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

Angeliki Lazaridou [1] Ivan Titov [2] Caroline Sporleder [3]

[1]University of Trento [2]University of Amsterdam [3]Trier 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**?

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!

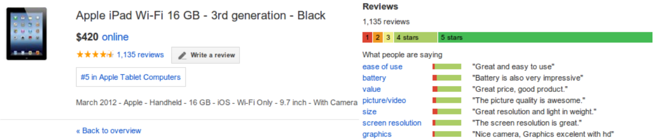
- **Aspect**-Based Sentiment Analysis becomes a popular task [Turney and Littman, 2002, Popescu and Etzioni, 2005, Mei et al., 2007, Titov and McDonald, 2008, Zhao et al., 2010] ...

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**

2. aggregate results for the product **across reviews**



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 uniformative and ambiguous.
  - Is the opinion about the view positive or negative?

Example

let's not talk about the ***view****.*

Discourse: We need more than content

- **Goal:** Identify **sentiments** and **aspects** ...
- Only content (i.e. **lexical features**) can be uniformative 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" constraints the sentiment between the two clauses to be the **same**.

Discourse: We need more than content

- **Goal:** Identify **sentiments** and **aspects** ...
- Only content (i.e. **lexical features**) can be uniformative 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" constraints the sentiment between the two clauses to be the **same**.
- Exploiting lexical local features while respecting constraints imposed by discourse is a promising direction.

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 tor 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 presense 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

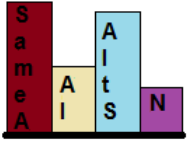
- For every EDU we need to infer:
    - the **sentiment**
    - the **aspect**
    - the discourse relation
    - the discourse cue signaling that relation
- We define a generative model $Pr(\theta, D)$ that explains the generation of a set of reviews
- The set of reviews D consists of:
    - the words 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=bathroom, y=positive



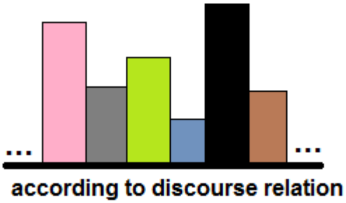
= AltSame

Generative story: Generate discourse cue

Example

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

Previous EDU: z=bathroom, y=positive
Current EDU: c=AltSame



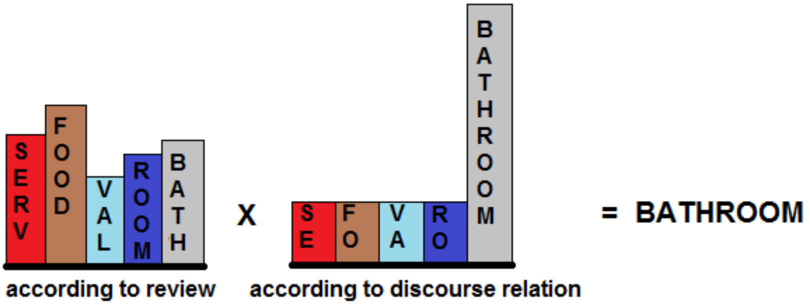
= "but it"

Generative story: Generate aspect

Example

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

Previous EDU: z=**bathroom**, y=positive
Current EDU: c=**AltSame**, cue=but it

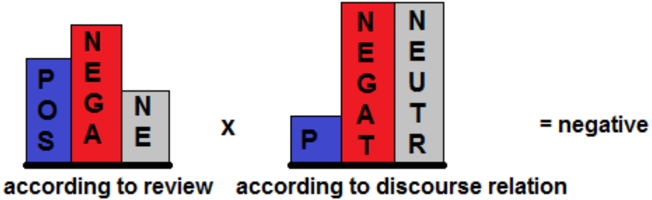


Generative story: Generate sentiment

Example

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

Previous EDU: z=bathroom, y=**positive**
Current EDU: c=**AltSame**, cue=but it, z=**bathroom**

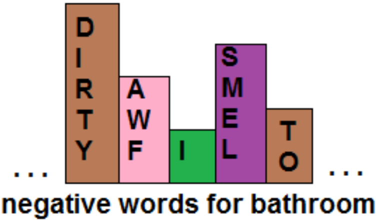


Generative story: Generate words

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=altSame, cue=but it, z=**bathroom**, y=**negative**



= dirty, is, very

Dataset

- 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 Kohen'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 partiotioning EDUs in clusters encoding **sentiment** and **aspect**
- Evaluation inspired by other 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 ⟨*negative*, *rooms*⟩
- 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

Model	Precision	Recall	F1
Random	3.9	3.8	3.8
SentAsp	15.0	10.2	9.2
Discourse	**16.5**	**13.8**	**10.8**

- *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]

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

Model	Unmarked	Marked
SentAsp	9.2	5.4
Discourse	**9.3**	**11.5**

- 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

Discourse relation	Cues
SameAlt	**the location is** , the room was, the hotel has, the hotel, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, **in addition**, good luck
AltSame	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, the only thing, even though, **unfortunately**, needless to say, fortunately
AltAlt	the room 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, and even though, i would definately

What do we really learn?

Task-specific discourse cues

Discourse relation	**Cues**
SameAlt	**the location is**, the room was, the hotel has, the hotel, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, in addition, **good luck**
AltSame	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, **the only thing**, even though, **unfortunately**, **needless to say**, fortunately
AltAlt	**the room 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, and even though, **i would definately**

What do we really learn?

“Traditional” discourse connectives

Discourse relation	Cues
SameAlt	the location is , the room was, the hotel has, the hotel, the hotel is, and the room, and the bed, breakfast was, our room was, the staff were, **in addition**, good luck
AltSame	**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, the only thing, **even though**, unfortunately, needless to say, fortunately
AltAlt	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, **and even though**, i would definately

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?

Features	**aspect+sentiment**	**aspect**	**sentiment**	**Marked only sentiment+aspect**
only unigrams	36.3	49.8	57.1	26.2
unigrams + SentAsp	38.0	50.4	59.3	27.8
unigrams + Discourse	**39.1**	**52.4**	**59.4**	**29.1**

- 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 in 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:

$\tilde{\varphi} \sim Dir(\nu)$ [distrib of disc rel]
for each discourse relation $c = 1, .., 4$:
  $\tilde{\phi}_c \sim \mathrm{DP}(\eta, G_o)$ [distrib of disc rel specific disc cues]
  $\tilde{\theta}_{c,k}$ - fixed [distrib of rel specific aspect transitions]
  $\tilde{\phi}_{c,y}$ - fixed [distrib of rel specific sent transitions]
for each aspect $k = 1, 2...K$:
  for each sentiment $y = -1, 0, +1$:
    $\phi_{k,y} \sim Dir(\boldsymbol{\lambda}_k)$ [unigram language models]
  for each global sentiment $\hat{y} = -1, 0, +1$:
    $\psi_{\hat{y},k} \sim Dir(\boldsymbol{\gamma})$ [sent distrib given overall sentiment]

Data Generation:

for each document d:
  $\hat{y}_d \sim Unif(-1, 0, +1)$ [global sentiment]
  $\theta_d \sim Dir(\alpha)$ [distr over aspects]
  for every EDU s:
    $c_{d,s} \sim \tilde{\varphi}$ [draw disc relation]
    if $c_{d,s} \neq NoRelation$
      $\tilde{w}_{d,s} \sim \tilde{\phi}_{c_{d,s}}$ [draw disc cue]
    $z_{d,s} \sim \theta_d * \tilde{\theta}_{c_{d,s},\, z_{d,s-1}}$ [draw aspect]
    $y_{d,s} \sim \psi_{\hat{y}_d, z_{d,s}} * \tilde{\psi}_{c_{d,s}, y_{d,s-1}}$ [draw sentiment level]
    for each word after disc cue:
      $w_{d,s} \sim \phi_{z_{d,s}, y_{d,s}}$ [draw words]

References I

Agirre, E. and Soroa, A. (2007).
Semeval-2007 task 02: Evaluating word sense induction and discrimination systems.
In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*.

Jo, Y. and Oh, A. H. (2011).
Aspect and sentiment unification model for online review analysis.
In *Proceedings of the fourth ACM international conference on Web search and data mining*.

Lin, C. and He, Y. (2009).
Joint sentiment/topic model for sentiment analysis.
In *Proceeding of CIKM*, pages 375–384.

Mei, Q., Ling, X., Wondra, M., Su, H., and Zhai, C. (2007).
Topic sentiment mixture: modeling facets and opinions in weblogs.
In *Proceedings of the 16th international conference on World Wide Web*.

Nakagawa, T., Inui, K., and Kurohashi, S. (2010).
Dependency tree-based sentiment classification using crfs with hidden variables.
In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

Polanyi, L. and Zaenen, A. (2004).
Contextual valence shifters.
In *Working Notes — Exploring Attitude and Affect in Text: Theories and Applications (AAAI Spring Symposium Series)*.

References II

Popescu, A.-M. and Etzioni, O. (2005).
Extracting product features and opinions from reviews.
In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., and Webber, B. (2008).
The penn discourse treebank 2.0.
In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.

Somasundaran, S., Namata, G., Wiebe, J., and Getoor, L. (2009).
Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification.
In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*.

Taboada, M., Voll, K., and Brooke, J. (2008).
Extracting sentiment as a function of discourse structure and topicality.

Täckström, O. and McDonald, R. (2011).
Semi-supervised latent variable models for sentence-level sentiment analysis.
In *Proceedings of ACL*, pages 569–574.

Titov, I. and McDonald, R. (2008).
Modeling online reviews with multi-grain topic models.
In *Proceedings of the 17th International Conference on World Wide Web (WWW)*.

References III

Tofiloski, M., Brooke, J., and Taboada, M. (2009).
A syntactic and lexical-based discourse segmenter.
In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*.

Turney, P. D. and Littman, M. L. (2002).
Unsupervised learning of semantic orientation from a hundred-billion-word corpus.

Zhao, W. X., Jiang, J., Yan, H., and Li, X. (2010).
Jointly modeling aspects and opinions with a maxent-lda hybrid.
In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

Zhou, L., Li, B., Gao, W., Wei, Z., and Wong, K.-F. (2011).
Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities.
In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.