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

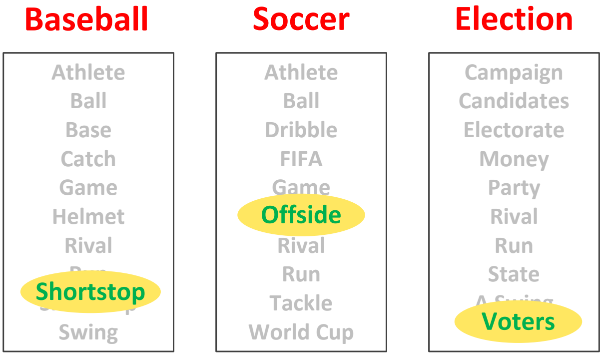
- Bayesian and deep models had explicit generative models
- Is it possible to find useful structure from matrix representations of data directly?
- Spectral methods: often very fast, but hard to engineer
- Like last week, a little out of place
- Today:
  - Anchor Words for Topic Models
  - Tensors
 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

Anchor Method: Definition

Baseball	Soccer	Election
Athlete	Athlete	Campaign
Ball	Ball	Candidates
Base	Dribble	Election
Catch	FIFA	Money
Game	Game	Party
Helmet	Offside	Rival
Rival	Rival	Run
Run	Run	State
Shortstop	Tackle	A Swing
Swing	World Cup	Voters

- Words are often shared among many topics

Anchor Method: Definition



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

Anchor Method: Big Idea

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

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

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

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

Anchor Method: Big Idea

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

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

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

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

- Easy: Bayes rule

Anchor Method: Why go backward?

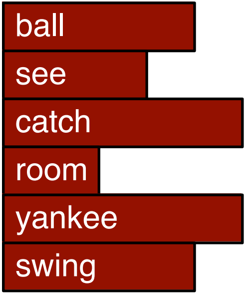
- Finding $C_{i,k}$ is easy if you know the anchor words (assume we do!)
- $Q_{i,j} = p(word_1 = i, word_2 = j)$ is the cooccurrence probability
- Anchor method is so efficient because it uses conditional word distribution

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

Anchor Method: Why go backward?

- Finding $C_{i,k}$ is easy if you know the anchor words (assume we do!)
- $Q_{i,j} = p(word_1 = i, word_2 = j)$ is the cooccurrence probability
- Anchor method is so efficient because it uses conditional word distribution

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

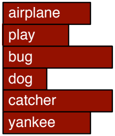


The conditional probability distribution $\bar{Q}_{\text{shortshop},*}$ looks a lot like the topic distribution!

What about other words?

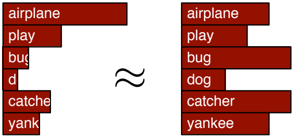
$$\bar{Q}_{\text{fly},*}$$

What about other words?



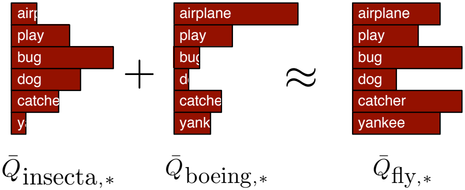
$\bar{Q}_{\text{fly},*}$

What about other words?

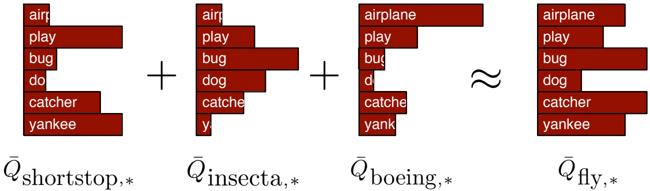


$\bar{Q}_{\text{boeing},*}$ $\bar{Q}_{\text{fly},*}$

What about other words?

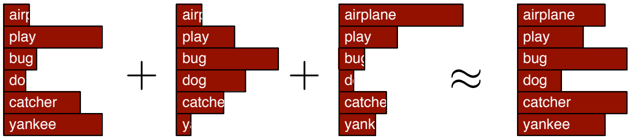


What about other words?



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

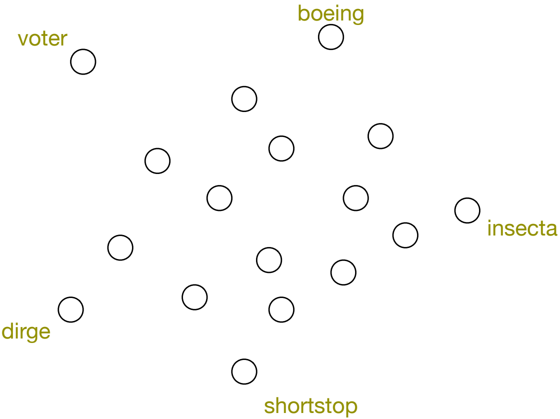
What about other words?



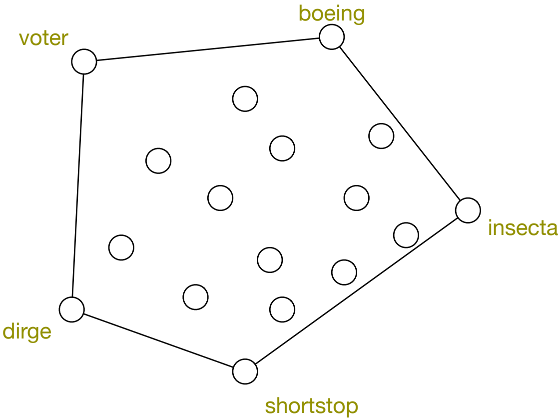
$$\bar{Q}_{\text{shortstop},*} + \bar{Q}_{\text{insecta},*} + \bar{Q}_{\text{boeing},*} \approx \bar{Q}_{\text{fly},*}$$

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

Topic Recovery

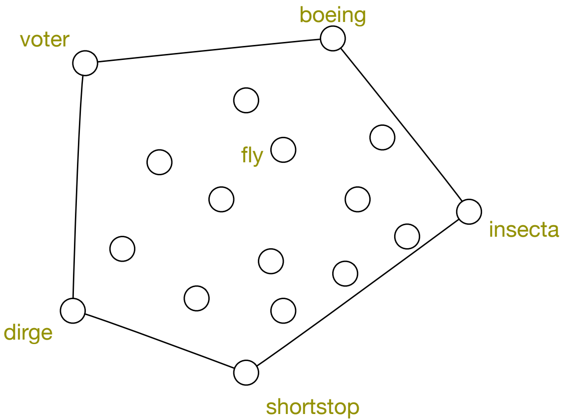


Topic Recovery



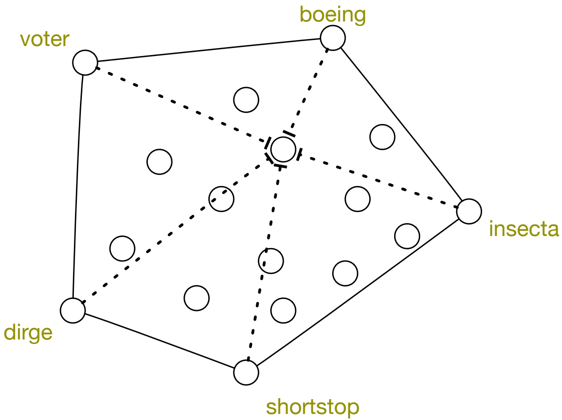
Let g_k be the anchor word for topic k

Topic Recovery



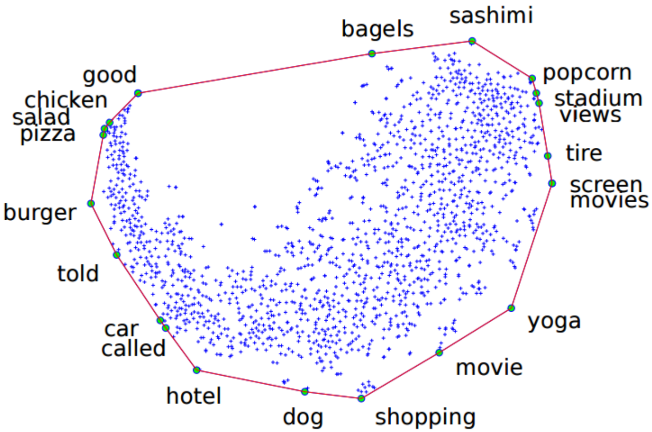
Let $C_{i,k}$ = p(topic=k | word=i), $C_{i,k} \geq 0$, $\sum_k C_{i,k} = 1$

Topic Recovery

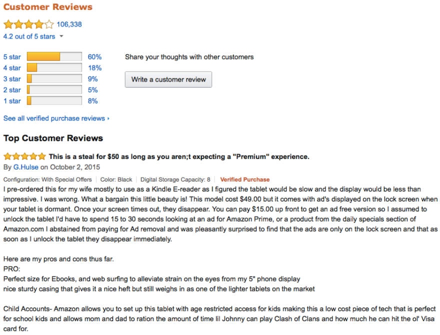


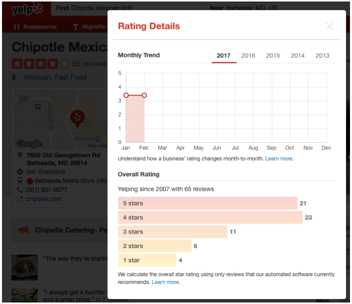
$$\bar{Q}_{i,j} = \sum_k C_{i,k} \bar{Q}_{g_k,j}$$

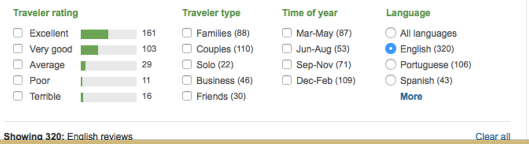
Finding Anchor Words



A Significant Portion of Text is Labeled







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

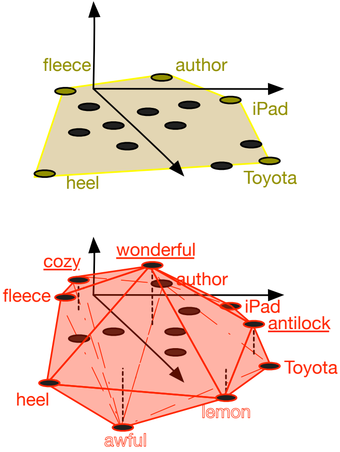
$$\bar{Q} \equiv \begin{bmatrix} p(w_1|w_1) \dots & & \\ & \vdots & \\ & & p(w_j|w_i) \end{bmatrix}$$

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

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

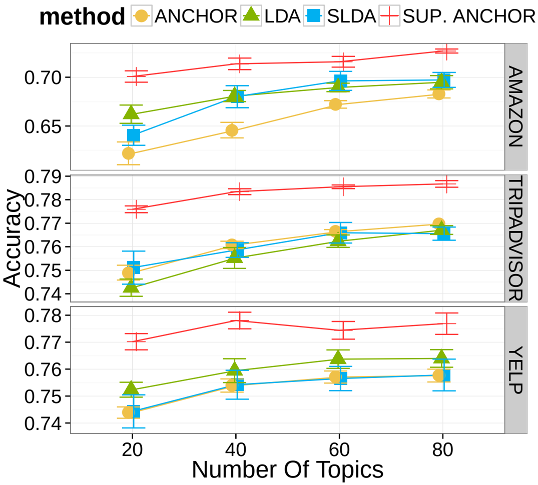
- **Goal**: Evaluate the new topics generated by the proposed model in a prediction task. We focus on binary classification in sentiment analysis datasets.
- Sentiment datasets.

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%

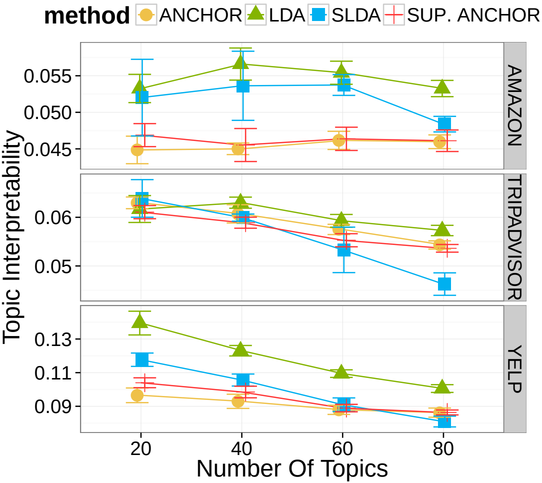
Runtime Analysis

- Total time for training and prediction on amazon dataset.

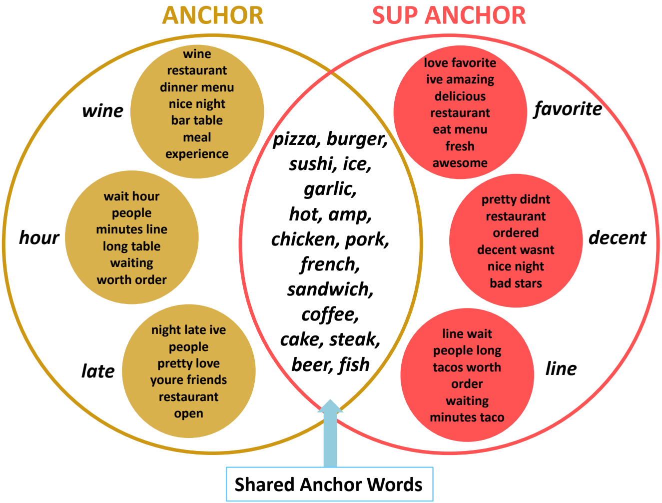
Prediction Accuracy



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