A Bayesian Model for Joint Unsupervised Induction of Sentiment, Aspect and Discourse Representations

Angeliki Lazaridou ¹  Ivan Titov ²  Caroline Sporleder ³

¹University of Trento ²University of Amsterdam ³Trier University

ACL 2013
What is Aspect-Based Sentiment Analysis

- Searching for a hotel in Sofia ...
  - Sheraton overall received positive reviews ...
  - ... but does it have a nice view?
Introduction

What is Aspect-Based Sentiment Analysis

- Searching for a hotel in Sofia ...
  - Sheraton overall received positive reviews ...
  - ... but does it have a nice view?
- Prohibitive number of reviews to go through!

Aspect-Based Sentiment Analysis becomes a popular task
[Turney and Littman, 2002, Popescu and Etzioni, 2005, Mei et al., 2007, Titov and McDonald, 2008, Zhao et al., 2010] ...
Why do we need Aspect-Based Sentiment Analysis

Having for every sentence or (even better!) for every phrase the sentiment and the aspect we could ...

1. structure single reviews

   “Lovely hotel”
   Reviewed May 6, 2013
   I stayed here in April 2013. The hotel is lovely, and kept in great condition. Centrally located. Perhaps the best aspect of the hotel is the friendly staff. They were very helpful with everything. It is one of the nicest hotels in Sofia and great value for money.
   Stayed April 2013, traveled on business

   Value  Location  Sleep Quality  Rooms  Cleanliness  Service

2. aggregate results for the product across reviews

   Apple iPad Wi-Fi 16 GB - 3rd generation - Black
   $420 online
   4.5 stars out of 5, 1,135 reviews
   Write a review
   #6 in Apple Tablet Computers
   March 2012 - Apple - Handheld - 16 GB - iOS - Wi-Fi Only - 9.7 inch - With Camera
   Reviews
   1, 1,35 reviews
   5 stars  4 stars  3 stars
   What people are saying
   ease of use  battery  value  picture/video size  screen resolution  graphics
   "Great and easy to use"  "Battery is also very impressive"  "Great price, good product."  "The picture quality is awesome."  "Great resolution and light in weight."  "The screen resolution is great."  "Nice camera, Graphics excellent with hd"

3. Just a step away from creating product summaries!
Discourse: We need more than content

- **Goal:** Identify *sentiments* and *aspects* ...
- Only content (i.e. *lexical features*) can be uninformative and ambiguous.
  - Is the opinion about the view *positive* or *negative*?

Example

*let’s not talk about the view.*
Introduction

Discourse: We need more than content

- **Goal:** Identify *sentiments* and *aspects* ...
- Only content (i.e. *lexical features*) can be uninformative and ambiguous.
  - Is the opinion about the view positive or negative?

Example

*and let’s not talk about the view*

- There exists some *linguistic structure* predictive of sentiment flow.
  - “*and*” constrains the sentiment between the two clauses to be the *same*. 
Goal: Identify sentiments and aspects ...
Only content (i.e. lexical features) can be uninformative and ambiguous.
- Is the opinion about the view positive or negative?

Example

*I’ve never seen such a fancy hotel room...and let’s not talk about the view*

- There exists some linguistic structure predictive of sentiment flow.
  - “and” constrains the sentiment between the two clauses to be the same.
- Exploiting lexical local features while respecting constraints imposed by discourse is a promising direction.
Introduction

Discourse in Sentiment Analysis so far...

- Use **polarity shifters** [Polanyi and Zaenen, 2004, Nakagawa et al., 2010]
- Use **discourse relations** as obtained from **discourse parsers** [Taboada et al., 2008] or by **mapping discourse connectives** to (a subset of) discourse relations [Zhou et al., 2011]
  - Pipeline process results in error propagation
  - Generic discourse relations model not so relevant phenomena for Sentiment Analysis
  - Fail to capture task-specific phenomena → the only thing and overall tell us something about **sentiment** and **aspect** transitions!
- [Somasundaran et al., 2009]
  - introduce task-specific discourse relations that enforce constraints on sentiment
  - proven very **helpful** for the task of Sentiment Analysis
  - still assume access to **perfect** oracle discourse information at test time
Desiderata for Discourse in Aspect-Based Sentiment Analysis

- Encode discourse information relevant for Aspect-Based Sentiment Analysis
- Capture transitions of sentiment and aspect
- Avoid defining mapping from discourse connectives to discourse relations
  - Induce discourse cues that are discriminative for the task
- Avoid gold standard annotation for discourse relations
  - Induce discourse relations jointly with sentiment and aspect
Joint induction of Sentiment, Aspect and Discourse Representations

Why should joint induction work anyway?

Example

*The bathroom was spacious with a lot of space to move, but it was very dirty*
Joint induction of Sentiment, Aspect and Discourse Representations

Why should joint induction work anyway?

Example

.....some aspect....., but it .....the same aspect....

- Induction of aspect and sentiment is **driven by** discourse
- What follows **but it** will probably refer to the same **aspect** but with different **sentiment**, i.e. **negative**
Joint induction of Sentiment, Aspect and Discourse Representations

Why should joint induction work anyway?

Example

... bathroom ...., X ...very dirty....

- Induction of aspect and sentiment is **driven by** discourse
  - What follows **but it** will probably refer to the same **aspect** but with different **sentiment**, i.e. **negative**
- Aspect and sentiment **can signal** the presence of discourse relations and discourse cues
  - Different **sentiments** but the same **aspect** around **but it** signal that probably it serves as a discourse connective for some discourse relation
Modeling Discourse Structure

- Discourse relations can exist between linguistically meaningful adjacent fragments, Elementary Discourse Units (EDUs).
  - Discourse segmentation is obtained automatically.

- **Main Idea:** Each relation between the current and the previous EDU encodes soft constraints on its sentiment and aspect.

- Discourse framework inspired by [Somasundaran et al., 2009]

  - **AltSame** Favors changing **sentiment** but keeping same **aspect**
  - **AltAlt** Favors changing **sentiment** and **aspect**
  - **SameAlt** Favors keeping same **sentiment** but changing **aspect**

- Constraints on **sentiment** and **aspect** are operationalized by modeling their transitions as a function of the different discourse relations.
A Bayesian model of Discourse, Sentiment and Aspect

For every EDU we need to infer:
- the sentiment
- the aspect
- the discourse relation
- the discourse cue signaling that relation

We define a generative model $Pr(\theta, D)$ that explains the generation of a set of reviews.

The set of reviews $D$ consists of:
- the words of the reviews
- the global sentiment of the review (practically the only supervision!)

Bayesian model implies marginalizing out model parameters (i.e. unknown distributions):
$$Pr(z, y, cue, rel|D) = \int Pr(z, y, cue, rel|D, \theta)d\theta$$

Inference is done via Collapsed Gibbs Sampling
Generative story: Generate discourse relation

Example

*The bathroom was spacious with a lot of space to move,*

Previous EDU: $z=\text{bathroom}$, $y=\text{positive}$
Generative story: Generate discourse cue

Example

*The bathroom* was spacious with a lot of space to move, *but it*

Previous EDU: \( z = \text{bathroom}, \ y = \text{positive} \)
Current EDU: \( c = \text{AltSame} \)

According to discourse relation

= "but it"
Generative story: Generate aspect

Example

*The bathroom was spacious with a lot of space to move, but it*

Previous EDU: $z=\text{bathroom}$, $y=\text{positive}$
Current EDU: $c=\text{AltSame}$, cue=but it

![Diagram showing aspects and their relation to discourse](image-url)
Generative story: Generate sentiment

Example

_The bathroom_ was spacious with a lot of space to move, _but it_

Previous EDU: \( z=\text{bathroom}, \ y=\text{positive} \)
Current EDU: \( c=\text{AltSame}, \ \text{cue}=\text{but it}, \ z=\text{bathroom} \)

\[
\begin{align*}
\text{POS} & \quad \text{NEG} & \quad \text{NEU} \\
\text{NEG} & \quad \text{NEG} & \quad \text{NEU}
\end{align*}
\]

\( \times \)

\[
\begin{align*}
\text{POS} & \quad \text{NEG} & \quad \text{NEUTR} \\
\text{POS} & \quad \text{NEG} & \quad \text{NEUTR}
\end{align*}
\]

\( = \text{negative} \)

according to review  according to discourse relation
**Generative story: Generate words**

**Example**

*The bathroom was spacious with a lot of space to move, but it was very dirty*

Previous EDU: $z=$bathroom, $y=$positive

Current EDU: $c=\text{altSame}$, cue=$\text{but it}$, $z=$bathroom, $y=$negative

= dirty, is, very

**negative words for bathroom**
13000 reviews collected from Trip Advisor
From sentences to 320000 EDUs
  - discourse segmentation done with SEGLEX [Tofiloski et al., 2009]
Creating a gold-standard for evaluation
  - 65 randomly selected reviews → 1541 EDUs
  - Aspect annotation (service, value, location, rooms, sleep quality, cleanliness, rest, amenities, food, recommendation) → very skewed distribution
  - Sentiment annotation (-1, +1 and 0) → fairly uniform distribution
  - 9 annotators, 61% IAA in terms of Cohen’s Kappa
Experimental Setup

- Sampler is let to run for 2000 iterations
- 10 aspects, 3 sentiments, 3 discourse relations
- Compare against the discourse-agnostic SentAsp
  - a cross-breed bayesian model between two state-of-the-art models: JST [Lin and He, 2009] and ASUM [Jo and Oh, 2011]
  - obtained by removing all discourse-related information from our model
Direct Clustering Evaluation: Setup

- The model results in partitioning EDUs in clusters encoding sentiment and aspect
- Evaluation inspired by other unsupervised tasks like Word Sense Induction [Agirre and Soroa, 2007]
- To evaluate, we need to find a mapping between induced clusters and classes
  - e.g cluster 3 is labeled as \(\langle\text{negative}, \text{rooms}\rangle\)
- 10-fold cross-validation
  - use 9 folds to induce a 1-1 mapping
  - evaluate the mapping on 10th fold
- Random Baseline: assigns a random label for sentiment and aspect respecting the distribution of labels in the training dataset
Direct Clustering Evaluation: Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>3.9</td>
<td>3.8</td>
<td>3.8</td>
</tr>
<tr>
<td>SentAsp</td>
<td>15.0</td>
<td>10.2</td>
<td>9.2</td>
</tr>
<tr>
<td>Discourse</td>
<td>16.5</td>
<td>13.8</td>
<td>10.8</td>
</tr>
</tbody>
</table>

- *Random* is very low, 28 labels in total → Challenging evaluation
- Latent information about discourse results in significantly higher performance over a discourse-agnostic model
Is our model able to do better in the cases where a discourse relation is explicit?

- **“Marked”**: EDUs that start with a “traditional” discourse connectives present in Penn Discourse Treebank [Prasad et al., 2008]

<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Aspect</th>
<th>Sentiment</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><em>but</em> certainly off its greatness</td>
<td>value</td>
<td>neg</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td><em>and</em> while small <em>they</em> are nice</td>
<td>rooms</td>
<td>pos</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td><em>but</em> it is not free for all guests</td>
<td>amenities</td>
<td>neg</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td><em>and</em> the water was brown</td>
<td>clean</td>
<td>neg</td>
<td>aspect ambiguity</td>
</tr>
<tr>
<td>5</td>
<td><em>and</em> no tea making facilities</td>
<td>rooms</td>
<td>neg</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td><em>when</em> i checked out</td>
<td>service</td>
<td>pos</td>
<td>uninformative EDUs</td>
</tr>
<tr>
<td>7</td>
<td><em>and</em> if you do not</td>
<td>service</td>
<td>neg</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td><em>when</em> we got home</td>
<td>clean</td>
<td>neu</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>Unmarked</th>
<th>Marked</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>SentAsp</em></td>
<td>9.2</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td><em>Discourse</em></td>
<td>9.3</td>
<td>11.5</td>
</tr>
</tbody>
</table>

- When no discourse relation is present, *Discourse* performs as good as *SentAsp* → if we drop discourse-related information one is left with *SentAsp*
- *Discourse* improves results over the challenging cases
  - Model able to leverage “traditional” discourse signal, although is application-specific
  - We are indeed modeling discourse-related information
### What do we really learn?

**Discourse cues predictive for the discourse class**

<table>
<thead>
<tr>
<th>Discourse relation</th>
<th>Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>SameAlt</td>
<td>the location is, the room was, the hotel has, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, <strong>in addition</strong>, good luck</td>
</tr>
<tr>
<td>AltSame</td>
<td>but, and, it was, and it was, and they, although, and it, but it, <strong>but it was</strong>, however, which was, which is, which, this is, this was, they were, the only thing, even though, <strong>unfortunately</strong>, needless to say, fortunately</td>
</tr>
<tr>
<td>AltAlt</td>
<td>the room was, the hotel is, the staff were, <strong>the only</strong>, the hotel is, but the, however, also, or, <strong>overall I</strong>, unfortunately, <strong>we will definitely</strong>, on the plus, the only downside, even though, and even though, i would definately</td>
</tr>
</tbody>
</table>
# What do we really learn?

## Task-specific discourse cues

<table>
<thead>
<tr>
<th>Discourse relation</th>
<th>Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>SameAlt</td>
<td>the location is, the room was, the hotel has, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, in addition, <strong>good luck</strong></td>
</tr>
<tr>
<td>AltSame</td>
<td>but, and, it was, and it was, and they, although, and it, but it, but it was, however, which was, which is, which, this is, this was, they were, <strong>the only thing</strong>, even though, <strong>unfortunately, needless to say</strong>, fortunately</td>
</tr>
<tr>
<td>AltAlt</td>
<td><strong>the room was</strong>, the hotel is, the staff were, <strong>the only</strong>, the hotel is, but the, however, also, or, <strong>overall I</strong>, unfortunately, <strong>we will definitely</strong>, <strong>on the plus, the only downside</strong>, even though, and even though, <strong>i would definately</strong></td>
</tr>
</tbody>
</table>
### What do we really learn?

#### “Traditional” discourse connectives

<table>
<thead>
<tr>
<th>Discourse relation</th>
<th>Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>SameAlt</td>
<td>the location is, the room was, the hotel has, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, <strong>in addition</strong>, good luck</td>
</tr>
<tr>
<td>AltSame</td>
<td><strong>but</strong>, <strong>and</strong>, it was, and it was, and they, <strong>although</strong>, and it, but it, but it was, <strong>however</strong>, which was, which is, which, this is, this was, they were, the only thing, <strong>even though</strong>, unfortunately, needlessly to say, fortunately</td>
</tr>
<tr>
<td>AltAlt</td>
<td>the rooms was, the hotel is, the staff were, the only, the hotel is, but the, however, also, or, overall I, unfortunately, we will definitely, on the plus, the only downside, even though, <strong>and even though</strong>, i would definitely</td>
</tr>
</tbody>
</table>
Features in Supervised Learning: Setup

- Supervised task: classify sentiment and aspect of EDUs
- Every EDU is represented by a bag-of-words concatenated with the latent sentiment and aspect as produced by the SentAsp and Discourse
- 3 Models:
  - only unigrams: only bag-of-words for EDUs
  - unigrams + SentAsp: bag-of-words and aspect and sentiment as predicted by SentAsp
  - unigrams + Discourse: bag-of-words and aspect and sentiment as predicted by Discourse
- SVM with polynomial kernel and 10-fold cross validation
How informative are the latent information produced by the topic?

<table>
<thead>
<tr>
<th>Features</th>
<th>aspect + sentiment</th>
<th>aspect</th>
<th>sentiment</th>
<th>Marked only sentiment + aspect</th>
</tr>
</thead>
<tbody>
<tr>
<td>only unigrams</td>
<td>36.3</td>
<td>49.8</td>
<td>57.1</td>
<td>26.2</td>
</tr>
<tr>
<td>unigrams + SentAsp</td>
<td>38.0</td>
<td>50.4</td>
<td>59.3</td>
<td>27.8</td>
</tr>
<tr>
<td>unigrams + Discourse</td>
<td><strong>39.1</strong></td>
<td><strong>52.4</strong></td>
<td><strong>59.4</strong></td>
<td><strong>29.1</strong></td>
</tr>
</tbody>
</table>

- Incorporating information from topic-model on *only unigrams* improves performance → The clusters are informative
- Results for sentiment prediction comparable to sentence-level results of [Täckström and McDonald, 2011]
- Features from *Discourse* result in higher performance both in the complete and *Marked* examples
Conclusions

- First research that treats the problem jointly in a weakly supervised framework
  - Completely unsupervised for the discourse!
- Modeling of discourse structure improves the results over state-of-the-art discourse-agnostic models
- Induction of meaningful discourse structure for the task of Aspect-Based Sentiment Analysis
- Qualitative analysis showed that our discourse framework has linguistic basis
Future Work

- Induce discourse segmentation within our model.
- Experiment with more discourse relations
  - Model constraints that signaled by the previous EDU

Example

In addition to our spacious room, the shower was fantastic.

- Can we model implicit discourse relations?
Thank you for your attention!
The generative story for the joint model

Global parameters:

\[ \tilde{\varphi} \sim \text{Dir}(\nu) \quad \text{[distrib of disc rel]} \]

for each discourse relation \( c = 1, \ldots, 4 \):

\[ \tilde{\phi}_c \sim \text{DP}(\eta, G_0) \quad \text{[distrib of disc rel specific disc cues]} \]

\[ \tilde{\theta}_{c,k} - \text{fixed} \quad \text{[distrib of rel specific aspect transitions]} \]

\[ \tilde{\psi}_{c,y} - \text{fixed} \quad \text{[distrib of rel specific sent transitions]} \]

for each aspect \( k = 1, 2, \ldots, K \):

for each sentiment \( y = -1, 0, +1 \):

\[ \phi_{k,y} \sim \text{Dir}( \lambda_k ) \quad \text{[unigram language models]} \]

for each global sentiment \( \tilde{y} = -1, 0, +1 \):

\[ \psi_{\tilde{y},k} \sim \text{Dir}( \gamma ) \quad \text{[sent distrib given overall sentiment]} \]

Data Generation:

for each document \( d \):

\[ \tilde{y}_d \sim \text{Unif}(-1, 0, +1) \quad \text{[global sentiment]} \]

\[ \theta_d \sim \text{Dir}(\alpha) \quad \text{[distr over aspects]} \]

for every EDU \( s \):

\[ c_{d,s} \sim \tilde{\varphi} \quad \text{[draw disc relation]} \]

if \( c_{d,s} \neq \text{NoRelation} \)

\[ \tilde{w}_{d,s} \sim \tilde{\phi}_{c_{d,s}} \quad \text{[draw disc cue]} \]

\[ z_{d,s} \sim \theta_d \star \tilde{\phi}_{c_{d,s}, z_{d,s-1}} \quad \text{[draw aspect]} \]

\[ y_{d,s} \sim \psi_{\tilde{y}_d, z_{d,s}} \star \tilde{\psi}_{c_{d,s}, y_{d,s-1}} \quad \text{[draw sentiment level]} \]

for each word after disc cue:

\[ z_{d,s} \sim \phi_{z_{d,s}, y_{d,s}} \quad \text{[draw words]} \]


A syntactic and lexical-based discourse segmenter.

Unsupervised learning of semantic orientation from a hundred-billion-word corpus.

Jointly modeling aspects and opinions with a maxent-lda hybrid.

Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities.