VODUM: a Topic Model Unifying Viewpoint, Topic and Opinion Discovery

Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, Karen Pinel-Sauvagnat

VODUM: a Topic Model Unifying Viewpoint, Topic and Opinion Discovery

Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, and Karen Pinel-Sauvagnat

Université Paul Sabatier, IRIT, 118 Route de Narbonne, F-31062 Toulouse CEDEX 9, France

{thonet,cabanac,boughanem,sauvagnat}@irit.fr

Abstract. The surge of opinionated on-line texts provides a wealth of information that can be exploited to analyze users' viewpoints and opinions on various topics. This article presents VODUM, an unsupervised Topic Model designed to jointly discover viewpoints, topics, and opinions in text. We hypothesize that partitioning topical words and viewpoint-specific opinion words using part-of-speech helps to discriminate and identify viewpoints. Quantitative and qualitative experiments on the Bitterlemons collection show the performance of our model. It outperforms state-of-the-art baselines in generalizing data and identifying viewpoints. This result stresses how important topical and opinion words separation is, and how it impacts the accuracy of viewpoint identification.

1 Introduction

The surge of opinionated on-line texts raised the interest of researchers and the general public alike as an incredibly rich source of data to analyze contrastive views on a wide range of issues, such as policy or commercial products. This large volume of opinionated data can be explored through text mining techniques, known as Opinion Mining or Sentiment Analysis. In an opinionated document, a user expresses her **opinions** on one or several **topics**, according to her **viewpoint**. We define the key concepts of topic, viewpoint, and opinion as follows. A topic is one of the subjects discussed in a document collection. A viewpoint is the standpoint of one or several authors on a set of topics. An opinion is a wording that is specific to a topic and a viewpoint. For example, in the manually crafted sentence *Israel occupied the Palestinian territories of the Gaza strip*, the topic is *the presence of Israel on the Gaza strip*, the viewpoint is *pro-Palestine* and an opinion is *occupied*. Indeed, when mentioning the action of building Israeli communities on disputed lands, the pro-Palestine side is likely to use the verb *to occupy*, whereas the pro-Israel side is likely to use the verb *to settle*. Both sides discuss the same topic, but they use a different wording that conveys an opinion.

The contribution of this article is threefold:

1. We first define the task of **Viewpoint and Opinion Discovery**, which consists in analyzing a collection of documents to identify the viewpoint of

each document, the topics mentioned in each document, and the viewpoint-specific opinions for each topic.
2. To tackle this issue, we propose the *Viewpoint and Opinion Discovery Unification Model* (VODUM), an unsupervised approach to jointly model viewpoints, topics, and opinions.
3. Finally, we quantitatively and qualitatively evaluate our model VODUM on the Bitterlemons collection, benchmarking it against state-of-the-art baselines and degenerate versions of our model to analyze the usefulness of VODUM's specific properties.

The remainder of this paper is organized as follows. Section 2 presents related work and state-of-the-art Viewpoint and Opinion Discovery Topic Models. Our model's properties and inference process are described in Section 3. Section 4 details the experiments performed to evaluate VODUM. We conclude and give future directions for this work in Section 5.

2 Related Work: Viewpoint and Opinion Discovery Topic Models

Viewpoint and Opinion Discovery is a sub-task of Opinion Mining, which aims to analyze opinionated documents and infer properties such as subjectivity or polarity. We refer the reader to [6] for a general review of this broad research topic. While most Opinion Mining works first focused on product reviews, more recently, a surge of interest for sociopolitical and debate data led researchers to study tasks such as Viewpoint and Opinion Discovery. The works described in this section relate to LDA [2] and more generally to probabilistic Topic Models, as a way to model diverse latent variables such as viewpoints, topics, and opinions.

Several works modeled viewpoint-specific opinions [7,3] but they did not learn the viewpoint assignments of documents. Instead, they assumed these assignments to be known beforehand and leveraged them as prior information fed to their models. Some authors proposed a Topic Model to analyze culture-specific viewpoints and their associated wording on common topics [7]. In [3], the authors jointly modeled topics and viewpoint-specific opinions. They distinguished between topical words and opinion words based on part-of-speech: nouns were assumed to be topical words; adjectives, verbs, and adverbs were assumed to be opinion words.

Other works discovered document-level viewpoints in a supervised or semi-supervised fashion [4,8,12]. In [4], document-level and sentence-level viewpoints were detected using a supervised Naive Bayes approach. In [8], the authors defined the Topic-Aspect Model (TAM) that jointly models topics, and aspects, which play the role of viewpoints. Similarly, the Joint Topic Viewpoint (JTV) model was proposed in [12] to jointly model topics and viewpoints. However, both TAM and JTV inferred parameters were only integrated as features into a SVM classifier to identify document-level viewpoints. TAM was extended to perform contrastive viewpoint summarization [9], but this extension was still weakly supervised as it leveraged a sentiment lexicon to identify viewpoints.

Ref.	Model is used without supervision	Topical words and opinion words are partitioned	Viewpoint assignments are learned	Model is independent of structure-specific properties
[7]	+	−	−	+
[3]	+	+	−	+
[4,8,12,9]	−	−	+	+
[10,11]	+	−	+	−
VODUM	+	+	+	+

Table 1: Comparison of our model VODUM against related work approaches.

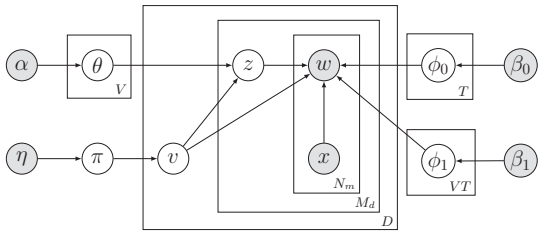

Fig. 1: Graphical model of VODUM.

The task of viewpoint identification was also studied for user generated data such as forums, where users can debate on controversial issues [10,11]. These works proposed Topic Models that however rely on structure-specific properties exclusive to forums (such as threads, posts, users, interactions between users), which cannot be applied to infer general documents' viewpoint.

The specific properties of VODUM compared to related work are summarized in Table 1. VODUM is totally unsupervised. It separately models topical words and opinion words. Document-level viewpoint assignments are learned. VODUM is also structure-independent and thus broadly applicable. These properties are further detailed in Section 3.

3 Viewpoint and Opinion Discovery Unification Model

3.1 Description

VODUM is a probabilistic Topic Model based on LDA [2]. VODUM simultaneously models viewpoints, topics, and opinions, i.e., it identifies topical words and viewpoint-specific topic-dependent opinion words. The graphical model of VODUM and the notation used in this article are provided in Fig. 1 and Table 2, respectively. The specific properties of VODUM are further detailed below.

Topical Words and Opinion Words Separation. In our model, topical words and opinion words are partitioned based on their part-of-speech, in line with several viewpoint modeling and Opinion Mining works [3,13,5]. Here, nouns are assumed to be topical words; adjectives, verbs and adverbs are assumed to be opinion words. While this assumption seems coarse, let us stress that a more accurate definition of topical and opinion words (e.g., by leveraging sentiment lexicons) could be used, without requiring any modification of our model. The

D, M_d, N_m	Number of documents in the collection, number of sentences in document d and number of words in sentence m, respectively
W	Number of words in the vocabulary
W_0, W_1	Number of topical words and opinion words in the vocabulary, respectively
T, V	Number of topics and viewpoints, respectively
$\mathcal{W}_0$, $\mathcal{W}_1$	Set of topical words and opinion words in the vocabulary, respectively
$w_{d,m,n}$	The n-th word of the m-th sentence from the d-th document
$x_{d,m,n}$	The part-of-speech category (0 or 1) of $w_{d,m,n}$
$z_{d,m}$	The topic assigned to the m-th sentence of the d-th document
v_d	The viewpoint assigned to the d-th document
$\boldsymbol{w}$, $\boldsymbol{x}$, $\boldsymbol{z}$, $\boldsymbol{v}$	Vector of all words, part-of-speech categories, topic assignments and viewpoint assignments, respectively
ϕ_0	$T \times W$ matrix of viewpoint-independent distributions over topical words
ϕ_1	$V \times T \times W$ matrix of viewpoint-dependent distributions over opinion words
θ	$V \times T$ matrix of viewpoint-dependent distributions over topics
π	V matrix of the distribution over viewpoints
β_0, β_1, α, η	Symmetric Dirichlet prior for ϕ_0, ϕ_1, θ and π, respectively
$n^{(i)}$	Number of documents in the collection assigned to viewpoint i
$n_i^{(j)}$	Number of sentences in the collection assigned to viewpoint i and topic j
$n_{0,j}^{(k)}$	Number of instances of topical word k assigned to topic j
$n_{1,i,j}^{(k)}$	Number of instances of opinion word k assigned to viewpoint i and topic j

Table 2: Notation for our model VODUM.

part-of-speech tagging pre-processing step is further described in Section 4.2. The part-of-speech category is represented as an observed variable x which takes a value of 0 for topical words and 1 for opinion words. Topical words and opinion words are then drawn from distributions ϕ_0 and ϕ_1, respectively.

Sentence-level Topic Variables. Most Topic Models define word-level topic variables (e.g., [2,8,3]). We hypothesize that using sentence-level topic variables, denoted by z, better captures the dependency between the opinions expressed in a sentence and the topic of the sentence. Indeed, coercing all words from a sentence to be related to the same topic reinforces the co-occurrence property leveraged by Topic Models. As a result, the topics induced by topical word distributions ϕ_0 and opinion word distributions ϕ_1 are more likely to be aligned.

Document-level Viewpoint Variables. Viewpoint variables v are defined at the document level and drawn from the distribution π. In previous works, viewpoint variables were allocated to words [8,9] or authors [10,11]. While it is reasonable to suppose that an author writes all her documents with the same viewpoint, the authorship information is not always available. On the other hand, allocating each word of a document to a potentially different viewpoint is meaningless. We thus modeled document-level viewpoint variables.

Viewpoint-specific Topic Distributions. In VODUM, topic distributions, denoted by θ in the graphical model, are viewpoint-specific instead of being document-specific as in other Topic Models [2,8,12]. This assumption comes from the observation in [10] that different viewpoints have different dominating topics. For example, opponents of same-sex marriage are more likely to mention religion than the supporting side.

1. Draw a viewpoint distribution π from Dirichlet(η).
2. Draw a viewpoint-independent topical word distribution $\phi_{0,j}$ from Dirichlet(β_0) for all topics $j \in [\![1, T]\!]$.
3. Draw a viewpoint-dependent opinion word distribution $\phi_{1,i,j}$ from Dirichlet(β_1) for all viewpoints $i \in [\![1, V]\!]$ and all topics $j \in [\![1, T]\!]$.
4. Draw a topic distribution θ_i from Dirichlet(α) for all viewpoints $i \in [\![1, V]\!]$.
5. For each document $d \in [\![1, D]\!]$
   (a) Draw a viewpoint v_d from Multinomial(π).
   (b) For each sentence $m \in [\![1, M_d]\!]$
      i. Draw a topic $z_{d,m}$ from Multinomial(θ_{v_d}).
      ii. For each word $n \in [\![1, N_m]\!]$
         A. Choose a part-of-speech category $x_{d,m,n}$ from $\{0, 1\}$, where category 0 denotes topical words (nouns) and category 1 denotes opinion words (adjectives, verbs and adverbs).
         B. If $x_{d,m,n} = 0$, draw a topical word $w_{d,m,n}$ from Multinomial($\phi_{0,z_{d,m}}$), else if $x_{d,m,n} = 1$, draw an opinion word $w_{d,m,n}$ from Multinomial($\phi_{1,v_d,z_{d,m}}$).

Fig. 2: Generative process for a collection as modeled by VODUM.

Similarly to LDA and other probabilistic Topic Models, the probability distributions ϕ_0, ϕ_1, θ, and π are Multinomial distributions with symmetric Dirichlet priors β_0, β_1, α, and η, respectively.

The virtual generation of a document as modeled by VODUM is the following. The author writes the document according to her own viewpoint. Depending on her viewpoint, she selects for each sentence of the document a topic that she will discuss. Then, for each sentence, she chooses a set of topical words to describe the topic that she selected for the sentence, and a set of opinion words to express her viewpoint on this topic. Formally, the generative process of a document collection is performed as described in Fig. 2. In Section 3.2, we detail how we infer parameters ϕ_0, ϕ_1, θ, and π.

3.2 Model Inference

As for other probabilistic Topic Models, the exact inference of VODUM is not tractable. We thus rely on approximate inference to compute parameters ϕ_0, ϕ_1, θ, and π, as well as the document-level viewpoint assignments v. We chose collapsed Gibbs sampling as it was shown to be quicker to converge than approximate inference methods such as variational Bayes [1].

Collapsed Gibbs Sampling. Collapsed Gibbs sampling is a Markov chain Monte Carlo algorithm that generates a set of samples drawn from a posterior probability distribution, i.e., the probability distribution of latent variables (v and z in our model) given observed variables (w and x in our model). It does not require the actual computation of the posterior probability, which is usually intractable for Topic Models. Only the marginal probability distributions of latent variables (i.e., the probability distribution of one latent variable given all other latent variables and all observed variables) need to be computed in order to perform collapsed Gibbs sampling. For each sample, the collapsed

Gibbs sampler iteratively draws assignments for all latent variables using their marginal probability distributions, conditioned on the previous sample's assignments. The marginal probability distributions used to sample the topic assignments and viewpoint assignments in our collapsed Gibbs sampler are described in (1) and (2), respectively. The derivation is omitted due to space limitation. The notation used in the equations is defined in Table 2. Additionally, indexes or superscripts $-d$ and $-(d,m)$ exclude the d-th document and the m-th sentence of the d-th document, respectively. Similarly, indexes or superscripts d and (d,m) include only the d-th document and the m-th sentence of the d-th document, respectively. A superscript $(\cdot)$ denotes a summation over the corresponding superscripted index.

$$p(z_{d,m}=j|v_d=i,\boldsymbol{v}_{-d},\boldsymbol{z}_{-(d,m)},\boldsymbol{w},\boldsymbol{x}) \propto \frac{n_i^{(j),-(d,m)}+\alpha}{n_i^{(\cdot),-(d,m)}+T\alpha} \times \frac{\prod_{k\in\mathcal{W}_0}\prod_{a=0}^{n_{0,j}^{(k),(d,m)}-1} n_{0,j}^{(k),-(d,m)}+\beta_0+a}{\prod_{b=0}^{n_{0,j}^{(\cdot),(d,m)}-1} n_{0,j}^{(\cdot),-(d,m)}+W_0\beta_0+b} \times \frac{\prod_{k\in\mathcal{W}_1}\prod_{a=0}^{n_{1,i,j}^{(k),(d,m)}-1} n_{1,i,j}^{(k),-(d,m)}+\beta_1+a}{\prod_{b=0}^{n_{1,i,j}^{(\cdot),(d,m)}-1} n_{1,i,j}^{(\cdot),-(d,m)}+W_1\beta_1+b} \tag{1}$$

$$p(v_d=i|\boldsymbol{v}_{-d},\boldsymbol{z},\boldsymbol{w},\boldsymbol{x}) \propto \frac{n^{(i),-d}+\eta}{n^{(\cdot),-d}+V\eta} \times \frac{\prod_{j=1}^{T}\prod_{a=0}^{n_i^{(j),d}-1} n_i^{(j),-d}+\alpha+a}{\prod_{b=0}^{n_i^{(\cdot),d}-1} n_i^{(\cdot),-d}+T\alpha+b} \times \prod_{j=1}^{T} \frac{\prod_{k\in\mathcal{W}_1}\prod_{a=0}^{n_{1,i,j}^{(k),d}-1} n_{1,i,j}^{(k),-d}+\beta_1+a}{\prod_{b=0}^{n_{1,i,j}^{(\cdot),d}-1} n_{1,i,j}^{(\cdot),-d}+W_1\beta_1+b} \tag{2}$$

Parameter Estimation. The alternate sampling of topics and viewpoints using (1) and (2) makes the collapsed Gibbs sampler converge towards the posterior probability distribution. The count variables $n^{(i)}$, $n_i^{(j)}$, $n_{0,j}^{(k)}$ and $n_{1,i,j}^{(k)}$ computed for each sample generated by the collapsed Gibbs sampler are used to estimate distributions π, θ, ϕ_0 and ϕ_1 as described in (3), (4), (5) and (6), respectively.

$$\pi^{(i)} = \frac{n^{(i)}+\eta}{n^{(\cdot)}+V\eta} \tag{3}$$

$$\theta_i^{(j)} = \frac{n_i^{(j)}+\alpha}{n_i^{(\cdot)}+T\alpha} \tag{4}$$

$$\phi_{0,j}^{(k)} = \begin{cases} \frac{n_{0,j}^{(k)}+\beta_0}{n_{0,j}^{(\cdot)}+W_0\beta_0} & \text{if } k\in\mathcal{W}_0 \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

$$\phi_{1,i,j}^{(k)} = \begin{cases} \frac{n_{1,i,j}^{(k)}+\beta_1}{n_{1,i,j}^{(\cdot)}+W_1\beta_1} & \text{if } k\in\mathcal{W}_1 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

4 Experiments

We investigated the following hypotheses in our experiments:

- **(H1)** Using viewpoint-specific topic distributions (instead of document-level topic distributions, e.g., as in TAM, JTV, and LDA) has a positive impact on the ability of the model to identify viewpoints.
- **(H2)** The separation between opinion words and topical words has a positive impact on the ability of the model to identify viewpoints.
- **(H3)** Using sentence-level topic variables improves the ability of the model to identify viewpoints.
- **(H4)** Using document-level viewpoint variables helps the model to identify viewpoints.
- **(H5)** VODUM outperforms state-of-the-art models (e.g., TAM, JTV, and LDA) in the modeling and viewpoint identification tasks.

Note that an issue similar to (H1) was already addressed in [10,11]. The authors did not evaluate, however, the impact of this assumption on the viewpoint identification task. The rest of this section is organised as follows. In Section 4.1, we detail the baselines we compared VODUM against. Section 4.2 describes the dataset used for the evaluation and the experimental setup. In Section 4.3, we report and discuss the results of the evaluation.

4.1 Baselines

We compared VODUM against state-of-the-art models and degenerate versions of our model in order to answer the research questions underlying our five hypotheses. The state-of-the-art models we considered are TAM [8,9], JTV [12] and LDA [2]. These are used to investigate (H5). The four degenerate versions of VODUM are defined to evaluate the impact of each of our model's properties in isolation. The degenerate versions and their purpose are detailed below.

In VODUM-D, topic distributions are defined at the document level. In VODUM, topic distributions are instead viewpoint-specific and independent of documents. VODUM-D has been defined to study (H1).

VODUM-O assumes that all words are opinion words, i.e., all words are drawn from distributions that depend both on viewpoint and topic. On the contrary, VODUM distinguishes opinion words (drawn from viewpoint-specific topic-dependent distributions) from topical words (drawn from topic-dependent distributions). Comparing VODUM against VODUM-O answers (H2).

VODUM-W defines topic variables at the word level, instead of sentence level as in VODUM. This essentially allows a document to be potentially associated with more topics (one topic per word as opposed to one per sentence), and loosens the link between opinion words and topical words. VODUM-W has been defined to tackle (H3).

VODUM-S models sentence-level viewpoint variables, while, in VODUM, viewpoint variables are defined at the document level. Therefore, a document modeled by VODUM-S can contain sentences assigned to different viewpoints. Comparing VODUM against VODUM-S addresses (H4).

4.2 Dataset and Experimental Setup

We evaluated our model on a collection of articles published in the Bitterlemons e-zine.[1] It contains essays written by Israeli and Palestinian authors, discussing the Israeli-Palestinian conflict and related issues. It was first introduced in [4] and then used in numerous works that aim to identify and model viewpoints in text (e.g., [9,12]). This collection contains 297 essays written by Israeli authors and 297 written by Palestinian authors. Before using the collection, we performed the following pre-processing steps using the Lingpipe[2] Java library. We first filtered out tokens that contain numerical characters. We then applied stop word removal and Porter stemming to the collection. We also performed part-of-speech tagging and annotated data with the binary part-of-speech categories that we defined in Section 3.1. Category 0 corresponds to topical words and contains common nouns and proper nouns. Category 1 corresponds to opinion words and contains adjectives, verbs, adverbs, qualifiers, modals, and prepositions. Tokens labeled with other part-of-speech were filtered out.

We implemented our model VODUM and the baselines based on the JGibbLDA[3] Java implementation of collapsed Gibbs sampling for LDA. The source code of our implementation and the formatted data (after all pre-processing steps) are available at `https://github.com/tthonet/VODUM`.

In the experiments, we set the hyperparameters of VODUM and baselines to the following values. The hyperparameters in VODUM were manually tuned: $\alpha = 0.01$, $\beta_0 = \beta_1 = 0.01$, and $\eta = 100$. The rationale behind the small α (θ's hyperparameter) and the large η (π's hyperparameter) is that we want a sparse θ distribution (i.e., each viewpoint has a distinct topic distribution) and a smoothed π distribution (i.e., a document has equal chance to be generated under each of the viewpoints). We chose the same hyperparameters for the degenerate versions of VODUM. The hyperparameters of TAM were set according to [9]: $\alpha = 0.1$, $\beta = 0.1$, $\delta_0 = 80.0$, $\delta_1 = 20.0$, $\gamma_0 = \gamma_1 = 5.0$, $\omega = 0.01$. For JTV, we used the hyperparameters' values described in [12]: $\alpha = 0.01$, $\beta = 0.01$, $\gamma = 25$. We manually adjusted the hyperparameters of LDA to $\alpha = 0.5$ and $\beta = 0.01$. For all experiments, we set the number of viewpoints (for VODUM, VODUM-D, VODUM-O, VODUM-W, VODUM-S, and JTV) and the number of aspects (for TAM) to 2, as documents from the Bitterlemons collection are assumed to reflect either the Israeli or Palestinian viewpoint.

4.3 Evaluation

We performed both quantitative and qualitative evaluation to assess the quality of our model. The quantitative evaluation relies on two metrics: held-out perplexity and viewpoint identification accuracy. It compares the performance of our model VODUM according to these metrics against the aforementioned baselines. In addition, the qualitative evaluation consists in checking the coherence

[1] `http://www.bitterlemons.net/`
[2] `http://alias-i.com/lingpipe/`
[3] `http://jgibblda.sourceforge.net/`

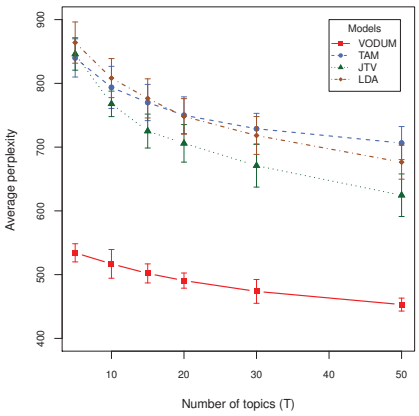

Fig. 3: Held-out perplexity of VODUM, TAM, JTV, and LDA computed for 5, 10, 15, 20, 30, and 50 topics (lower is better). Error bars denote standard deviation in the 10-fold cross-validation.

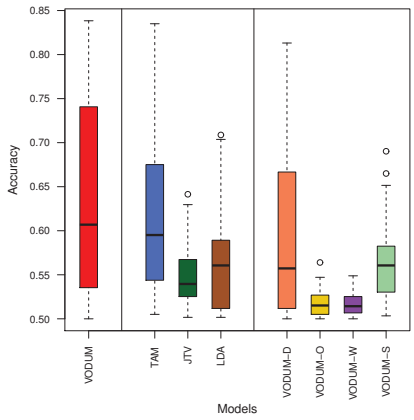

Fig. 4: Viewpoint identification accuracy of VODUM, TAM, JTV, LDA, VODUM-D, VODUM-O, VODUM-W, and VODUM-S (higher is better). Each boxplot is drawn from 50 samples.

of topical words and the related viewpoint-specific opinion words inferred by our model. These evaluations are further described below.

Held-out Perplexity. Held-out perplexity is a metric that is often used to measure the generalization performance of a Topic Model [2]. Perplexity can be understood as the inverse of the geometric mean per-word likelihood. As computing the perplexity of a Topic Model is intractable, an estimate of the perplexity is usually computed using the parameters' point estimate provided by a Gibbs sampler, as shown in Section 3.2. The process is the following: the model is first learned on a training set (i.e., inference is performed to compute the parameters of the model), then the inferred parameters are used to compute the perplexity of the test set (i.e., a set of held-out documents). A lower perplexity for the test set, which is equivalent to a higher likelihood for the test set, can be interpreted as a better generalization performance of the model: the model, learned on the training set, is less "perplexed" by the test set. In this experiment, we aimed to investigate (H5) and compared the generalization performance of our model VODUM against the state-of-the-art baselines. We performed a 10-fold cross-validation as follows. The model is trained on nine folds of the collection for 1,000 iterations and inference on the remaining, held-out test fold is performed for another 1,000 iterations. For both training and test, we only considered the final sample, i.e., the 1,000$^{\text{th}}$ sample. We finally report the held-out perplexity averaged on the final samples of the ten possible test folds. As the generalization performance depends on the number of topics, we computed the held-out perplexity of models for 5, 10, 15, 20, 30, and 50 topics.

The results of this experiment (Fig. 3) support (H5): for all number of topics VODUM has a significantly lower perplexity than TAM, JTV, and LDA. This implies that VODUM's ability to generalize data is better than baselines'. JTV presents slightly lower perplexity than TAM and LDA, especially for larger number of topics. TAM and LDA obtained comparable perplexity, TAM being slightly better for lower number of topics and LDA being slightly better for higher number of topics.

Viewpoint Identification. Another important aspect of our model is its ability to identify the viewpoint under which a document has been written. In this experiment, we aim to evaluate the viewpoint identification accuracy (VIA) of our model VODUM against our baselines, in order to investigate (H1), (H2), (H3), (H4), and (H5). As the Bitterlemons collection contains two different viewpoints (Israeli or Palestinian), viewpoint identification accuracy is here equivalent to binary clustering accuracy: each document is assigned to viewpoint 0 or to viewpoint 1. The VIA is then the ratio of well-clustered documents. As reported in [9], the viewpoint identification accuracy presents high variance for different executions of a Gibbs sampler, because of the stochastic nature of the process. For each model evaluated, we thus performed 50 executions of 1,000 iterations, and kept the final ($1{,}000^{\text{th}}$) sample of each execution, resulting in a total of 50 samples. In this experiment, we set the number of topics for the different models as follows: 12 for VODUM, VODUM-D, VODUM-O, VODUM-W, and VODUM-S. The number of topics for state-of-the-art models was set according to their respective authors' recommendation: 8 for TAM (according to [9]), 6 for JTV (according to [12]). For LDA, the number of topics was set to 2: as LDA does not model viewpoints, we evaluated to what extent LDA is able to match viewpoints with topics.

VODUM, VODUM-D, VODUM-O, and VODUM-W provide documents' viewpoint assignment for each sample. We thus directly used these assignments to compute the VIA. However, VODUM-S only has sentence-level viewpoint assignments. We assigned each document the majority viewpoint assignment of its sentences. When the sentences of a document are evenly assigned to each viewpoint, the viewpoint of the document was chosen randomly. We proceeded similarly with TAM, JTV, and LDA, using their majority word-level aspect, viewpoint, and topic assignments, respectively, to compute the document-level viewpoint assignments. The results of the experiments are given in Fig. 4. The boxplots show that our model VODUM overall performed the best in the viewpoint identification task. More specifically, VODUM outperforms state-of-the-art models, thus supporting (H5). Among state-of-the-art models, TAM obtained the best results. We also observe that JTV did not outperform LDA in the viewpoint identification task. This may be due to the fact that the dependency between topic variables and viewpoint variables was not taken into account when we used JTV to identify document-level viewpoints – word-level viewpoint assignments in JTV are not necessarily aligned across topics. The observations of the degenerate versions of VODUM support (H1), (H2), (H3), and (H4). VODUM-O

Middle East conflicts Topical words	israel	palestinian	syria	jihad	war	iraq	dai	suicid	destruct	iran
Middle East conflicts Opinion words (I)	islam	isra	terrorist	recent	militari	intern	like	heavi	close	american
Middle East conflicts Opinion words (P)	need	win	think	sai	don	strong	new	sure	believ	commit

Table 3: Most probable topical and opinion (stemmed) words inferred by VODUM for the topic manually labeled as *Middle East conflicts*. Opinion words are given for each viewpoint: Israeli (I) and Palestinian (P).

and VODUM-W performed very poorly compared to other models. The separation of topical words and opinion words, as well as the use of sentence-level topic variables – properties that were removed from VODUM in VODUM-O and VODUM-W, respectively – are then both absolutely necessary in our model to accurately identify documents' viewpoint, which confirms (H2) and (H3). The model VODUM-S obtained reasonable VIA, albeit clearly lower than VODUM. Document-level viewpoint variables thus lead to a better VIA than sentence-level viewpoint variables, verifying (H4). Among the degenerate versions of VODUM, VODUM-D overall yielded the highest VIA, but still slightly lower than VODUM. We conclude that the assumption made in [10,11], stating that the use of viewpoint-specific topic distributions (instead of document-specific topic distributions as in VODUM-D) improves viewpoint identification, was relevant, which in turn supports (H1).

Qualitative Evaluation. The qualitative evaluation of our model VODUM consists in studying the coherence of the topical words and the related viewpoint-specific opinion words. More specifically, we examine the most probable words in our model's viewpoint-independent distribution over words ϕ_0 and each viewpoint-specific distribution over words ϕ_1. This evaluation of VODUM is performed on the sample that obtained the best VIA. We report in Table 3 the most probable words for a chosen topic, which we manually labeled as *Middle East conflicts*. The most probable topical words are coherent and clearly relate to Middle East conflicts with words like *syria*, *jihad*, *war*, and *iraq*. The second and third rows of Table 3 show the opinion words used by the Israeli and Palestinian viewpoints, respectively. Not surprisingly, words like *islam*, *terrorist*, and *american* are used by the Israeli side to discuss Middle East conflicts. On the other hand, the Palestinian side remains nonspecific on the conflicts with words like *win*, *strong*, and *commit*, and does not mention Islam or terrorism. These observations confirm that the topical words and the opinion words are related and coherent.

5 Conclusion and Research Directions

This article introduced VODUM, an unsupervised Topic Model that enables viewpoint and opinion discovery in text. Throughout the experiments, we showed that our model outperforms state-of-the-art baselines, both in generalizing data

and identifying viewpoints. We also analyzed the importance of the properties specific to our model. The results of the experiments suggest that the separation of opinion words and topical words, as well as the use of sentence-level topic variables, document-level viewpoint variables, and viewpoint-specific topic distributions improve the ability of our model to identify viewpoints. Moreover, the qualitative evaluation confirms the coherence of topical words and opinion words inferred by our model.

We expect to extend the work presented here in several ways. As the accuracy of viewpoint identification shows a high variance between different samples, one needs to design a method to automatically collect the most accurate sample or to deduce accurate viewpoint assignments from a set of samples. VODUM can also integrate sentiment labels to create a separation between positive and negative opinion words, using sentiment lexicons. This could increase the discrimination between different viewpoints and thus improve viewpoint identification. A viewpoint summarization framework can as well benefit from VODUM, selecting the most relevant sentences from each viewpoint and for each topic by leveraging VODUM's inferred parameters.

References

1. Asuncion, A., Welling, M., Smyth, P., Teh, Y.W.: On Smoothing and Inference for Topic Models. In: Proc. of UAI '09. pp. 27–34 (2009)
2. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. In: Proc. of NIPS '01. pp. 601–608 (2001)
3. Fang, Y., Si, L., Somasundaram, N., Yu, Z.: Mining Contrastive Opinions on Political Texts using Cross-Perspective Topic Model. In: Proc. of WSDM '12. pp. 63–72 (2012)
4. Lin, W.H., Wilson, T., Wiebe, J., Hauptmann, A.: Which Side are You on? Identifying Perspectives at the Document and Sentence Levels. In: Proc. of CoNLL '06. pp. 109–116 (2006)
5. Liu, B., Hu, M., Cheng, J.: Opinion Observer: Analyzing and Comparing Opinions on the Web. In: Proc. of WWW '05. pp. 342–351 (2005)
6. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval 2(1–2), 1–135 (2008)
7. Paul, M.J., Girju, R.: Cross-Cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models. In: Proc. of EMNLP '09. pp. 1408–1417 (2009)
8. Paul, M.J., Girju, R.: A Two-Dimensional Topic-Aspect Model for Discovering Multi-Faceted Topics. In: Proc. of AAAI '10. pp. 545–550 (2010)
9. Paul, M.J., Zhai, C., Girju, R.: Summarizing Contrastive Viewpoints in Opinionated Text. In: Proc. of EMNLP '10. pp. 66–76 (2010)
10. Qiu, M., Jiang, J.: A Latent Variable Model for Viewpoint Discovery from Threaded Forum Posts. In: Proc. of NAACL HLT '13. pp. 1031–1040 (2013)
11. Qiu, M., Yang, L., Jiang, J.: Modeling Interaction Features for Debate Side Clustering. In: Proc. of CIKM '13. pp. 873–878 (2013)
12. Trabelsi, A., Zaiane, O.R.: Mining Contentious Documents Using an Unsupervised Topic Model Based Approach. In: Proc. of ICDM '14. pp. 550–559 (2014)
13. Turney, P.D.: Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. In: Proc. of ACL '02. pp. 417–424 (2002)