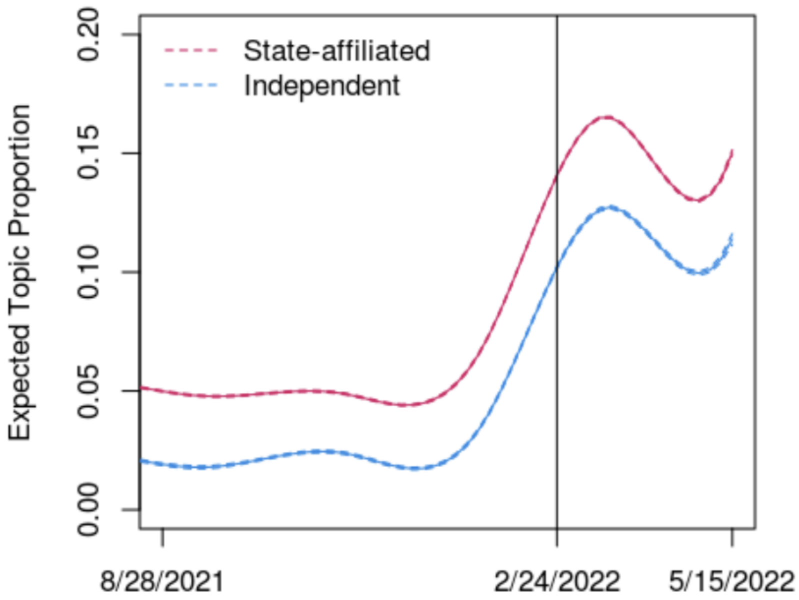
Nice-Matin = local newspaper
Le Monde = national newspaper

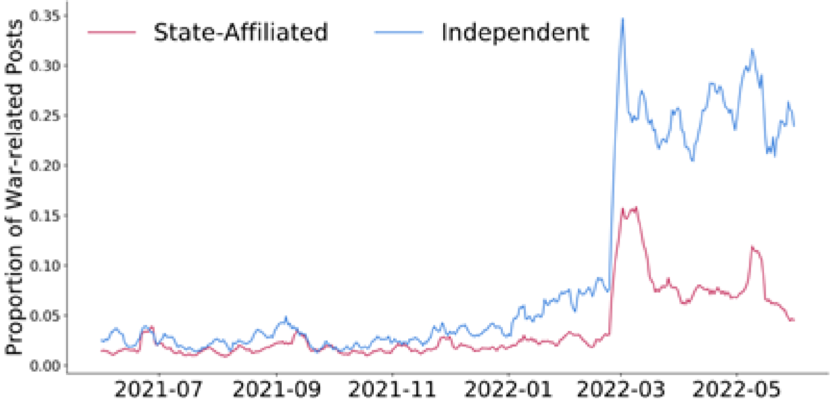
Topic 6: "Lethal Regulation"



https://www.structuraltopicmodel.com/
[Chandelier et al. 2018]

STM topic with the highest probability of Ukraine and military related





Topic might be capturing explicitly pro-Russia narrative

stm: R Package for Structural Topic Models

Margaret E. Roberts
UCSD

Brandon M. Stewart
Princeton

Dustin Tingley
Harvard

- Extremely popular go-to tool for computational social science (Cited 1000+ times)
- Flexible inclusion of covariates
- Tools for visualizing topic outputs
    - E.g. expected proportions, selecting example documents for each topic, representing topics with top words
- [Implemented in R package]

Today’s takeaways

- Key ideas behind variational inference
- Agenda setting and framing
- STM: example of adoption NLP method for social-oriented analysis

Next class:

- Word Embeddings

Logistics: HW1 has been released!

References

1. Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3.Jan (2003): 993-1022.
2. Roberts, Margaret E., et al. "The structural topic model and applied social science." *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*. Vol. 4. No. 1. 2013.

More links:

- https://www.youtube.com/watch?v=smfWKhDcaoA