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

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

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

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

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

$$C_{i,k} = p(\text{topic} = k|\text{word} = i)$$
**Anchor Method: Big Idea**

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

\[
A_{i,k} = p(\text{word} = i | \text{topic} = k)
\]

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

\[
C_{i,k} = p(\text{topic} = k | \text{word} = i)
\]

- 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
  \[ \tilde{Q}_{i,j} = p(\text{word}_2 = j | \text{word}_1 = i) \]
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

$$\tilde{Q}_{i,j} = p(word_2 = j|word_1 = i)$$

The conditional probability distribution $\tilde{Q}_{\text{short shop},*}$ looks a lot like the topic distribution!
What about other words?

\[ Q_{fly,*} \]
What about other words?
What about other words?
What about other words?

\[ \bar{Q} \text{insecta,*} + \bar{Q} \text{boeing,*} \approx \bar{Q} \text{fly,*} \]
What about other words?

\[ \bar{Q}_{\text{shortstop},*} + Q_{\text{insecta},*} + Q_{\text{boeing},*} \approx Q_{\text{fly},*} \]

\[ \bar{Q}_{i,j} = \sum_{k} C_{i,k} \bar{Q}_{g_k,j} \]
What about other words?

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

\[
\bar{Q}_{i,j} = \sum_k C_{i,k} \bar{Q}_{g_k,j}
\]
Topic Recovery

- voter
- boeing
- insecta
- dirge
- shortstop
Let $g_k$ be the anchor word for topic $k$
Let $C_{i,k} = \Pr(\text{topic}=k | \text{word}=i)$, $C_{i,k} \geq 0$, $\sum_k C_{i,k} = 1$
Topic Recovery

\[ \tilde{Q}_{i,j} = \sum_k C_{i,k} \tilde{Q}_{g_k,j} \]
Finding Anchor Words
A Significant Portion of Text is Labeled

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

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 $49.99 but it comes with a 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 feel 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 Ill Johnny can play Clash of Clans and how much he can hit the of Visa card for.
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

\[ \tilde{Q} \equiv \begin{bmatrix} p(w_1|w_1) & \ldots \\ \vdots \\ p(w_j|w_i) \end{bmatrix} \]

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

New column(s) encoding word-sentiment relationship

\[ S_{i,} = \sum_{g_k \in \mathcal{G}} C_{i,k} S_{g_k,} \]
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.**

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Train</th>
<th>Test</th>
<th>Tokens</th>
<th>Vocab</th>
<th>+1</th>
</tr>
</thead>
<tbody>
<tr>
<td>amazon</td>
<td>13,300</td>
<td>3,314</td>
<td>1,031,659</td>
<td>2,662</td>
<td>52.2%</td>
</tr>
<tr>
<td>tripadvisor</td>
<td>115,384</td>
<td>28,828</td>
<td>12,752,444</td>
<td>4,867</td>
<td>41.5%</td>
</tr>
<tr>
<td>yelp</td>
<td>13,955</td>
<td>3,482</td>
<td>1,142,555</td>
<td>2,585</td>
<td>27.7%</td>
</tr>
</tbody>
</table>
Total time for training and prediction on Amazon dataset.
Prediction Accuracy

![Graph showing prediction accuracy for different methods and datasets. The x-axis represents the number of topics, ranging from 20 to 80. The y-axis represents accuracy, ranging from 0.74 to 0.79. The methods include Anchor, LDA, SLDA, and Supervised Anchor. The datasets are Amazon, TripAdvisor, and Yelp. The graph shows the accuracy for each method across different numbers of topics.]
Topic Coherence

![Topic Interpretability Chart]

**Method**
- ANCHOR
- LDA
- SLDA
- SUP. ANCHOR

**Topics Coherence**

- **Number of Topics**
- **Amazon**
- **Trip Advisor**
- **Yelp**
Anchor Words and Their Topics

**ANCHOR**
- **wine**: wine, restaurant, dinner menu, nice night, bar table, meal experience
- **hour**: wait hour, people, minutes line, long table, waiting, worth order
- **late**: night, late, ive, people, pretty love, your friends, restaurant, open

**SUP ANCHOR**
- **favorite**: love, favorite, ive, amazing, delicious, restaurant, eat menu, fresh, awesome
- **decent**: pretty, didnt, restaurant, ordered, decent, wasnt, nice night, bad stars
- **line**: line, wait, people, long, tacos, worth, order, waiting, minutes, taco

**Shared Anchor Words**
- pizza, burger, sushi, ice, garlic, hot, amp, chicken, pork, french, sandwich, coffee, cake, steak, beer, fish
Ongoing Work

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