Content Modeling for Social Media Text

Christina Sauper

Joint work with Aria Haghighi and Regina Barzilay
This place is pure mediocrity.

My boyfriend and I went there on a Saturday night at 9pm. We entered the (basically half-empty) restaurant and asked for a table for two. The host asked if we had a reservation, we said we were guests of the hotel, he said he would "check". After a while he ushered us into the place. There were literally 8 empty tables next to us the entire time we were there. I'm glad they could fit us into their busy schedule.

This set the tone for the entire evening. They were understaffed and although our waitress was nice, she wasn't very attentive - it took forever to get drink orders, bread, etc.

We started with the oysters and the tuna tartare. The oysters were really good, but if you're a decent restaurant in Boston and you have crappy oysters, you should just close up shop now. The tuna was okay, but I question it being served with potato chips.

My boyfriend had the steak and fries and I had the pork with cheddar grits and mushrooms. The grits and mushrooms were awesome, but the pork was flavorless and not very tender. BF said the steak and fries was on the extremely average side. Not only that, when it was delivered to our table, the waitress looked askance at it and said "Do you want that cooked more?" It wasn't rare, it was raw. She took it back to the kitchen and we waited a few minutes until it returned, re-cooked.

The biggest disappointment was dessert. I ordered the espresso creme brulee (I am a big dessert person). The waitress brought out the brulee, without the biscotti it comes with. I asked about it, and she looked puzzled and took it back. She returned saying they had run out of biscotti and had given me a chocolate-chip cookie cut into thirds instead.

Overall, I'd never go back and I would actively encourage you to avoid this place as well. When I spend $100 on dinner, I expect to enjoy myself and eat well. I'd rather spend $30 on tasty Indian food or $50 on sushi than eat mediocre nouvelle American for that much. A big disappointment.
Brasserie fare with a busy backdrop
By Alison Arnett, Globe Staff

First you have to get past the menu. It reads like a kid's game - maybe something to do with Barbie dolls or Miss Kitty. Crispy, soupy, leafy, crabby - crabby? The designations are cute but distracting and make it difficult to take the food seriously.

Then there's the look of Bambara, which is sort of generic hotel design (why do hotel restaurants always have such busy patterns in the carpets and furnishings?), ameliorated by some interesting details. In a comfortable room with deeply cushioned booths and high bar tables, there is so much going on decor-wise, from several styles of light fixtures, metal railings, pretty salt and pepper shakers continually tipped over by diners and the wait staff, and sheets of paper laid over tablecloths. (I know this is supposed to be brasserie-style, but nothing else looks like a brasserie, and the paper gets caught on diners' sleeves and is tacky.) All this takes away the pleasing sense of outdoors from the tall windows.

Then there are service disappointments - soup served without a spoon, tea and coffee orders mixed up, and more important, waits between courses that seem frozen in time. On a fairly busy Saturday evening, the courses lurched along like a tortoise, slow and steady. On a weeknight, the pauses for someone to inquire about wine, take the order, and bring appetizers were noticeable. But the wait for the main courses was painful. This was particularly egregious because the chef - whom I recognized from a press packet sent to me - was chatting with a group near the bar.

Too bad, because many of the dishes by chef Tom Berry, an alumnus of Ming Tsai's Blue Ginger, are quite delicious. Although Bambara and the Hotel Marlowe, in which it is located, are part of a group, each restaurant has its own concept and each chef designs his or her own menu, according to a manager. Berry, who is from New Hampshire, reveals his affinity for seafood in many dishes, and the menu is especially awash in crab. A fat crab cake falls apart when a fork breaks through the crisped exterior, revealing that it is almost all flaky crab meat. Its sweetness is nicely balanced with chunks of golden beets and an arugula salad. Soft-shell crab is dusted with ancho chili powder to give an intriguing flavor without much heat to the crispy fried crustacean.
Social media text challenges

• Less apparent structure

Brasserie fare with a busy backdrop
By Alison ArneS, Globe Staff

First you have to get past the menu. It reads like a kid’s game - maybe something to do with Barbie dolls or Miss Kitty. Crispy, soupy, leafy, crabby - crabby? The designations are cute but distracting and make it difficult to take the food seriously.

Then there’s the look of Bambara, which is sort of generic hotel design (why do hotel restaurants always have such busy patterns in the carpets and furnishings?), ameliorated by some interesting details. In a comfortable room with deeply cushioned booths and high bar tables, there is so much going on decor-wise, from several styles of light fixtures, metal railings, pretty salt and pepper shakers continually tipped over by diners and the wait staff, and sheets of paper laid over tablecloths. (I know this is supposed to be brasserie-style, but nothing else looks like a brasserie, and the paper gets caught on diners’ sleeves and is tacky.) All this takes away the pleasing sense of outdoors from the tall windows.

Then there are service disappointments - soup served without a spoon, tea and coffee orders mixed up, and more important, waits between courses that seem frozen in time. On a fairly busy Saturday evening, the courses lurched along like a tortoise, slow and steady. On a weekend, the pauses for someone to inquire about wine, take the order, and bring appetizers were noticeable. But the wait for the main courses was painful. This was particularly egregious because the chef - whom I recognized from a press packet sent to me - was chatting with a group near the bar.

Too bad, because many of the dishes by chef Tom Berry, an alumnus of Ming Tsai’s Blue Ginger, are quite delicious. Although Bambara and the Hotel Marlowe, in which it is located, are part of a group, each restaurant has its own concept and each chef designs his or her own menu, according to a manager. Berry, who is from New Hampshire, reveals his affinity for seafood in many dishes, and the menu is especially awash in crab. A fat crab cake falls apart when a fork breaks through the crisped exterior, revealing that it is almost all flaky crab meat. Its sweetness is nicely balanced with chunks of golden beets and an arugula salad. Soft-shell crab is dusted with ancho chili powder to give an intriguing flavor without much heat to the crispy fried crustacean.

This place is pure mediocrity.

My boyfriend and I went there on a Saturday night at 9pm. We entered the (basically half-empty) restaurant and asked for a table for two. The host asked if we had a reservation, we said we were guests of the hotel, he said he would “check”. After a while he ushered us into the place. There were literally 8 empty tables next to us the entire time we were there. I’m glad they could fit us into their busy schedule.

This set the tone for the entire evening. They were understaffed and although our waitress was nice, she wasn’t very attentive - it took forever to get drink orders, bread, etc.

We started with the oysters and the tuna tartare. The oysters were really good, but if you’re a decent restaurant in Boston and you have crappy oysters, you should just close up shop now. The tuna was okay, but I question it being served with potato chips.

My boyfriend had the steak and fries and I had the pork with cheddar grits and mushrooms. The grits and mushrooms were awesome, but the pork was flavorless and not very tender. BF said the steak and fries was on the extremely average side. Not only that, when it was delivered to our table, the waitress looked askeance at it and said “Do you want that cooked more?” It wasn’t rare, it was raw. She took it back to the kitchen and we waited a few minutes until it returned, re-cooked.

The biggest disappointment was dessert. I ordered the espresso creme brulee (I am a big dessert person). The waitress brought out the brulee, without the biscotti it comes with. I asked about it, and she looked puzzled and took it back. She returned saying they had run out of biscotti and had given me a chocolate-chip cookie cut into thirds instead.

Overall, I’d never go back and I would actively encourage you to avoid this place as well. When I spend $100 on dinner, I expect to enjoy myself and eat well. I’d rather spend $30 on tasty Indian food or $50 on sushi than eat mediocre nouvelle American for that much. A big disappointment.
Social media text challenges

• Less apparent structure
• Novel words

this asian inspired burger can say "konichi-hot-damn-that-is-an-awesome-tastic-burger-get-in-my-mouth-faster-little-heavenly-morsel"

All of the pasta dishes we ordered were delectable, saporofic and fantastically fabulicious.
Social media text challenges

• Less apparent structure
• Novel words
• Poor spelling and grammar

soo our food came out and i had order da eggplant .... and it was nicey n oily.. yuck... ugh... our server came around mayb once? or twice to refil our drinks and she put the wrong drink in my friend drink.. hmmmmmmmmmm
Social media text challenges

- Less apparent structure
- Novel words
- Poor spelling and grammar
- Dependence on outside context

Spencer's on game day is a completely different experience then Spencer's on a non-game day.

So it's not as close to home as "that other place" but that other place isn't as good as it used to be.

I ordered a small iced almond latte, which ended up being the same size as a grande at the-evil-chain-which-shall-not-be-named.
Social media text challenges

• Less apparent structure
• Novel words
• Poor spelling and grammar
• Dependence on outside context

Content modeling can help!
Options for content modeling

Many possible choices of content models

• Summarization or extraction:

<table>
<thead>
<tr>
<th>Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK food. Nice atmosphere. I would go mostly for the ambiance.</td>
</tr>
<tr>
<td>We had an early dinner after returning from Logan airport. The staff was friendly and seated us early.</td>
</tr>
<tr>
<td>Overall, food was unimpressive. The best thing about this restaurant is the ambiance and hotel decor, as it is situated within Hotel Marlowe, which has a chic lobby. It beats dining at the Cheesecake Factory.</td>
</tr>
</tbody>
</table>

• Sentiment Analysis:

<table>
<thead>
<tr>
<th>Opinionated sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>OK food. Nice atmosphere. I would go mostly for the ambiance.</td>
</tr>
<tr>
<td>We had an early dinner after returning from Logan airport. The staff was friendly and seated us early.</td>
</tr>
<tr>
<td>Overall, food was unimpressive. The best thing about this restaurant is the ambiance and hotel decor, as it is situated within Hotel Marlowe, which has a chic lobby. It beats dining at the Cheesecake Factory.</td>
</tr>
</tbody>
</table>
Classical discourse structure

• Topics
  – Latent Semantic Analysis (Deerwester & al. 1990)
  – Latent Dirichlet Allocation (Blei & al. 2003)

• Schemas or Templates
  – Event chains (Chambers & Jurafsky 2008)

• Low-level relations
  – Rhetorical Structure Theory (Mann & Thomson 1988)
  – Rhetorical structure relations (Marcu & Echihabi 2002)

These representations are task-specific!
Contributions

• Learning relevant document structure *automatically* with task parameters improves task performance

• Modeling the structure of relations in text allows informative aggregation across multiple documents

• Idiosyncrasies of social media text can be overcome with a rich model of content structure
Contributions

• Combining these ideas yields real-world benefit
Task-specific content modeling for single documents
Finding structure in a single document

• Goal: Given a review, identify phrases corresponding to several pre-defined aspects

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food</td>
<td>The pizza tasted like cardboard! Calzones were amazing</td>
</tr>
<tr>
<td>Ambiance</td>
<td>Stylish décor! They have great jazz on Saturday nights</td>
</tr>
<tr>
<td>Service</td>
<td>Fast and friendly service I waited 45 minutes for the waiter</td>
</tr>
</tbody>
</table>

• Context can help!

...from local artists. The food was great! I ordered the veal parm and loved it. The service on the other hand...


Finding structure in a single document

• Sequence Labeling Task

I ordered lunch from them the other day and I was pleasantly surprised. Our waiter dazzled me with his blue eyes and genuine smile, and all the waiters were extremely professional and efficient.

• Content Model

I ordered lunch from them the other day and I was pleasantly surprised. Our waiter dazzled me with his blue eyes and genuine smile, and all the waiters were extremely professional and efficient.
The Big Disconnect

Discourse Modeling

- Topic models
- Rhetorical Structure Theory

Analysis Applications

- Information extraction
- Sentiment analysis
Approach Overview

• Jointly learn content and task parameters
  – Content model is shaped by task

• Incorporate unlabeled data in a principled way
  – More unlabeled data, better performance
Phrase extraction

Goal: Extract phrases from review text for pre-specified aspects

Input: User-generated review text: labeled training data, unlabeled text

Output: Labeled phrases for each aspect, formed by word labeling

<table>
<thead>
<tr>
<th>FOOD</th>
<th>SERVICE</th>
<th>ATMOSPHERE</th>
<th>PRICE</th>
<th>OVERALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>I came here with my husband for the tasting menu, and we were not disappointed. We got to sit at the chef’s table, which overlooked the kitchen. The service was polite and knowledgeable, the atmosphere was elegant and energetic and the food was wonderfully creative and delicious.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Model Summary

Text

Task labels

have great food and

Task parameters
Model Summary

Content structure

Text

Task labels

have great food and

T

S

Y

θ

ϕ

Content parameters

Task parameters

have great food and
Multi-aspect phrase extraction

- **Content model**
  - sentence-level HMM

- **Task model**
  - token-level linear chain CRF

They have great food and awesome service
Augmenting a CRF with Topics
Factorization

\[ P(T, s, y) = P_\theta(T, s) P_\phi(y|T, s) \]

\[ = \prod_{i=1}^{n} P_\theta(T_{i+1}|T_i) \left( P_\theta(s_i|T_i) P_\phi(y^*_i|s_i, T_i) \right) \]

Product over sentences

\[ \text{Topic transitions} \quad \text{Content model emission} \quad \text{CRF} \]

| \( \theta \) | Topic params |
| \( T \) | Sentence topics |
| \( S \) | Sentences |
| \( \phi \) | Task params |
| \( y \) | Task labels |
Factorization

\[ P(T, s, y) = P_\theta(T, s)P_\phi(y|T, s) \]

\[ = \prod_{i=1}^{n} P_\theta(T_{i+1}|T_i) \left( P_\theta(s_i|T_i)P_\phi(y_i^*|s_i, T_i) \right) \]
Joint learning

\[ \mathcal{L}(\theta, \phi) = \sum_{(s, y^*)} \log P(s, y^*) \]

\[ = \sum_{(s, y^*)} \log \left( \sum_{T} P(T, s, y^*) \right) \]
Joint learning

E-step:

\[ P(T | s, y^*) \propto P(T, s, y^*) \]

\[ = \prod_{i=1}^{n} P_\theta(T_{i+1} | T_i) \left( P_\theta(s_i | T_i) P_\phi(y_i^* | s_i, T_i) \right) \]

- Can be computed using Forward-Backward algorithm

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>Topic params</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T )</td>
<td>Sentence topics</td>
</tr>
<tr>
<td>( S )</td>
<td>Sentences</td>
</tr>
<tr>
<td>( \phi )</td>
<td>Task params</td>
</tr>
<tr>
<td>( y )</td>
<td>Task labels</td>
</tr>
</tbody>
</table>
Joint learning

M-step:

\[ \theta \text{ : Standard normalization of counts from E-step} \]

\[ \phi \text{ : weighted conditional likelihood objective} \]

\[
\max_{\phi} \sum_{i} \mathbb{E}_{T_i} \log P(y_i^*|s_i, T_i)
\]

| \( \theta \) | Topic params |
| \( \phi \) | Task params |
| **T** | Sentence topics |
| **S** | Sentences |
| **Y** | Task labels |
Supervised objective

\[ \mathcal{L}(\theta, \phi) = \sum_{(s, y^*)} \log P_{\theta, \phi}(s, y^*) \]

Labeled data for task parameters

| \theta | Topic params |
| \phi   | Task params  |
| \text{T} | Sentence topics |
| \text{S}  | Sentences    |
| \text{y}  | Task labels  |
Semi-supervised objective

\[
\mathcal{L}(\theta, \phi) = \sum_{(s,y^*)} \log P_{\theta,\phi}(s, y^*) + \sum_{s} \log P_{\theta}(s)
\]

- Labeled data for task parameters
- Unlabeled data for content parameters

\[\theta\] Topic params \[T\] Sentence topics \[S\] Sentences

\[\phi\] Task params \[Y\] Task labels
Data sets

- Standard review domains
- Annotated with Mechanical Turk

Amazon TV reviews
- Train 35
- Test 24
- Unlabeled 12,684

Yelp restaurant reviews
- Train 48
- Test 48
- Unlabeled 33,015

- Doctor-created summaries of patient visits
- Annotated by doctors

Medical summary text
- Train 47
- Test 47
- Unlabeled 206
Systems

• **NoCM**  Just the CRF, no content model

• **IndepCM**  Estimate content parameters and task params independently

• **JointCM**  Estimate content and task parameters jointly using EM
Results

Token F-measure evaluation

<table>
<thead>
<tr>
<th>NoCM</th>
<th>IndepCM</th>
<th>JointCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>Yelp</td>
<td>Medical</td>
</tr>
<tr>
<td>18.9</td>
<td>28.2</td>
<td>43.4</td>
</tr>
<tr>
<td>24.5</td>
<td>37.9</td>
<td>55.5</td>
</tr>
<tr>
<td>28.8</td>
<td>39.2</td>
<td>56.6</td>
</tr>
</tbody>
</table>
Impact of unlabeled data

Setup: Amazon data set
  – Complete annotated data set
  – Varying amounts of unlabeled data
Compensating for annotation sparsity

Setup: Amazon data set
- Half of the annotated data set
- Varying amounts of unlabeled data

<table>
<thead>
<tr>
<th>Unlabeled Data</th>
<th>NoCM</th>
<th>15.1</th>
<th>20.8</th>
<th>22.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>15.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,575</td>
<td></td>
<td></td>
<td>20.8</td>
<td></td>
</tr>
<tr>
<td>3,150</td>
<td></td>
<td></td>
<td></td>
<td>22.6</td>
</tr>
</tbody>
</table>
Multi-aspect sentiment ranking

• **Goal:** Predict rating (1-10) for several aspects

  **Input:** Review text, labeled training data
  **Output:** Numeric score for each aspect

  **Example Input:**
  
  I had a wonderful meal here last year and I’m counting down the days until I go back.
  The service was polite and knowledgeable, the atmosphere was elegant and energetic, and the food was wonderfully creative and delicious.
  We got to sit at the chef’s table, which consisted of 4 bar stools along the kitchen. It overlooked the people who were in charge of plating appetizers. Such a stressful job, I don’t think I could ever work in a kitchen.
  Although the tasting menu was ubiquitous (and delicious looking), we opted to create a little pig feast of our own. We had the crispy pig tails, trio of pates, crispy pig ears, and the pig’s...

  **Example Output:**
<table>
<thead>
<tr>
<th>Aspect</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Movie</td>
<td>7</td>
</tr>
<tr>
<td>Audio Quality</td>
<td>9</td>
</tr>
<tr>
<td>Video Quality</td>
<td>8</td>
</tr>
<tr>
<td>Extras</td>
<td>6</td>
</tr>
</tbody>
</table>

• **Approach:**
  
  **Task model:** Independent linear regression
  **Content model:** Paragraph-level HMM

  E- and M-steps differ slightly; same objective
Multi-aspect sentiment ranking

- IGN.com DVD Reviews
  - 600 train, 65 test

L_2 Error (lower is better)

<table>
<thead>
<tr>
<th>NoCM</th>
<th>IndepCM</th>
<th>JointCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.15</td>
<td>2.80</td>
<td>2.65</td>
</tr>
</tbody>
</table>
Multi-aspect sentiment ranking

- IGN.com DVD Reviews
  - 600 train, 65 test

![L2 Error Chart]

<table>
<thead>
<tr>
<th>Method</th>
<th>L2 Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoCM</td>
<td>3.15</td>
</tr>
<tr>
<td>IndepCM</td>
<td>2.80</td>
</tr>
<tr>
<td>JointCM</td>
<td>2.65</td>
</tr>
<tr>
<td>GoldCM</td>
<td>2.48</td>
</tr>
</tbody>
</table>
Informative aggregation across multiple documents
Review Aggregation

- Opinions vary widely
  - Need aggregate statistics
- Histograms show sentiment distribution, but it’s not enough
Aspect-based Analysis

Pre-defined domain-specific product aspects

→ Coarse level analysis
The fried oysters were very good
The catfish tasted dry and bland and boring
The star of the plate was the grits

The gnocchi with mushrooms was outstanding
The catfish approaches perfection
The shrimp and grits are nothing less than spectacular
The fried oysters were very good

The catfish tasted dry and bland and boring

The star of the plate was the grits

The gnocchi with mushrooms was outstanding

The catfish approaches perfection

The shrimp and grits are nothing less than spectacular
Informative Aggregation

Useful information:

– What’s the best dish at this restaurant?

– What do people dislike about this restaurant?

– Which dishes do people disagree about?
We had a great time last night at this restaurant. The sushi was so incredibly fresh. We had a bad experience at the bar, though. My chocolate martini was absolutely terrible. We will be back, but we’ll skip the drinks.

Wow, I can’t believe how much this place has changed! They used to be mediocre, but now they never fail to amaze. We started off at the bar with awesome sake bombs. When we got to our table, the sushi was fantastic.

I have such mixed things to say about this restaurant. On one hand, their sushi is unquestionably the best in the city. On the other, the atmosphere isn’t that great. Plus, their drinks are completely watered down.

Japanese Restaurant

Sushi
100% positive

Drinks
33% positive

Relevant aspects

User sentiment

Informative Aggregation

Aggregation of *product-specific* aspects
Corpus-driven Aspect Definition

Define aspects *dynamically* based on reviews

<table>
<thead>
<tr>
<th>Japanese Restaurant</th>
<th>Bakery</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>We had a great time last night at this restaurant. The sushi was so incredibly fresh.</strong></td>
<td><strong>We had a great time last night at this restaurant. The sushi was so incredibly fresh.</strong></td>
</tr>
<tr>
<td><strong>We had a bad experience at the bar, though. My chocolate martini was absolutely terrible.</strong></td>
<td><strong>We had a bad experience at the bar, though. My chocolate martini was absolutely terrible.</strong></td>
</tr>
<tr>
<td><strong>Wow, I can’t believe how much this place has changed! They used to be mediocre, but now they never fail to amaze. We started off at the bar with awesome sake bombs. When we got to our table, the sushi was fantastic.</strong></td>
<td><strong>Wow, I can’t believe how much this place has changed! They used to be mediocre, but now they never fail to amaze. We started off at the bar with awesome sake bombs. When we got to our table, the sushi was fantastic.</strong></td>
</tr>
<tr>
<td><strong>I have such mixed things to say about this restaurant. On one hand, their sushi is unquestionably the best in the city. On the other, the atmosphere isn’t that great. Plus, their drinks are completely watered down.</strong></td>
<td><strong>I have such mixed things to say about this restaurant. On one hand, their sushi is unquestionably the best in the city. On the other, the atmosphere isn’t that great. Plus, their drinks are completely watered down.</strong></td>
</tr>
</tbody>
</table>

- Sushi
- Sake
- Dessert
- Cookies
- Cakes
- Pies

→ Aspects specific to each product
Corpus-driven Aspect Definition

Allows comparison across multiple reviews

---

Bakery

I buy all of my baked goods from this bakery. Their bread is so delicious! It's also good for all kinds of baked goods. They also have some truly beautiful cakes on display. Even their cookies are great!

I picked up a birthday cake for my son here yesterday. It was the most amazing cake I've ever seen! The decorations were outstanding, and all the kids loved the chocolate icing. I'll definitely come back!

This place is nice for some baked goods, but some things are really nasty. The loaf of bread I bought was stale! They were happy to take it back and give me another, but I'll be watching next time.

---

...truly beautiful cakes on display.  ...most amazing cake I’ve ever seen!

---

– Consensus (both positive and negative)

What’s the best/worst aspect of this product?
## Corpus-driven Aspect Definition

Allows comparison across multiple reviews

<table>
<thead>
<tr>
<th>Bakery</th>
<th>Bakery</th>
<th>Bakery</th>
</tr>
</thead>
<tbody>
<tr>
<td>I buy all of my baked goods from this bakery. Their bread is so delicious! It's also good for all kinds of baked goods. They also have some truly beautiful cakes on display. Even their cookies are great!</td>
<td>I picked up a birthday cake for my son here yesterday. It was the most amazing cake I've ever seen! The decorations were outstanding, and all the kids loved the chocolate icing. I'll definitely come back!</td>
<td>This place is nice for some baked goods, but some things are really nasty. The loaf of bread I bought was stale! They were happy to take it back and give me another, but I'll be watching next time.</td>
</tr>
</tbody>
</table>

- **Consensus** (both positive and negative)
  What’s the best/worst aspect of this product?

- **Conflicts of opinion**
  What aspects do people disagree about?

Their **bread** is so delicious!  The **loaf of bread** I bought was stale!
Task: Input

Input:

– Food-related snippets from restaurant reviews
  • Concise description of a user’s opinion

– Automatically extracted from full review text
  We went to the restaurant, and the sushi was incredibly fresh.

– Segmented by restaurant, but no additional annotation

<table>
<thead>
<tr>
<th>Japanese Restaurant</th>
<th>Bakery</th>
</tr>
</thead>
<tbody>
<tr>
<td>the sushi was so incredibly fresh</td>
<td>I’d recommend the apple pie</td>
</tr>
<tr>
<td>best chicken katsu in town</td>
<td>the bread was disappointingly stale</td>
</tr>
<tr>
<td>drinks are fun, fresh, and delicious</td>
<td>chocolate torte is the stuff of dreams</td>
</tr>
</tbody>
</table>
Task: Output

Output:

- Relevant aspects for each restaurant
- Aspect label for each snippet
- Sentiment label for each snippet

<table>
<thead>
<tr>
<th>Mexican Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Burrito</strong></td>
</tr>
<tr>
<td>+ they had a decent burrito</td>
</tr>
<tr>
<td>- the burrito was mediocre at best</td>
</tr>
<tr>
<td>- the burrito was heavily cilantroed</td>
</tr>
<tr>
<td><strong>Salsa</strong></td>
</tr>
<tr>
<td>+ the salsa is incredible</td>
</tr>
<tr>
<td>+ the mango salsa is perfectly diced</td>
</tr>
<tr>
<td>+ hola free chips &amp; salsa</td>
</tr>
</tbody>
</table>
Possible Solution

Use clustering based on lexical similarity

Problem:
Clusters and aspects are not aligned!
Our Solution

• Jointly model aspect and sentiment

• Leverage data to distinguish sentiment and aspect

<table>
<thead>
<tr>
<th>Bakery</th>
<th>Japanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review 1</td>
<td>pies, delicious, fresh, salmon, fantastic, smooth</td>
</tr>
<tr>
<td></td>
<td>cookies, cakes, pies, maki, sake, beautiful, fresh</td>
</tr>
<tr>
<td>Review 2</td>
<td>cakes, fantastic, pies, salmon, maki, beautiful, fresh</td>
</tr>
<tr>
<td></td>
<td>bread, beautiful, stale, maki, delicious, bland</td>
</tr>
<tr>
<td>Review 3</td>
<td>cakes, beautiful, bread, maki, delicious, bland</td>
</tr>
</tbody>
</table>
Model: Overview

• Each snippet has an aspect and a sentiment

• Each word is drawn from a topic distribution:
  – Aspects are specific to a single product
    
    pizza dessert burrito
  – Sentiment is global across all products
    
    great horrible amazing
  – Background distribution is global
    
    was our food

• Transition distribution encodes word topic transitions

  They had wonderful appetizers.
Model: Generative Story

1. Global distributions

2. Restaurant-level distributions

3. Snippet-level latent structure

4. Words
Model: Generative Story

Globally,

a. **Background distribution** $\theta_B$
   word distribution for stop words and in-domain white noise

b. **Sentiment distributions** $\theta^+_S, \theta^-_S$
   word distributions over positive and negative sentiment words
   small bias for seed words

c. **Transition distribution** $\Lambda$
   first-order Markov distribution of word topic transitions
Model: Generative Story

For each restaurant $i$,

a. Aspect distributions $\theta_{A}^{i,k}$
   - word distribution for each aspect

b. Aspect-sentiment binomials $\phi^{i,k}$
   - probability of positive vs. negative sentiment for each aspect

c. Aspect multinomial $\psi^{i}$
   - probability of each aspect
Model: Generative Story

For each snippet $j$ from restaurant $i$,

a. Aspect $Z^{i,j}_{A}$
   chosen from aspect multinomial $\psi^i$

b. Sentiment $Z^{i,j}_{S}$
   chosen from aspect-sentiment binomial $\phi^{i,k}$

c. Sequence of word topics $Z^{i,j,w}_{W}$
   Background, Aspect, or Sentiment
   selected from transition distribution $\Lambda$

Word topic sequence $\Lambda$
Model: Generative Story

For each word $w$, 

a. Word 
chosen from topic-specific distribution 
based on word topic sequence 

Word topic sequence 

| B | A | B | S | S |
---|---|---|---|---|

The pizza was really great
Standard Variational Inference

• Desired posterior:

\[
P(\psi, \theta_A, \theta_B, \theta_S, \phi, Z|s)
\]
Standard Variational Inference

• Desired posterior:

\[ P(\psi, \theta_A, \theta_B, \theta_S, \phi, Z|s) \]

• Optimizing directly is intractable

• Instead, optimize variational objective with mean-field factorization:

\[ \arg\min_{Q(\cdot)} KL(P(\psi, \theta_A, \theta_B, \theta_S, \phi, Z|s) \parallel Q(\psi, \theta_A, \theta_B, \theta_S, \phi, Z)) \]

\[ \text{s.t. } Q(\cdot) \text{ factorizes} \]
Standard Variational Inference

- **Mean-field factorization:**

\[
Q(\psi, \theta_A, \theta_B, \theta_S, \phi, Z) = q(\theta_B) \left( \prod_{a=1}^{N} q(\theta^a_S) \right) \left( \prod_{i}^{n} q(\psi^i) \left( \prod_{k=1}^{K} q(\theta^{i,k}_A)q(\phi^{i,k}) \right) \left( \prod_{j} q(Z_{S}^{i,j})q(Z_{A}^{i,j}) \prod_{w} q(Z_{W}^{i,j,w}) \right) \right)
\]

- **Optimization strategy:**

minimize each \( q(\cdot) \) while holding others constant (coordinate descent)
Data Set

Food-related snippets from Yelp restaurant reviews
- 13,879 total snippets
- 328 restaurants
- 42.1 snippets per restaurant
- 7.8 words per snippet

Seed words for sentiment distributions
- 42 positive, 33 negative
- Relevant to domain (e.g., “delicious”)
Experiments: Aspect Clustering

• Gold standard
  – Clusters over snippets from 20 restaurants

• MUC cluster evaluation metric
  – Based on number of cluster merges and splits required to achieve gold data
Experiments: Aspect Clustering

• Baseline: Clustering weighted by TF*IDF
  – CLUSTER-ALL uses all unigrams
  – CLUSTER-NOUN uses nouns only

• All systems allowed 10 clusters per restaurant
Experiments: Aspect Clustering

MUC $F_1$

- Cluster-All cannot distinguish between aspect and sentiment
  - The sushi was the best I’d ever had

- Cluster-Noun has a limited set of information
  - Pizza tasted like cardboard
  - Blackened chicken was delicious
Experiments: Sentiment Analysis

• Gold standard
  – 645 snippets (533 train / 112 test)
  – Neutral / ambiguous snippets eliminated
  – Manually labeled POSITIVE or NEGATIVE

• Baselines
  – DISCRIMINATIVE: binary classifier over unigrams
    • Small: 100 training snippets
    • Large: 533 training snippets
  – SEED: positive and negative seed word counts
Experiments: Sentiment Analysis

- Seed performs well due to domain-specific words
- Discriminative improves with additional training data but plateaus
Experiments: Sentiment Analysis

Accuracy with additional training data

- **Our model**: 80.4
- **Seed**: 77.7

Accuracy values:
- 74.1
- 76.8
- 79.5
- 78.6
Error Analysis

- Goal:
  - Identify common errors

- Procedure:
  - Examine 102 clusters
  - Manually annotate correctness of aspect and sentiment
Error Analysis

Similar number of aspect and sentiment errors

Aspect errors
- Similar aspect words in different contexts
  - the blackened chicken was meh
  - chicken enchiladas are yummy
  - the cream cheese wasn’t bad
  - ice cream was just delicious

Sentiment errors
- Rare sentiment words
  - belgian frites are very crave-able
- Negation, sometimes
  - the cream cheese was n’t bad
Aspect identification with shared aspects

• For some domains, aspects are shared corpus-wide
• Medical text: doctor-patient visit summaries

Dear Dr. Smith,
I had the pleasure of seeing your patient, John Doe on September 20. ... He was exposed to lead-based paint chips. ...
John speaks in 2-word phrases. ...

**HEENT** within normal limits. **Heart** rate and rhythm normal. The **blood lead level** was 16 mcg/dl, the **ZPP** was 35/78. ...
Assessment: Low body burden lead poisoning.
Plan: John will return to the PEHC in one month. ...

• Each document summarizes one visit
• Each patient has similar data mentioned, but mentioned only *once*
• Focus on lab/exam results
Applying the model

- Patients replace restaurants
- Aspects represent tests or physical observations
- Goal: automatic aspect identification/clustering

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lungs</td>
<td>clear bilaterally to auscultation lungs were normal</td>
</tr>
<tr>
<td>Patient’s blood lead level</td>
<td>was 35 Pb: 15</td>
</tr>
<tr>
<td>He was</td>
<td>113 cm in height</td>
</tr>
<tr>
<td>Patient was</td>
<td>100 cm tall</td>
</tr>
<tr>
<td>Heart normal</td>
<td></td>
</tr>
<tr>
<td>Cardio</td>
<td>no murmurs</td>
</tr>
<tr>
<td>Heart</td>
<td>clear S1 S2</td>
</tr>
</tbody>
</table>
Data set

- 6,198 snippets from 271 summaries
- Gold clustering annotated by doctors

- Evaluation metric: MUC cluster evaluation

- Baselines: clustering over unigrams
  - CLUSTER-ALL: all unigrams
  - CLUSTER-NOUN: only nouns
Results: Aspect clustering

- **CLUSTER-ALL** outperforms **CLUSTER-NOUN**, unlike in restaurant domain
- Our model retains high performance
Conclusions

• Relevant content models can be learned jointly with an application task to improve performance

• Learning the structure of relations in text beyond single documents facilitates informative aggregation

• Social media presents several challenges which can be overcome by content modeling
Thanks!

Demo  http://condensr.com

Code, papers:

Incorporating Content Structure for Text Analysis Tasks
   EMNLP 2010  http://groups.csail.mit.edu/rbg/code/content_structure

Content Models with Attitude
   ACL 2011  http://groups.csail.mit.edu/rbg/code/content_attitude

Automatic Aggregation by Joint Modeling of Aspects and Values
   JAIR, in revision  http://groups.csail.mit.edu/rbg/code/review-aggregation