Extracting and Understanding Contrastive Opinion through Topic Relevant Sentences

Ebuka Ibeke and **Chenghua Lin** and **Adam Wyner** and **Mohamad Hardyman Barawi**
Department of Computing Science
University of Aberdeen, UK
{e.e.ibeke,chenghua.lin,azwyner,r01mhbb}@abdn.ac.uk

Abstract

Contrastive opinion mining is essential in identifying, extracting and organising opinions from user generated texts. Most existing studies separate input data into respective collections. In addition, the relationships between the topics extracted and the sentences in the corpus which express the topics are opaque, hindering our understanding of the opinions expressed in the corpus. We propose a novel unified latent variable model (contraLDA) which addresses the above matters. Experimental results show the effectiveness of our model in mining contrasted opinions, outperforming our baselines.

1 Introduction

Recent text mining applications have uncovered public opinions and social trends. This is partially driven by large corpora of opinionated documents in the web. Contrastive opinion mining is the discovery of opposing opinions and sentiments held by individuals or groups about a given topic. The usefulness of contrastive opinion mining spans across many applications such as discovering the public's stand on major socio-political events (Fang et al., 2012), observing heated debates over controversial issues (Lippi and Torroni, 2016), and product review sites (Lerman and McDonald, 2009). Considering the volume of reviews, it is highly desirable to acquire an overview of the major viewpoints from large amounts of text data automatically, allowing one to convert data into actionable knowledge for timely decision-making.

Recently, there have been some studies on mining contrastive viewpoints or opinions from text (Paul and Girju, 2009; Fang et al., 2012; Elahi and Monachesi, 2012; Gutiérrez et al., 2016). However, these studies assume that input data are separated into different collections beforehand, e.g., news articles from CNN vs. those from Fox News about the same set of events. While this assumption might hold for some practical scenarios, one quite often needs to analyse contrastive opinions contained in a single collection such as an open-ended discussion about government policy or commercial products in order to understand the viewpoints and their connections across the collection.

In addition, it is natural that debates on some topics are more prominent, indicating the importance of the topic. Therefore, being able to understand the prominence of a topic and the levels of contrastive sentiment will help one to prioritise actions. Finally, existing models generally interpret contrastive opinions solely in terms of the extracted topic words, which are not adequate to help us accurately understand the opinions presented in the corpus since the topic words only express shallow semantics. Understanding the dependency between the sentences in the corpus and the topic of discussion would be illuminating. The representative sentences also help to clarify the coherence of the extracted topics.

In this paper, we address the aforementioned issues by proposing a novel unified latent variable model (contraLDA) for mining contrastive opinion from text collections. The proposed model contributes the following: (1) automatically discovers contrastive opinion from both single and multiple text collections; (2) quantifies the strength of opinion contrastiveness towards the topic of interest, which could allow one to swiftly flag issues that require immediate attention; and (3) adopts the sentence extraction approach in (Barawi et al., 2017) to extract relevant sentences related to topics, making sentiment-bearing top-

ics clearer to users. Experimental results show that our model outperforms several baseline models in terms of extracting coherent and distinctive sentiment-bearing topics which express contrastive opinions. The topic relevant sentences extracted by our approach further help us effectively understand and interpret sentiment-bearing topics.
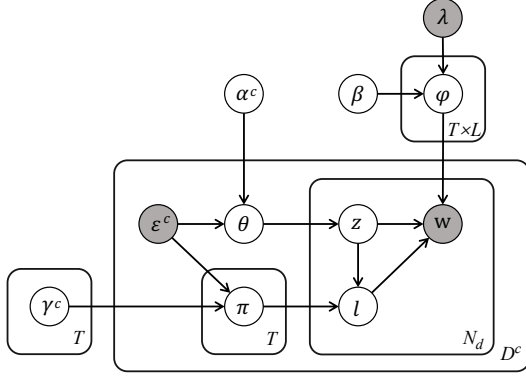
2 Methodology

Figure 1: The contraLDA Model.

We propose a model called contraLDA which offers a unified framework for mining contrastive opinions from text, where the source of text could be either a single collection or multiple collection of text. The graphical model of contraLDA is shown in Figure 1. Given a collection of documents $\boldsymbol{D}$, assume that $\boldsymbol{D}$ can be divided in to C classes: $\boldsymbol{D} = \{D^c\}_{c=1}^C$ with D^c documents per class, each document d in class c is a sequence of N_d words, each word in the document is an item from a vocabulary with V distinct terms, and c is the class index. Also assuming that L and T are the total number of sentiment labels and topics, respectively, the complete procedure for generating a word w_n in contraLDA is as follows: first, one draws a topic z from the class-constrained topic distribution $\boldsymbol{\theta}_d^c$. Following that, one draws a sentiment label l from the topic specific, class-constrained sentiment distribution $\boldsymbol{\pi}_{d,z}^c$. Finally, one draws a word from the per-corpus word distribution $\boldsymbol{\varphi}_{z,l}$ conditioned on both topic z and sentiment label l. Note that documents of all collections share the same $\boldsymbol{\varphi}$, and we can fully keep track of which collection a document belongs to based on its class index c. It is also important to note that the number of classes C plays a key role in controlling the operation mode of contraLDA. That is when $C = 1$, contraLDA is essentially modelling

a single collection of text without any class membership information. In the scenario where $C > 1$, contraLDA will be switching to model multiple collections of text, e.g., documents annotated with class labels, or articles from New York Times and Xinhua News about the same set of events. We summarise the generative process of contraLDA as follows:

- For each topic $z \in \{1, \cdots, T\}$
  - For each sentiment label $l \in \{1, \cdots, S\}$
    * Draw $\boldsymbol{\varphi}_{z,l} \sim \mathrm{Dir}(\boldsymbol{\beta}_{z,l})$.

- For each document $d \in \boldsymbol{D}$
  - choose a distribution $\boldsymbol{\theta}_d^c \sim \mathrm{Dir}(\boldsymbol{\epsilon}_z^c \cdot \boldsymbol{\alpha})$.
  - For each sentiment label l under document d,
    * Choose a distribution $\boldsymbol{\pi}_{d,z}^c \sim \mathrm{Dir}(\boldsymbol{\epsilon}_l^c \cdot \boldsymbol{\gamma})$.
  - For each word $n \in \{1, \cdots, N_d^c\}$ in document d
    * Choose a topic $z_n \sim \mathrm{Mult}(\boldsymbol{\theta}_d^c)$,
    * Choose a sentiment label $l_n \sim \mathrm{Mult}(\boldsymbol{\pi}_{d,z_n}^c)$,
    * Choose a word $w_n \sim \mathrm{Mult}(\boldsymbol{\varphi}_{z_n,l_n})$.

2.1 Incorporating Supervised Information.

The contraLDA model can be trained flexibly depending on the type of supervision information available. Specifically, if there are only labelled features available (e.g., sentiment lexicon, or topic seed words), our model will incorporate the labelled features to constrain the Dirichlet prior of topic-word distributions, which essentially plays a role in governing the model inference. If there is fully labelled data available, e.g., labelled documents, our model will account for the full supervision from document labels during the generative process, where each document can associate with a single class label or multiple class labels. However, if the dataset contains both labelled and unlabelled data, our model will account for the available labels during the generative process as well as incorporate the labelled features as above to constrain the Dirichlet prior.

When labelled data is available, contraLDA incorporates supervised information by constraining that a training document can only be generated from the topic set with class labels corresponding to the document's observed label set. This is achieved by introducing a dependency link from the document label matrix $\boldsymbol{\epsilon}$ to the Dirichlet priors $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$. Suppose a corpus has three topical labels denoted by $\boldsymbol{Z} = \{z_1, z_2, z_3\}$ and for each label z_k there are two sentiment labels denoted by $\boldsymbol{l} = \{l_1, l_2\}$. Given observed label matrix $\boldsymbol{\epsilon}_c = \{\boldsymbol{\epsilon}_z^c, \boldsymbol{\epsilon}_l^c\} = \{(1,0,1),(1,0)\}$ which indicates

that d is associated with topic labels z_1, z_3 as well as sentiment label l_1, we can encode the label information into contraLDA as

$$\boldsymbol{\alpha}_d^c = \boldsymbol{\epsilon}_z^c \cdot \boldsymbol{\alpha} \qquad (1)$$

$$\boldsymbol{\gamma}_d^c = \boldsymbol{\epsilon}_l^c \cdot \boldsymbol{\gamma} \qquad (2)$$

This ensures that d can only be generated from topics associated with observed class labels from $\boldsymbol{\epsilon}$. If there are no labelled documents available, contraLDA will incorporate labelled features from $\boldsymbol{\lambda}$ (e.g., sentiment lexicons) for constraining the Dirichlet priors $\boldsymbol{\beta}$ using the same strategy described in (Lin and He, 2009; Lin et al., 2012a).

2.2 Inference.

From the contraLDA graphical model depicted in Figure 1, we can write the joint distribution of all observed and hidden variables which can be factored into three terms:

$$P(\mathbf{w}, \mathbf{z}, \mathbf{l}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c}) =$$
$$P(\mathbf{w}|\mathbf{z}, \mathbf{l}, \boldsymbol{\beta})P(\mathbf{l}|\mathbf{z}, \boldsymbol{\gamma}, \boldsymbol{c})P(\mathbf{z}|\boldsymbol{\alpha}, \boldsymbol{c}) \qquad (3)$$

The main objective of inference in contraLDA is then to find a set of model parameters that can best explain the observed data, namely, the class-constrained topic proportion $\boldsymbol{\theta}^c$, the class-constrained topic label specific sentiment proportion $\boldsymbol{\pi}^c$, and the per-corpus word distribution $\boldsymbol{\varphi}$. To compute these target distributions, we need to calculate the posterior distribution of the model. As the posterior is intractable, we use a collapsed Gibbs sampler to approximate the posterior based on the full conditional distribution for each word token in position t. By evaluating the model joint distribution in Eq. 3, we can yield the full conditional distribution as follows

$$P(z_t = k, l_t = j|\mathbf{w}, \mathbf{z}^{-t}, \mathbf{l}^{-t}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c}) \propto$$
$$\frac{N_{k,j,w_t}^{-t} + \beta_{k,j,i}}{N_{k,j}^{-t} + \sum_i \beta_{k,j,i}} \cdot \frac{N_{d,k}^{-t} + \alpha_{d,k}^c}{N_d^{-t} + \sum_k \alpha_{d,k}^c}$$
$$\cdot \frac{N_{d,k,j}^{-t} + \gamma_{d,k,j}^c}{N_{d,k}^{-t} + \sum_j \gamma_{d,k,j}^c}. \qquad (4)$$

where the superscript $-t$ denotes a quantity that excludes data from tth position, $N_{k,j,w}$ is the number of times word w appeared in topic k with sentiment label j, $N_{k,j}$ is the number of times words are assigned to topic k and sentiment label j, $N_{d,k}$

is the number of times topic k is assigned to some word tokens in document d, N_d is the total number of words in document d, $N_{d,k,j}$ is the number of times a word from document d is associated with topic k and sentiment label j.

Using Eq. 4, we can obtain sampling assignments for contraLDA model, based on which model parameters can be estimated as
$$\varphi_{k,j,i} = \frac{N_{k,j,i} + \beta_{k,j,i}}{N_{k,j} + \sum_i \beta_{k,j,i}}, \quad \theta_{d,k,j}^c = \frac{N_{d,k} + \alpha_{k,j}^c}{N_d + \sum_k \alpha_{d,k}^c}$$
and $\qquad \pi_{d,k}^c = \frac{N_{d,k,j} + \gamma_{d,k,j}^c}{N_{d,k} + \sum_j \gamma_{d,k,j}^c}$.

2.3 Modelling the associations between sentiment-bearing topics and sentences.

Our model adopts a computational mechanism (Barawi et al., 2017) that can uncover the association between an opinionated (or sentiment-bearing) topic and the underlying sentences of a corpus. First, we preserve the sentential structure of each document during the corpus preprocessing step (see §3 for more details). Second, modelling topic-sentence relevance is essentially equivalent to calculating the probability of a sentence given a sentiment-bearing topic $p(\text{sent}|z, l)$. The posterior inference of our model, based on Gibbs sampling, can recover the hidden sentiment label and topic label assignments for each word in the corpus. Such label-word assignment information provides a means for re-assembling the relevance between a word and a sentiment-bearing topic. By leveraging the sentential structure information and gathering the label assignment statistics for each word of a sentence, we can derive the probability of a sentence given a sentiment-bearing topic as

$$p(\text{sent}|z, l) = \frac{p(z, l|\text{sent}) \cdot p(\text{sent})}{p(z, l)}$$
$$\propto p(z, l|\text{sent}) \cdot p(\text{sent}) \qquad (5)$$

where

$$p(z, l|\text{sent}) = \frac{\sum_{w, z', l'} \varphi_{z', l', w}}{\sum_{w \in \text{sent}} \varphi_{z', l', w}}, \qquad (6)$$

$$p(\text{sent}) = \sum_z \sum_l \prod_{w \in \text{sent}} \varphi_{z, l, w}. \qquad (7)$$

Also $p(l, z)$ is discounted as it is a constant when comparing sentential labels for the same sentiment-bearing topic. The extracted sentences for each sentiment-bearing topic are ranked based on their probability scores.

3 Experimental Setup

Dataset. We evaluate the performance of our model[1] for contrastive opinion mining on the El Capitan dataset[2] (Ibeke et al., 2016) which consists of reviews manually annotated (with 18 topic labels and 3 sentiment labels in total) for various opinion mining tasks. The dataset consists of 2,232 customer reviews, with topic and sentiment annotations at both the review and sentence levels. For the sentiment labels, we only concentrate on positive and negative sentiment labels with the 2.3% of neutral reviews being ignored, since the aim of this study is to mine contrastive opinion from text. The dataset has 10,348 sentences with an average length of 17.3 words.

Preprocessing. We preprocessed the experimental dataset by first performing automatic sentence segmentation[3] in order to preserve the sentential structure information of each document. We then remove punctuation, numbers, non-alphabet characters, stop words, lowercase all words, and perform stemming.

4 Experimental Results

Topic coherence. We first quantitatively measure the coherence of the extracted topics by our model and compare the results with a number of baselines, namely, LDA (Blei et al., 2003), ccLDA (Paul and Girju, 2009), TAM (Paul and Girju, 2010), and VODUM (Thonet et al., 2016). We employ normalised pointwise mutual information (NPMI) (Bouma, 2009) which outperforms other metrics in measuring topic coherence (Newman et al., 2010; Aletras and Stevenson, 2013). We run our model and the baseline models with two sentiment labels (i.e., positive and negative), and vary the topic number setting $T \in \{5, 10, 20, 30, 40, 50\}$, setting $\beta = 0.01$ (Steyvers and Griffiths, 2007) and $\alpha = 0.1$. Our model learns α directly from data using maximum-likelihood estimation (Lin et al., 2012b).

As can be seen from Figure 2a, there is a general pattern for all tested models, where the coherence score of the extracted topics decreases as a larger number of topics K being modelled. This is inline with the observations of (Mimno et al.,

2011; Gutiérrez et al., 2016), who discovered that as the number of topics increases, lower-likelihood topics tend to be more incoherent, resulting in lower coherence score for topics. In terms of individual models, our model consistently achieves a higher coherent score than all baseline models. For instance, when compared with the best baseline VODUM, our model gives over 8% averaged improvement. This demonstrates the capability of the proposed contraLDA in extracting coherent and meaningful topics.

Analysis of opinion contrastiveness. We further study the problem of quantifying the strength of opinion contrastiveness towards the topic of interest, which allows one to swiftly flag topics or issues that require immediate attention. We approach this by computing the prominence score for each sentiment-bearing topic extracted by contraLDA given a corpus $\boldsymbol{c}$ using

$$P(z, l|\boldsymbol{c}) = \frac{1}{|D|} \sum_{d=1}^{D} P(l|z, d)P(z|d)$$

$$= \frac{1}{|D|} \sum_{d=1}^{D} \theta_{d,z} \cdot \pi_{d,z,l}, \qquad (8)$$

where D is the total number of documents in the corpus. Thus the prominence for topic z in a corpus can be derived as

$$P(z) = \sum_{l} P(z, l). \qquad (9)$$

Figure 2b shows some contrastive opinion topic pairs ordered by their prominence in the corpus. Modelling topic prominence and sentiment contrastiveness provides a quick overview of the notable topics and the sentiments towards them. We can easily identify that the most heated topics are `update` and `performance`. In terms of opinion contrastiveness, we see that `Speed` received quite balanced positive and negative sentiment magnitude. `Performance` and `Update` are skewed towards the negative sentiment, indicating that a majority of customers experienced a performance drop after upgrading to El Capitan.

Contrastive opinion topic analysis. In this experiment, we qualitatively evaluate our model in the task of discovering contrastive opinions.

The top panel of Table 1 shows contrastive opinion topic pairs extracted by our model. Note that `Performance`, `Office` and `Yosemite` are label information from the El Capitan

---

[1] While our model can be applied to both single and multiple data collections, due to page limits, we only show the experimental results on a single dataset.

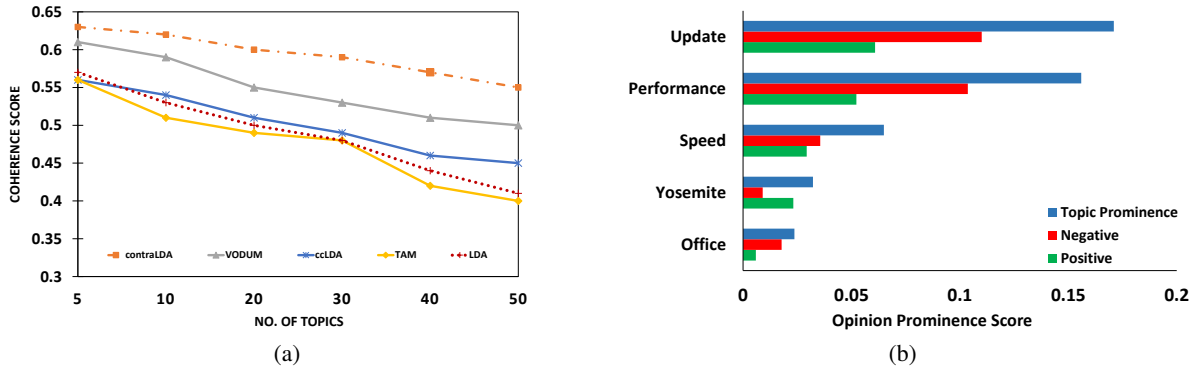[2] https://github.com/eibeke/El-Capitan-Dataset

[3] http://www.nltk.org/

Figure 2: Topic coherence analysis using NPMI (a); Analysis of topic prominence and sentiment contrastiveness (b). NB: blue bar indicates the overall prominence of contrastive topic pair; green bar indicates the strength of a positive sentiment topic, and red bar for negative sentiment topic.

Performance		Office		Yosemite	
+	-	+	-	+	-
work	crash	offic	offic	yosemit	yosemit
run	work	microsoft	use	work	upgrad
perform	time	compat	work	time	destroy
faster	app	quick	microsoft	downgrade	slow
app	use	fine	ms	restor	work
smooth	slow	work	crash	issu	mac
new	mac	updat	issu	instal	bad
pro	open	upgrad	word	machin	problem
macbook	freez	new	excel	macbook	maverick
better	just	didn	appl	revert	appl

Performance +	So much better than before, and apps run faster too.
Performance -	Computer slows down dramatically, programs freeze.
Office +	Office 2016 opens quickly with no issues.
Office -	Update:Office apps tend to crash after the update!
Yosemite +	So I downgraded back to Yosemite and - hey presto!
Yosemite -	My 2010 iMac was destroyed by Yosemite.

Table 1: Contrastive opinion topic examples and the top rated sentence for each topic.

dataset. A topic pair, e.g., (Performance+, Performance−), expresses contrastive opinions towards the same topic Performance, with '+' and '-' indicating the topic sentiment orientation. For instance, the two topics under Performance+ suggests that some people feel the system *performs better* and *app runs faster*, whereas Performance− seems to show highly contrastive opinion that people have bad experience after upgrade, e.g., *app crashes* or *freezes*, *mac* becomes *slow*. However, it is still impossible to accurately interpret the extracted topics solely based on its multinomial distribution, especially when one is unfamiliar with the topic domain. We bridge this gap by extracting the most relevant sentences for a given topic, which can greatly facilitate sentiment-bearing topic interpretation (as described in § 2.3).

The bottom panel of Table 1 shows the extracted top sentences (ranked based on Eq. 5) for each topic. For instance, the extracted top sentences for the Office topic show that some customers recorded an improvement with their office app (e.g., "*Office* 2016 opens *quickly* with no issues"), while others are unhappy with the office app (e.g., "Update: *Office* apps tend to *crash* after the update"). We see that the top sentences can effectively bridge the gap between the topic word distributions and the opinion encoded within the topic, and hence can greatly help facilitate sentiment-bearing topic understanding and interpretation.

5 Conclusion

We presented the contraLDA model which detects contrastive opinions both in single and multiple data collections, and determines the sentiments of the extracted opinions. Our model effectively mines coherent topics and contrastive opinions from text. Experimental results show that our model outperforms baselines in extracting coherent topics. In addition, we presented a mechanism for extracting sentences from corpus that are relevant to sentiment-bearing topics, which helps understanding and interpretation of the topics discovered. We plan to further investigate our approach on datasets from more domains.

Acknowledgments

References

Nikolaos Aletras and Mark Stevenson. 2013. Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS)*. pages 13–22.

Mohamad H. Barawi, Chenghua Lin, and Advaith Siddharthan. 2017. Automatically labelling sentiment-bearing topics with descriptive sentence labels. In *In Proceedings of the 22nd International Conference on Natural Language and Information Systems (NLDB)*. pages 299–312.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3:993–1022.

Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *In The German Society for Computational Linguistics and Language Technology (GSCL)* pages 31–40.

Mohammad F. Elahi and Paola Monachesi. 2012. An examination of cross-cultural similarities and differences from social media data with respect to language use. In *The Emotion and Sentiment Analysis Workshop in the 10th Language Resources and Evaluation Conference (LREC)*. pages 4080–4086.

Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. 2012. Mining contrastive opinions on political texts using cross-perspective topic model. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM)*. pages 63–72.

ED Gutiérrez, Ekaterina Shutova, Patricia Lichtenstein, Gerard de Melo, and Luca Gilardi. 2016. Detecting cross-cultural differences using a multilingual topic model. *Transactions of the Association for Computational Linguistics (TACL)* 4:47–60.

Ebuka Ibeke, Chenghua Lin, Chris Coe, Adam Wyner, Dong Liu, Mohamad H. Barawi, and Noor F.A. Yusof. 2016. A curated corpus for sentiment-topic analysis. In *The Emotion and Sentiment Analysis Workshop in the 10th Language Resources and Evaluation Conference (LREC), Slovenia*.

Kevin Lerman and Ryan McDonald. 2009. Contrastive summarization: an experiment with consumer reviews. In *Proceedings of human language technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (ACL)*. pages 113–116.

Chenghua Lin and Yulan He. 2009. Joint sentiment/topic model for sentiment analysis. In *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, pages 375–384.

Chenghua Lin, Yulan He, Richard Everson, and Stefan Ruger. 2012a. Weakly supervised joint sentiment-topic detection from text. *IEEE Transactions on Knowledge and Data Engineering* 24(6):1134–1145.

Chenghua Lin, Yulan He, Carlos Pedrinaci, and John Domingue. 2012b. Feature lda: a supervised topic model for automatic detection of web api documentations from the web. In *International Semantic Web Conference (ISWC)*. Springer, pages 328–343.

Marco Lippi and Paolo Torroni. 2016. Argument mining from speech: Detecting claims in political debates. In *In Association for the Advancement of Artificial Intelligence (AAAI)*. pages 2979–2985.

David Mimno, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pages 262–272.

David Newman, Jey H. Lau, Karl Grieser, and Timothy Baldwin. 2010. Automatic evaluation of topic coherence. In *Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (ACL)*. pages 100–108.

Michael Paul and Roxana Girju. 2009. Cross-cultural analysis of blogs and forums with mixed-collection topic models. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*. pages 1408–1417.

Michael Paul and Roxana Girju. 2010. A two-dimensional topic-aspect model for discovering multi-faceted topics. *In Association for the Advancement of Artificial Intelligence (AAAI)* .

Mark Steyvers and Tom Griffiths. 2007. Probabilistic topic models pages 424–440.

Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat. 2016. Vodum: a topic model unifying viewpoint, topic and opinion discovery. In *European Conference on Information Retrieval*. Springer, pages 533–545.