Discovering Correspondence of Sentiment Words and Aspects

Geli Fei[1], Zhiyuan (Brett) Chen[1], Arjun Mukherjee[2], Bing Liu[1]

[1]Department of Computer Science, University of Illinois at Chicago, Chicago, IL, USA
gfei2@uic.edu, czyuanacm@gmail.com, liub@cs.uic.edu
[2]Department of Computer Science, University of Houston, TX, USA
arjun@cs.uh.edu

Abstract. Extracting aspects and sentiments is a key problem in sentiment analysis. Existing models rely on joint modeling with supervised aspect and sentiment switching. This paper explores unsupervised models by exploiting a novel angle – correspondence of sentiments with aspects via topic modeling under two views. The idea is to split documents into two views and model the topic correspondence across the two views. We propose two new models that work on a set of document pairs (documents with two views) to discover their corresponding topics. Experimental results show that the proposed approach significantly outperforms strong baselines.

1 Introduction

Finding topics in documents is an important problem for many NLP applications. One of the effective techniques is topic modeling. Over the years, many topic models have been proposed which are extensions or variations of the basic models such as PLSA [11] and LDA [3]. These models typically take a set of documents and discover a set of topics from the documents. They may also produce some other types of auxiliary information at the same time using joint modeling. Topic models are also widely used in sentiment analysis for finding product or service aspects (or features) and opinions about them.

In this paper, we focus on discovering aspect specific sentiments. For example, in the restaurant reviews, positive sentiment words such as “tasty” and “delicious” are usually associated with the *food* aspect, and positive sentiment words “friendly” and “helpful” are about the *staff* aspect. Modeling aspects (topical words) with aspect specific sentiments (opinion words) is very useful. First, it can improve opinion target detection [16, 23]. For example, in the sentence, “*I had sushi in Sakura’s on Washington St, which was really tasty*,” there is a positive opinion indicated by “tasty”. However, it is not easy to determine the target of the positive opinion – “sushi”, “Sakura’s” or “Washington St.” However, if we know that “sushi” appears in a food topic and “tasty” appears in its corresponding sentiment topic, then we know that “tasty” is about “sushi”, and not about “Sakura’s” or “Washington St”. Second, the results can also help co-reference resolution. For instance, in “*I had sushi in Sakura’s on Washington St. It was really tasty*”, it is not easy to know what “it” refers to. Topical correspondence can resolve “it” to “sushi.”

This paper proposes two paired topic models to discover correspondence between two views, i.e., source view and target view, which correspond to aspect and sentiment, respectively.

The first model is a directional model, ASL (Aspect Sentiment LDA) that explicitly models topic correspondence via conditioning target topics on source topics. ASL does not consider target topics while inducing source topics, which is a weakness.

This motivates us to propose the second model IASL (Integrated Aspect and Sentiment LDA), which additionally improves the topic discovery in source and target documents. Unlike ASL, IASL is not directional. It can improve ASL because in inducing the source topics it also considers words in the target documents and vice versa. Existing models and ASL are unable to do this. Technically, IASL merges the source and target views into one virtual document and uses an indicator variable to tell the model whether a word is from the source or the target view during inference. This merging allows improved joint modeling that yields much better results.

However, to apply the proposed models, we need document pairs, but each review or review sentence is not a pair. We split each review or review sentence into two parts: a sub-document consisting of sentiment words and a sub-document consisting of non-sentiment words. Thus, given a large number of online reviews, the proposed ASL and IASL models can find the corresponding aspect and sentiment topics. Our experimental results using reviews from both hotel and restaurant domains show that the proposed models are highly effective and significantly outperform relevant existing models PLTM [19] and ASUM [13].

2 Related Work

Topic models have been applied to numerous applications and domains. In sentiment analysis, they have been used to find aspect and/or sentiment terms/words. Related work in this thread include those in e.g., [18, 26, 15, 5, 27, 29, 21, 22, 7]. Although many can model aspects and sentiments jointly, they need supervised aspect/sentiment labels, and none of these above models work on documents that come with two views. One representative model, ASUM in [13], models both aspect and sentiments in reviews. It can, to some extent, extract aspects that are specific to sentiment labels. Thus, we use it as one of our baselines in the experiments.

Although in the context of sentiment analysis, we are not aware of any topic models that work on documents pairs, there are several models in other application fields that have resemblances to our dual view aspect-sentiment topic models. These include the works in [1, 2, 18, 4, 28, 17, 14, 10] that also work on pairs, e.g., authors and papers, images and tags, etc. In [1], Corr-LDA was proposed to work on image and caption pairs. In the model, the generation of caption words is conditioned on image regions. The condition part has some similarity to our ASL model, but ASL’s target topics are conditioned on source topics. The model in [4] used unaligned multi-lingual documents to find shared topics and to pair related documents across languages. Rosen-Zvi et al. [24] proposed an author-topic model, which takes a collection of author lists and their articles as input to find each author's topical interests. Mimno et al. [19] proposed the PLTM model which can take a set of tuples, where each tuple has several corresponding documents. Compared to our IASL model, none of these existing models makes use of word co-occurrences in the target documents in inducing the source topics and vice versa. Among these existing models, PLTM is the closest in function to our models, so we consider it as another baseline model.

3 ASL Model

Given a set of document pairs (where the document is split into two views – source view: containing non-sentiment words and target view containing sentiment words), ASL finds a set of topics (called source topics) from the source documents and at the same time, for each source topic, finds the corresponding target topic in the target documents. The modeling of topic correspondence in ASL is directional, meaning that the discovery of target topics is conditioned on source topics. We use a directed link from the source topic node to the target topic node to explicitly model the dependency of target topics on source topics. Also, the topic discovery in the source documents of ASL is independent and is the same as LDA.

ASL assumes that there are K source topics in the source documents and O target topics in the target documents. It posits K source topic-word distributions in the source view, O target topic-word distributions in the target view, and K source-target topic distributions, i.e., each of the K source topics has a distribution over O target topics, i.e., a distribution over distributions as each target topic is already a distribution over words. Through parameter estimation, we aim to discover the K source topics and the most probable source-topic-specific target topics.

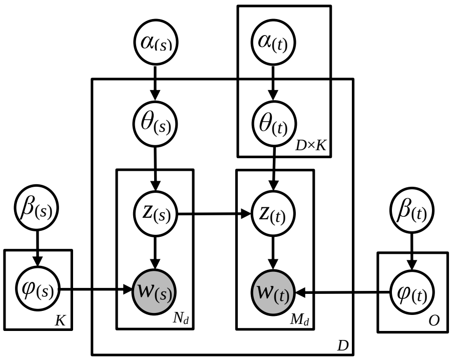

Figure 1: The Plate Notation of ASL

3.1 Generative Process

We follow standard notations. $\varphi_{(s),k}$ and $\varphi_{(t),o}$ denote the source and target topic distributions k and o respectively. $\theta_{(s),d_{(s)}}$ denotes the document-topic distribution of source document $d_{(s)}$. $\theta_{(t),k,d_{(t)}}$ denotes the source-document topic k-specific topic distribution of target document $d_{(t)}$. z and w represent the standard latent topic and observed word in respective views. α, β denote the respective Dirichlet hyperparameters. Subscripts $(s)/(t)$ indicate whether a variable lies in the source or target views of a document. We detail the generative process of ASL (Figure 1) as follows:

The generative process of ASL for a corpus of D document pairs is as follows:

1. For each source document $d_{(s)} \in \{1, \dots, D\}$:

   Draw document-topic distribution $\theta_{(s),d_{(s)}} \sim Dir\ (\alpha_{(s)})$.

2. For each target document $d_{(t)} \in \{1, \dots, D\}$:

   For each source topic $k \in \{1, \dots, K\}$:

   Draw document-source topic-target topic distribution $\theta_{(t),k,d_{(t)}} \sim Dir(\alpha_{(t),k})$.

3. For each source topic $k \in \{1, \dots, K\}$:

   Draw topic-word distribution $\varphi_{(s),k} \sim Dir\ (\beta_{(s)})$.

4. For each target topic $o \in \{1, \dots, O\}$:

   Draw topic-word distribution $\varphi_{(t),o} \sim Dir\ (\beta_{(t)})$.

5. For each source document $d_{(s)} \in \{1, \dots D\}$:

   For each word $w_{(s),d_{(s)},n} \in d_{(s)}$:

   Choose a source topic $z_{(s),d_{(s)},n} \sim Multi\ (\theta_{(s),d_{(s)}})$.

   Choose a source word $w_{(s),d_{(s)},n} \sim Multi\ (\varphi_{(s),z_{(s),d_{(s)},n}})$.

6. For each target document $d_{(t)} \in \{1, \dots D\}$:

   For each word $w_{(t),d_{(t)},m} \in d_{(t)}$:

   Choose a target topic $z_{(t),d_{(t)},m} \sim Multi\ (\theta_{(t),d_{(t)},z_{(s),d_{(s)},*}})$.

   Choose a target word $w_{(t),d_{(t)},m} \sim Multi\ (\varphi_{(t),z_{(t),d_{(t)},m}})$.

Note that in step 6, we use $z_{(s),d_{(s)},*}$ which denotes the dependency of all target topics on source topic s, as while sampling a target topic word, we need to consider all the (source) topic assignment of words in the corresponding source document.

3.2 Inference

We use Gibbs sampling for inference. In sampling, we sample source topics to generate stable source topics first and then sample target topics. As source topics are independent of target topics, we can sample them independently and wait till source topics stabilize to shape target topics by reducing total Gibbs iterations and autocorrelation. The source topic sampling is similar to LDA:

$$p\left(z_{(s),d_{(s)},v}\middle|z_{-(s),d_{(s)},v},w_{(s)},\alpha_{(s)},\beta_{(s)}\right)$$
$$\propto\left(c_{z_{(s),d_{(s)},v},d_{(s)},*}^{-(d_{(s)},v)}+\alpha_{(s),z_{(s),d_{(s)},v}}\right)$$
$$\times\frac{\left(c_{z_{(s),d_{(s)},v},*,w_{(s),d_{(s)},v}}^{-(d_{(s)},v)}+\beta_{(s),w_{(s),d_{(s)},v}}\right)}{\sum_{i=1}^{I}\left(c_{z_{(s),d_{(s)},v},*,i}^{-(d_{(s)},v)}+\beta_{(s),i}\right)}$$

After the source topics stabilize through Gibbs sampling iterations, we sample target topics in the second step using the following Gibbs sampler.

$$p\left(z_{(t),d_{(t)},y}\middle|z_{-(t),d_{(t)},y},z_{(s)},w_{(t)},\alpha_{(t)},\beta_{(t)}\right)\propto$$

$$\prod_{k=1}^{K}\frac{c_{d,k,*,z_{(t),d_{(t)},y},*}^{-(d_{(t)},y)}+\alpha_{(t),k,z_{(t),d_{(t)},y}}}{\sum_{o=1}^{O}\left(c_{d,k,*,o,*}^{-(d_{(t)},y)}+\alpha_{(t),k,o}\right)}\times\frac{c_{z_{(t),d_{(t)},y},*,w_{(t),d_{(t)},y}}^{-(d_{(t)},y)}+\beta_{(t),w_{(t),d_{(t)},y}}}{\sum_{j=1}^{J}\left(c_{z_{(t),d_{(t)},y},*,j}^{-(d_{(t)},y)}+\beta_{(t),j}\right)}$$

$c_{k,u,*}$ is the number of times topic k is assigned to words in document u; $c_{k,*,a}$, the number of times topic k is assigned to word a; $c_{d,k,*,o,*}$, the number of times source topic k is assigned to source words while target topic o is assigned to target words in document pair d. $c^{\neg(u,v)}$ discounts word v in document u.

4 IASL Model

In ASL, the discovery of source topics is independent of target topics. However, the target view can help shape better source topics. Hence, we now propose an integrated model IASL to jointly model both source and target views. Specifically, it merges the source and target documents in each document pair into a virtual document. An indicator variable, y (which is observed) is used to indicate whether a word is from the source or the target. By doing so, we effectively increase the word co-occurrence, which consequently results in better topics and better topic correspondence. IASL assumes that there are K one-to-one correspondence of topics between the source and target. Each source (target) topic is a distribution over the vocabulary in the source (target) documents. The plate notation is shown in Figure 2.

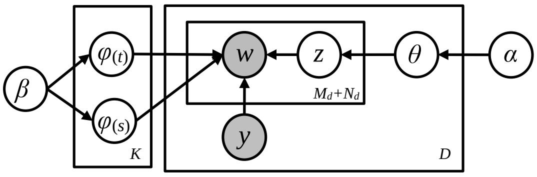

Figure 2: The Plate Notation of IASL

4.1 Generative Process

The generative process of IASL is as follows:

```
For each source and target topic $k \in \{1, \ldots, K\}$:
    draw $\varphi_{(s),k} \sim Dir\ (\beta)$
    draw $\varphi_{(t),k} \sim Dir\ (\beta)$
For each virtual document $d$:
    draw $\theta_d \sim Dir\ (\alpha)$.
For each word $w_{d,u} \in d$:
    draw its topic $z_{d,u} \sim Multi\ (\theta_d)$.
    if $y$ indicates $w_{d,u}$ being sentiment word:
        draw word $w_{d,u} \sim Multi\ (\varphi_{(s),z_{d,u}})$.
    if $y$ indicates $w_{d,u}$ being aspect word:
        draw word $w_{d,u} \sim Multi\ (\varphi_{(t),z_{d,u}})$.
```

4.2 Inference

The Gibbs sampler for assigning topics to words in the documents takes the following form. K is the number of topics and I is the size of vocabulary in the source and target documents combined.

$$p\left(z_{d,u}\middle|z_{-(d,u)},y,w\right) \propto \frac{c_{k,d,*}^{-(d,u)}+\alpha}{\sum_{k=1}^{K}\left(c_{k,d,*}^{-(d,u)}+\alpha\right)} \times \frac{c_{(s),k,*,w_{d,u}}^{-(d,u)}+\beta}{\sum_{i=1}^{I}\left(c_{(s),k,*,w_i}^{-(d,u)}+\beta\right)} \times \frac{c_{(t),k,*,w_{d,u}}^{-(d,u)}+\beta}{\sum_{i=1}^{I}\left(c_{(t),k,*,w_i}^{-(d,u)}+\beta\right)}$$

where $z_{d,u}$ and $z_{-(d,u)}$ represent topic assignment for $w_{d,u}$ in the virtual document d and topic assignment for words except $w_{d,u}$ respectively. $c_{k,d,*}^{-(d,u)}$ represents the number of times topic k assigned to words in document d except $w_{d,u}$. $c_{(s),k,*,w_i}^{-(d,u)}$ and $c_{(t),k,*,w_i}^{-(d,u)}$ represent the number of times topic k assigned to source word w_i except $w_{d,u}$ and the number of times topic k assigned to target word w_i except $w_{d,u}$ respectively.

5 Experiments

We now evaluate the proposed ASL and IASL models. Note that although ME-LDA [29] and ME-SAS [22] discover aspect specific sentiments, they belong to the family of supervised topic models and need labeled training data, seed sets, and also do not model documents in two views that place them in a different problem setting than ours. Instead, we compare our model with PLTM [19] which is unsupervised and designed for document tuples (similar to our dual views) and ASUM [13] which is a representative unsupervised joint aspect and sentiment model. The evaluation task is to use all the models to discover corresponding aspect and sentiment topics in online reviews. ASUM cannot automatically separate aspects from sentiment words, we will post-process its results in order to compare it with our models. The implementation of both PLTM and ASUM was obtained from their authors.

Datasets: We use two datasets: one hotel review dataset from TripAdvisor.com, and one restaurant review dataset from Yelp.com. Our hotel data contains 101,234 reviews and 692,783 sentences, and our restaurant data contains 25,459 reviews and 278,179 sentences. We ran the Stanford Parser to perform sentence detection and lemmatization. The data domain name is removed since it co-occurs with most words in the dataset, leading to an undesirable high overlap among topics/aspects.

Sentences as Documents: Standard topic models tend to produce topics that correspond to global properties of products instead of their aspects when applied to reviews [26]. We take the approach in [5, 7] and divide each review into sentences and treat each sentence as an independent document.

Document Pairs: Treating each sentence as a document, we split each sentence into the source and target pair. The source contains all non-sentiment (aspect) words and the target contains all sentiment words. The sentiment lexicon in [12] was used to find sentiment words in each sentence. We will post-process ASUM results using the same lexicon. Sentences with no sentiment words were ignored.

Parameter Settings: In all our experiments, posterior inference was drawn after 2000 Gibbs iterations with a burn-in of 200 iterations. Following [9], we fix the Dirichlet priors of our models as follows: for all document-topic distributions, we set $\alpha = 50/K$, where K is the number of topics. For ASL, we use the same number of topics for both the source and the target documents. And for all topic-word distributions, we set $\beta = 0.1$. We also experimented with other settings of these priors and did not notice much difference.

Setting the number of topics/aspects in topic models is tricky as it is difficult to know the exact number of topics in a corpus. While non-parametric Bayesian approaches [25] exist for estimating the number of topics, it’s not the focus of this paper. We empirically set the number of topics for both source and target documents to 15. Although 15 may not be optimal, since all models use the same number, there is no bias against any model.

Baseline Model Settings: The input of PLTM is the same as that of our proposed models. ASUM works on a single set of documents and assigns a topic to each sentence. So we treat one review as a document for ASUM input. Also, the output of ASUM is a set of topics called senti-aspects, which are jointly defined by aspects and sentiment labels. In this paper, we only consider positive and negative sentiment, and set 15 topics under each sentiment label (positive or negative). Since ASUM does not separate sentiment words from aspects, we separate them using the sentiment lexicon of [12] during post-processing. For both baselines, we set parameters as mentioned in their original papers.

TABLE 1. HOTEL DATA: ASPECT AND SENTIMENT TOPIC SUMMARY RESULTS

	PLTM			**ASL**			**IASL**		
	P@5	***P@10***	***P@15***	***P@5***	***P@10***	***P@15***	***P@5***	***P@10***	***P@15***
Aspect	0.82	0.75	0.72	0.87	0.83	0.81	0.90	0.85	0.78
Sentiment	0.78	0.76	0.72	0.77	0.78	0.77	0.92	0.86	0.78
Average	**0.80**	**0.76**	**0.72**	**0.82**	**0.81**	**0.79**	**0.91**	**0.86**	**0.78**

TABLE 2. RESTAURANT DATA: ASPECT AND SENTIMENT TOPIC SUMMARY RESULTS

	PLTM			**ASL**			**IASL**		
	P@5	***P@10***	***P@15***	***P@5***	***P@10***	***P@15***	***P@5***	***P@10***	***P@15***
Aspect	0.89	0.80	0.78	0.87	0.85	0.83	0.91	0.86	0.83
Sentiment	0.85	0.79	0.74	0.96	0.91	0.84	0.95	0.9	0.87
Average	**0.87**	**0.80**	**0.76**	**0.92**	**0.88**	**0.84**	**0.93**	**0.88**	**0.85**

TABLE 3. TOPIC COHERENCE OF SENTIEMNT TOPICS AND ASPECT TOPICS

(a) HOTEL

	PLTM	**ASUM**	**ASL**	**IASL**
Aspect	-380.04	-668.03	-378.67	-296.77
Sentiment	-448.83	-596.36	-420.85	-247.55
Average	**-414.43**	**-632.19**	**-399.76**	**-272.16**

(b) RESTAURANT

PLTM	**ASUM**	**ASL**	**IASL**
-382.13	-772.45	-387.27	-262.52
-438.34	-667.24	-404.04	-381.15
-410.235	**-719.84**	**-395.655**	**-321.835**

5.1 Results of Manual Labeling of Topics

Although statistical measures, such as perplexity, KL-divergence and topic coherence, have been used to evaluate topic models, they may not always conform to human notions of semantics [6] or an actual sentiment analysis application. Thus, we first report evaluation by human judges. In the next sub-section, we also report topic coherence results to evaluate our models statistically.

Topic labeling was done by two judges. The labeling was carried out in two stages sequentially: (1) labeling of topics (both aspect and sentiment word topics), and (2) labeling of topical words in each topic. After the first stage, agreement among judges was computed, and then the two judges discussed about the disagreed topics to reach a consensus. They then moved to the next stage to label the top ranked words in each topic. We measured inter-judge agreement using Kappa [8]. The Kappa score for topic labeling and topical words labeling were 0.853 and 0.917 respectively indicating strong agreements in the labeling.

Since we want to find topic correspondence, during labeling, in the first stage we first determine whether an aspect topic is good or bad and then for each aspect topic, we label its corresponding sentiment topic as good or bad. By good or bad, we mean whether a topic is coherent enough to have a distinct topic label from its top ranked words. If the judges could not label a topic due to its incoherence, it was labeled as bad; otherwise good. Further, if a real-world aspect for an aspect topic could not be identified (i.e., it was a bad topic), its corresponding sentiment topic was also not labeled and discarded. We choose this labeling scheme as our objective is to find topic correspondence, and also because that ASL is directional from source to target topics.

Unlike other models that generate sentiment topics, ASUM generates a set of senti-aspects, each of which is specific to a single sentiment label. Since it is hard to match positive senti-aspects with negative senti-aspects that are about the same aspects, we cannot directly compare the results of ASUM with other models. Hence, in this section, we will only label and compare topics of the other three models. We however compare all four models using statistical evaluation metrics in the next section.

Results and Discussions: We use *precision@n* (*P@n*) as our metric. *P@n* is the precision at rank *n* for a topic. The summaries (average over all topics) from both datasets are given in Table 1 and Table 2. We note that as ASL is a conditional model, it tries to ensure the source (aspect in our case) is not negatively affected by the goal of finding corresponding topics from the target. Hence, its aspect and sentiment topics are better than those of PLTM. IASL is able to achieve much better results for both the source (aspect) and target (sentiment) topics due to leveraging word collocations across both the source and target views. The improvements are particularly pronounced for the first 5 and the first 10 words (i.e., P@5 and P@10), which is important as in practical applications they are more trustworthy. To measure the significance of the improvements, we conducted paired t-tests. The tests showed that both ASL and IASL outperform PLTM significantly ($p < 0.01$). IASL also outperformed ASL significantly ($p < 0.05$).

Number of bad topics: Our labeling also gives the number of good and bad topics. For the hotel domain, PLTM has four bad aspect topics. For both ASL and IASL, there are only three bad aspect topics. For the restaurant domain, all three models give four bad topics. For both domains, for each good aspect topic, its corresponding sentiment topic is always good.

5.2 Statistical Evaluation

We use the *topic coherence* (TC) metric in [20] that correlates well with human semantics. Table 3 shows the TC values for aspect and sentiment topics for each dataset (averaged over all topics). Note that the output of ASUM is a set of senti-aspects (a set of words defined by both sentiment and aspect). So using 15 topics and 2 sentiment polarities, ASUM generates 30 senti-aspects. Again, ASUM cannot separate sentiment words from non-sentiment words. In order to compute topic coherence for both aspect topics and sentiment topics, we use the sentiment lexicon to separate each senti-aspect into sentiment topics and aspects. In computing TC values, we used the top 15 words from each topic, which is also the number of top topical words labeled in our human evaluation. TC gives a negative value and a better model should have a higher TC value. Also, TC gives a higher score if less frequent words appear at the top ranks of a topic. So in the statistical evaluation, we remove general seed words from each senti-aspect before separating aspects from sentiment topics in order to be fair to ASUM. This improved the average TC of sentiment topics of ASUM by more than 50. From Table 3, we can see that IASL has highest TC values, which dovetails with human labeling results in Table 1 and Table 2. Its values are markedly better

TABLE 4. EXAMPLE TOPICS EXTRACTED BY PLTM, ASL AND IASL

Bathroom					
PLTM		***ASL***		***IASL***	
bathroom	clean	area	clean	bathroom	small
area	comfortable	bathroom	comfortable	towel	dirty
bed	sink	large	adequate	floor	sink
large	vanity	long	sink	area	adequate
space	hang	provide	vanity	shower	hang
coffee	spacious	space	hang	wall	break
microwave	adequate	chair	spacious	bath	big
small	amenity	separate	quiet	space	vanity
fridge	available	small	friendly	hair	old
desk	safe	counter	poor	separate	stain
Location					
PTLM		***ASL***		***IASL***	
location	close	location	close	location	close
located	clean	located	clean	area	convenient
airport	convenient	minute	quiet	park	attraction
area	quiet	downtown	comfortable	minute	wonderful
shuttle	comfortable	drive	convenient	drive	new
downtown	wonderful	street	safe	downtown	ideal
drive	free	distance	attraction	short	decent
park	attraction	shopping	convenient	mile	clean
shopping	safe	short	reasonable	distance	quiet
price	convenient	main	affordable	airport	difficult

than PLTM and ASUM. ASL is also better than PLTM in all cases except a slight drop for aspect topics for restaurant.

Significance testing using paired t-test on the results of TC showed that IASL significantly improves PLTM, ASUM and ASL ($p < 0.03$). ASL also improves ASUM significantly ($p < 0.05$) but does not improve PLTM significantly ($p = 0.11$). The difference between the results here and those from the human labeling results is understandable because although TC correlates with human labeling well, they are not exactly same.

In summary, we conclude that both human evaluation and statistical evaluation show that the proposed models are more effective than the baseline models PLTM and ASUM. Also, IASL improves upon the other models by a large margin.

5.3 Case Study

This section shows some example aspect and sentiment topic pairs labeled by our human judges. Words in red are labeled as wrong by the judges. For the same reason as in human evaluation section, ASUM generates a set of senti-aspects, each of which is specific to a single sentiment label. It is thus hard to match positive senti-aspects with negative senti-aspects that are about the same aspects and to directly compare the results with other models. Table 4 lists two sets of topic pairs discovered by PLTM, ASL and IASL models. We can see that ASL performs better than PLTM in source topic (aspect) detection, and IASL outperforms both ASL and PLTM in both aspect topics and sentiment topics. The improvement is due to the proposed modeling tailored for document pairs.

6 Conclusion

This paper proposed two new topic models ASL and IASL to jointly model source and target topics and their correspondence for datasets involving document pairs. The ASL model is a directional topic model. The IASL model improves ASL by enabling the inference algorithm to leverage word collocations across both source and target documents while inducing topics. The proposed models have been evaluated on the task of finding sentiment and aspect topics and their correspondence using real-world reviews of hotels and restaurants. Experimental results showed that ASL and IASL outperformed the relevant baseline models PLTM and ASUM markedly.

Acknowledgements. This work was supported in part by a grant from National Science Foundation (NSF) under grant no. IIS-1407927, a NCI grant under grant no. R01CA192240, and a gift from Bosch. The content of the paper is solely the responsibility of the authors and does not necessarily represent the official views of the NSF, NCI, or Bosch.

References

1. Blei, D. and Jordan, M. 2003. Modeling annotated data. In Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 127–134, Toronto, Canada, 2003.
2. Blei, D. and Lafferty, J. 2006. Correlated topic models. In Advances in neural information processing systems 18.
3. Blei, D., Ng, A. and Jordan, M. 2003. Latent Dirichlet Allocation. Journal of Machine Learning Re-search, 3, 993–1022.
4. Boyd-Graber, J., and Blei, D. 2009. Multilingual topic models for unaligned text. In Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, pp. 75-82. AUAI Press, 2009.
5. Brody, S. and Elhadad, N. 2010. An unsupervised aspect-sentiment model for online reviews. In Proceedings of NAACL, pages 804–812.
6. Chang, J., Boyd-Graber, J., Chong, W., Gerrish, S. and Blei, M. 2009. Reading Tea Leaves: How Humans Interpret Topic Models. In Proceedings of NIPS, pages 288–296.
7. Chen, Z., Mukherjee, A., Liu, B., Hsu M., Castellanos, M. and Ghosh, R. 2013. Exploiting Domain Knowledge in Aspect Extraction. In Proceedings of EMNLP, pages 1655–1667.
8. Cohen, J. "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." Psychological bulletin 70, no. 4 (1968): 213.
9. Griffiths, T. and Steyvers, M. 2004. Finding Scientific Topics. PNAS, 101 Suppl, 5228–5235.
10. Gamon, M., Mukherjee A., and Pantel P. "Predicting interesting things in text." COLING, 2014.

11. Hofmann, T. 1999. Probabilistic Latent Semantic Analysis. In Proceedings of UAI, pages 289–296.
12. Hu, M. and Liu, B. 2004. Mining and Summarizing Customer Reviews. In Proceedings of KDD, pages 168–177.
13. Jo, Y. and Oh, A. 2011. Aspect and sentiment unification model for online review analysis. In Proceedings of WSDM, pages 815–824.
14. Kosuke, F., Eguchi, K. and Xing, E. 2012. Symmetric Correspondence Topic Models for Multilingual Text Analysis. In Advances in Neural Information Processing Systems 25, pp. 1295-1303. 2012.
15. Lin, C. and He, Y. 2009. Joint sentiment/topic model for sentiment analysis. In Proceedings of CIKM, pages 375–384.
16. Liu, B. 2010. Sentiment Analysis and Subjectivity. In Handbook of Natural Language Processing.
17. Miao, G., Guan, Z., Moser, L., Yan, X., Tao, S., Anerousis, N. and Sun, J. 2012. Latent association analysis of document pairs. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 1415-1423. ACM, 2012.
18. Mei, Q., Ling, X., Wondra, M., Su, H. and Zhai, C. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In Proceedings of WWW, pages 171–180.
19. Mimno, D., Wallach, H., Naradowsky, J., Smith, D. and McCallum, A. 2009. Polylingual topic models. In Proceedings of the 2009 Conference on Empiri-cal Methods in Natural Language Processing: Volume 2-Volume 2, pp. 880-889. Association for Computational Linguistics, 2009.
20. Mimno, D., Wallach, H., Talley, E., Leenders, M. and McCallum, A. 2011. Optimizing semantic coherence in topic models. In Proceedings of EMNLP, pages 262–272.
21. Moghaddam, S. and Ester, M. 2011. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In Proceedings of SIGIR, pages 665–674.
22. Mukherjee, A. and Liu, B. 2012. Aspect Extraction through Semi-Supervised Modeling. In Proceedings of ACL, pages 339–348.
23. Pang, B. and Lee, L. 2008. Opinion mining and sentiment analysis. Foundations and Trends in In-formation Retrieval, 2(1-2), 1–135.
24. Rosen-Zvi, M., Chemudugunta, C., Griffiths, T., Smyth, P. and Steyvers, M. 2010. Learning author-topic models from text corpora. ACM Trans. on Information Systems, 28(1), 1–38.
25. Teh, Y., Jordan, M., Beal, M. and Blei. D. 2006. Hierarchical Dirichlet Processes. In Journal of the American Statistical Association (JASA).
26. Titov, I. and Mcdonald, R. 2008. A joint model of text and aspect ratings for sentiment summarization. In Proceedings of ACL.
27. Wang, H., Lu, Y. and Zhai, C. 2010. Latent aspect rating analysis on review text data: a rating regression approach. In Proceedings of KDD, pages 783–792.
28. Zhang, D., Sun, J., Zhai, C., Bose, A. and Anerousis, N. 2010. PTM: Probabilistic topic mapping model for mining parallel document collections. In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1653-1656. ACM, 2010.
29. Zhao, W., Jiang, J., Yan, H. and Li, X. 2010. Jointly Modeling Aspects and Opinions with a MaxEnt-LDA Hybrid. In Proceedings of EMNLP, pages 56–65.