



## Spectral Methods

Advanced Machine Learning for NLP

Jordan Boyd-Graber

ANCHOR TOPIC MODELS

Slides adapted from Thang Nguyen

## What are Spectral Methods

---

- Bayesian and deep models had explicit generative models
- Is it possible to find useful structure from matrix representations of data directly?
- Spectral methods: often very fast, but hard to engineer
- Like last week, a little out of place
- Today:
  - Anchor Words for Topic Models
  - Tensors

## What are Spectral Methods

---

- Bayesian and deep models had explicit generative models
- Is it possible to find useful structure from matrix representations of data directly?
- Spectral methods: often very fast, but hard to engineer
- Like last week, a little out of place
- Today:
  - Anchor Words for Topic Models
  - Tensors
  - Projects / Presentations
  - FCQ

## Anchor Method: Definition

---

### Baseball

Athlete  
Ball  
Base  
Catch  
Game  
Helmet  
Rival  
**Run**  
Shortstop  
Swing

### Soccer

Athlete  
Ball  
Dribble  
FIFA  
Game  
Offside  
Rival  
**Run**  
Tackle  
World Cup

### Election

Campaign  
Candidates  
Election  
Money  
Party  
Rival  
**Run**  
State  
A Swing  
Voters

- Words are often shared among many topics

## Anchor Method: Definition

---

### Baseball

Athlete  
Ball  
Base  
Catch  
Game  
Helmet  
Rival  
Run  
**Shortstop**  
Swing

### Soccer

Athlete  
Ball  
Dribble  
FIFA  
Game  
**Offside**  
Rival  
Run  
Tackle  
World Cup

### Election

Campaign  
Candidates  
Electorate  
Money  
Party  
Rival  
Run  
State  
A Swing  
**Voters**

- Words are often shared among many topics
- **Anchor words**: words that unique to a topic

## Anchor Method: Big Idea

---

- Normally, we want to find  $p(\text{word}|\text{topic})$

$$A_{i,k} = p(\text{word} = i|\text{topic} = \mathbf{k})$$

- What we'll do instead is find  $p(\text{topic}|\text{word})$  (topic coefficient)

$$C_{i,k} = p(\text{topic} = \mathbf{k}|\text{word} = i)$$

## Anchor Method: Big Idea

---

- Normally, we want to find  $p(\text{word}|\text{topic})$

$$A_{i,k} = p(\text{word} = i|\text{topic} = \mathbf{k})$$

- What we'll do instead is find  $p(\text{topic}|\text{word})$  (topic coefficient)

$$C_{i,k} = p(\text{topic} = \mathbf{k}|\text{word} = i)$$

- Easy: Bayes rule

## Anchor Method: Why go backward?

---

- Finding  $C_{i,k}$  is easy if you know the anchor words (assume we do!)
- $Q_{i,j} = p(\text{word}_1 = i, \text{word}_2 = j)$  is the cooccurrence probability
- Anchor method is so efficient because it uses conditional word distribution

$$\bar{Q}_{i,j} = p(\text{word}_2 = j | \text{word}_1 = i)$$



## Anchor Method: Why go backward?

---

- Finding  $C_{i,k}$  is easy if you know the anchor words (assume we do!)
- $Q_{i,j} = p(\text{word}_1 = i, \text{word}_2 = j)$  is the cooccurrence probability
- Anchor method is so efficient because it uses conditional word distribution

$$\bar{Q}_{i,j} = p(\text{word}_2 = j | \text{word}_1 = i)$$



The conditional probability distribution  $\bar{Q}_{\text{shortshop},*}$  looks a lot like the topic distribution!

## What about other words?

---

$$\bar{Q}_{\text{fly},*}$$

## What about other words?

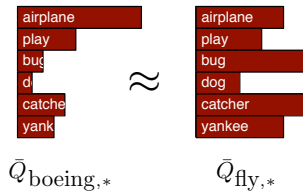
---

airplane  
play  
bug  
dog  
catcher  
yankee

$\bar{Q}$  fly,\*

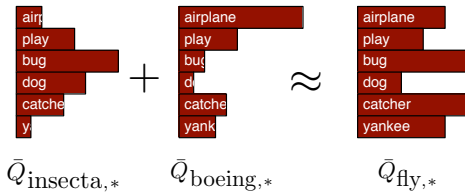
## What about other words?

---



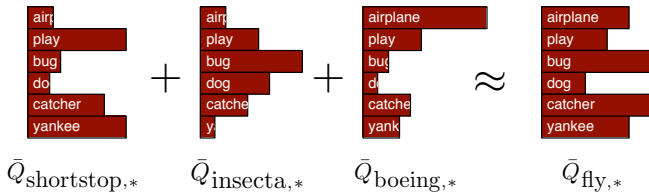
## What about other words?

---



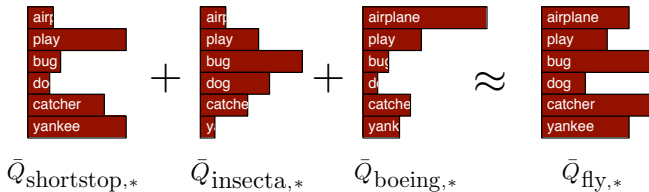
## What about other words?

---



$$\bar{Q}_{i,j} = \sum_k C_{i,k} \bar{Q}_{g_k,j}$$

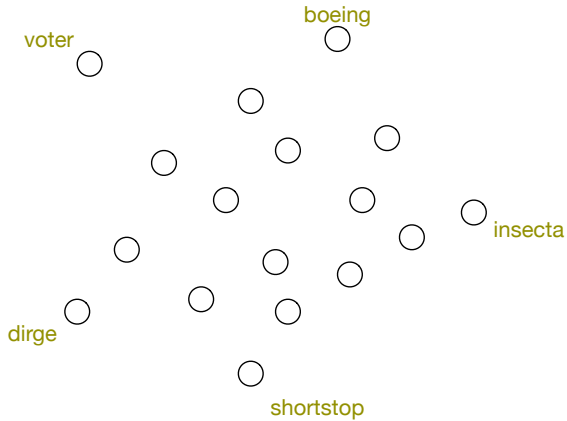
## What about other words?



$$\bar{Q}_{i,j} = \sum_k C_{i,k} \bar{Q}_{g_k,j}$$

## Topic Recovery

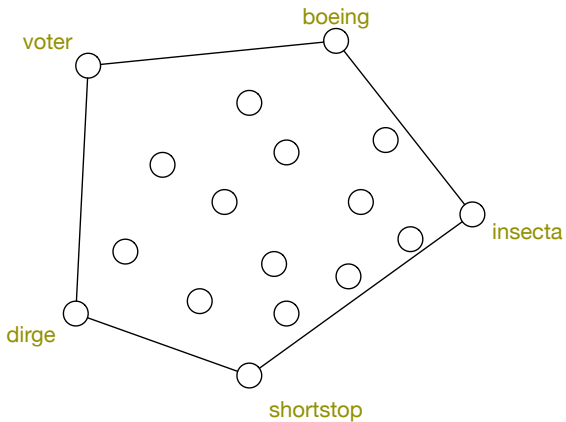
---





## Topic Recovery

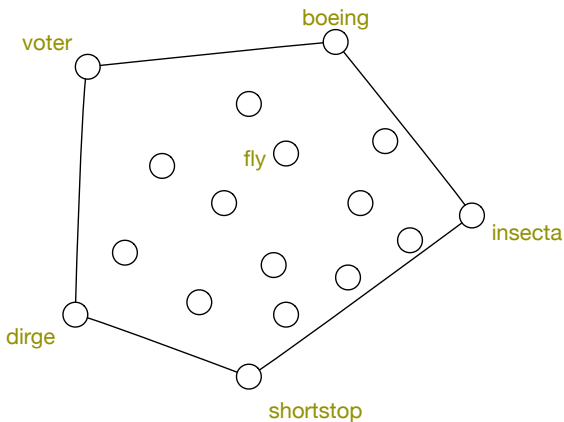
---



Let  $g_k$  be the anchor word for topic  $k$

## Topic Recovery

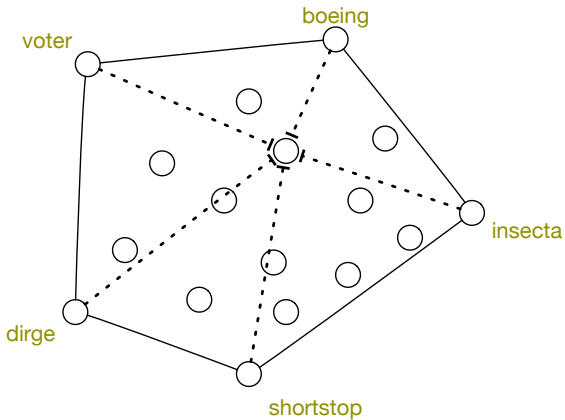
---



Let  $C_{i,k} = p(\text{topic}=k \mid \text{word}=i)$ ,  $C_{i,k} \geq 0$ ,  $\sum_k C_{i,k} = 1$

## Topic Recovery

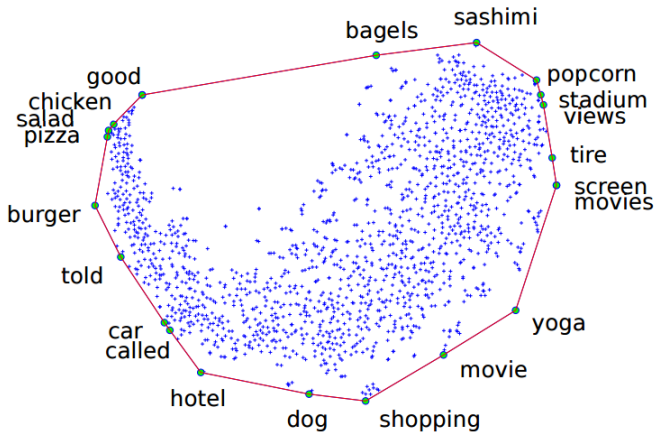
---



$$\bar{Q}_{i,j} = \sum_k c_{i,k} \bar{Q}_{g_k,j}$$

## Finding Anchor Words

---

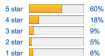


# A Significant Portion of Text is Labeled

## Customer Reviews

★★★★☆ 106,338

4.2 out of 5 stars



Share your thoughts with other customers

Write a customer review

See all verified purchase reviews

## Top Customer Reviews

★★★★★ This is a steal for \$50 as long as you aren't expecting a "Premium" experience.

By G.Hulse on October 2, 2015

Configuration: With Special Offers | Color: Black | Digital Storage Capacity: 8 | [Verified Purchase](#)

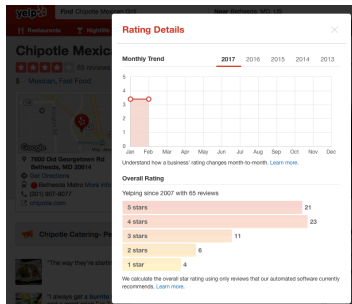
I pre-ordered this for my wife mostly to use as a Kindle E-reader as I figured the tablet would be slow and the display would be less than impressive. I was wrong. What a bargain this little beauty is! This model cost \$49.00 but it comes with ad's displayed on the lock screen when your tablet is dormant. Once your screen times out, they disappear. You can pay \$15.00 up front to get an ad free version so I assumed to unlock the tablet I'd have to spend 15 to 30 seconds looking at an ad for Amazon Prime, or a product from the daily specials section of Amazon.com I abstained from paying for Ad removal and was pleasantly surprised to find that the ads are only on the lock screen and that as soon as I unlock the tablet they disappear immediately.

Here are my pros and cons thus far.

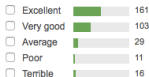
PRO:

Perfect size for Ebooks, and web surfing to alleviate strain on the eyes from my 5" phone display  
nice sturdy casing that gives it a nice heft but still weighs in as one of the lighter tablets on the market

Child Accounts- Amazon allows you to set up this tablet with age restricted access for kids making this a low cost piece of tech that is perfect for school kids and allows mom and dad to ration the amount of time JJ Johnny can play Clash of Clans and how much he can hit the of Visa card for.



## Traveler rating



## Traveler type



## Time of year



## Language



More

Showing 320: English reviews

Clear all

## Motivation

---

- Supervised topic models leverage latent document-level themes to capture nuanced sentiment, create sentiment-specific topics and improve sentiment prediction.
- Examples include Supervised LDA (Blei et al., 2007), Labelled LDA (Ramage et al., 2009), Med LDA (Zhu et al., 2009), etc.
- The downside is sluggish performance.

## Motivation

---

- Supervised topic models leverage latent document-level themes to capture nuanced sentiment, create sentiment-specific topics and improve sentiment prediction.
- Examples include Supervised LDA (Blei et al., 2007), Labelled LDA (Ramage et al., 2009), Med LDA (Zhu et al., 2009), etc.
- The downside is sluggish performance.
- **Create a supervised model based on Anchor Words?**

## Supervised Anchor Words: Idea

---

$$\bar{Q} \equiv \begin{bmatrix} p(w_1|w_1) \dots \\ \vdots \\ p(w_j|w_i) \end{bmatrix}$$

$$S \equiv \begin{bmatrix} p(w_1|w_1) \dots & p(y^{(l)}|w_1) \\ \vdots & \vdots \\ p(w_j|w_i) & p(y^{(l)}|w_i) \end{bmatrix}$$

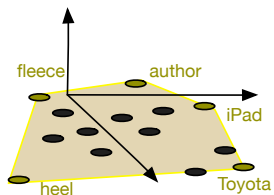
New column(s) encoding  
word-sentiment relationship



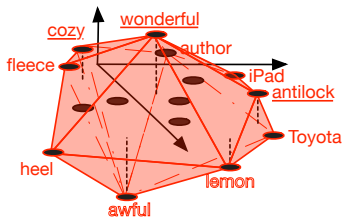
$$S_{i,\cdot} = \sum_{g_k \in \mathcal{G}} C_{i,k} S_{g_k,\cdot}$$



## Supervised Anchor Words: Intuition



- Adding sentiment related dimensions moves words UP or DOWN
- forming sentiment-specific points
- possibility of having different anchor words



## Evaluation of Supervised Anchor Words

---

- **Goal:** Evaluate the new topics generated by the proposed model in a prediction task. We focus on binary classification in sentiment analysis datasets.
- Sentiment datasets.

Corpus	Train	Test	Tokens	Vocab	+1
amazon	13,300	3,314	1,031,659	2,662	52.2%
tripadvisor	115,384	28,828	12,752,444	4,867	41.5%
yelp	13,955	3,482	1,142,555	2,585	27.7%

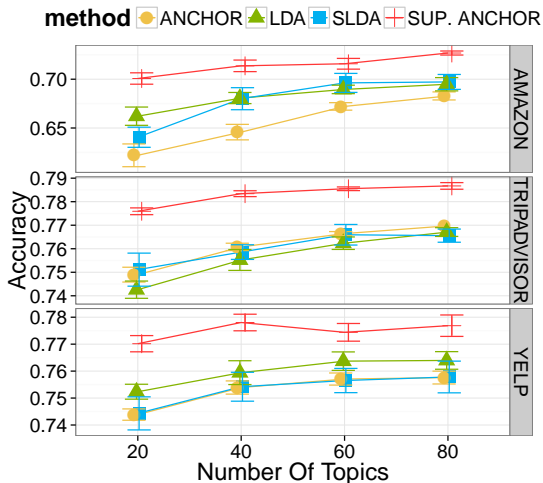
## Runtime Analysis

---

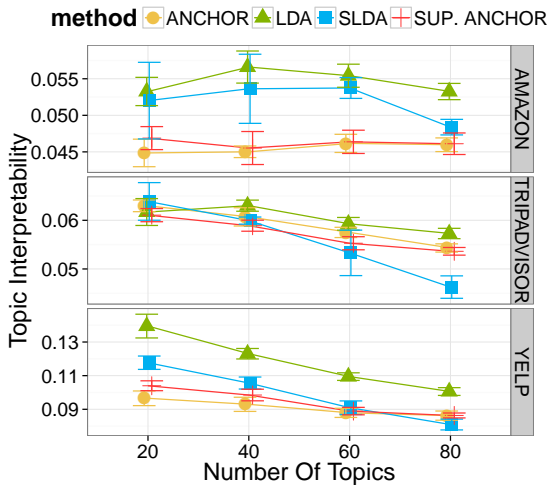


- Total time for training and prediction on amazon dataset.

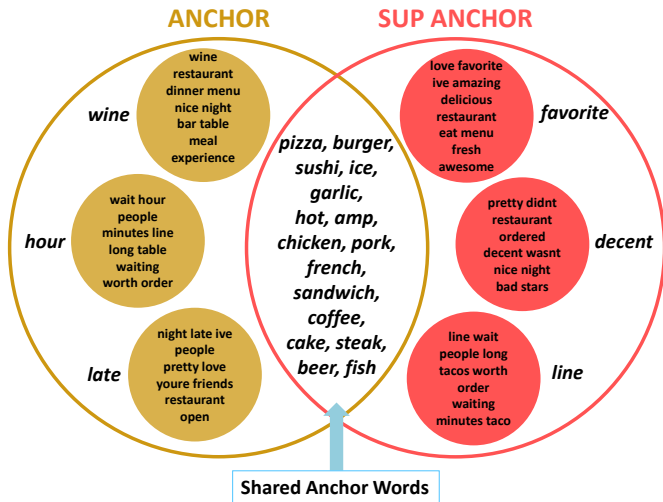
## Prediction Accuracy



## Topic Coherence



## Anchor Words and Their Topics



## Ongoing Work

---

- Near-instant updates
- Using multiple anchor words can improve coherence (and add interactivities)
- Downside: hard to create new models
- Hard to debug