Extraction and Clustering of Arguing Expressions in Contentious Text

Amine Trabelsi*, Osmar R. Zaïane

Department of Computing Science, University of Alberta, Edmonton, Canada T6G 2E8

Abstract

This work proposes an unsupervised method intended to enhance the quality of opinion mining in contentious text. It presents a Joint Topic Viewpoint (JTV) probabilistic model to analyse the underlying divergent arguing expressions that may be present in a collection of contentious documents. The conceived JTV has the potential of automatically carrying the tasks of extracting associated terms denoting an arguing expression, according to the hidden topics it discusses and the embedded viewpoint it voices. Furthermore, JTV's structure enables the unsupervised grouping of obtained arguing expressions according to their viewpoints, using a proposed constrained clustering algorithm which is an adapted version of the constrained k-means clustering (COP-KMEANS). Experiments are conducted on three types of contentious documents (polls, online debates and editorials), through six different contentious datasets. Quantitative evaluations of the topic modeling output, as well as the constrained clustering results show the effectiveness of the proposed method to fit the data and generate distinctive patterns of arguing expressions. Moreover, it empirically demonstrates a better clustering of arguing expressions over state-of-the art and baseline methods. The qualitative analysis highlights the coherence of clustered arguing expressions of the same viewpoint and the divergence of opposing ones.

Keywords: Contention Analysis, Topic Models, Arguing Expression Detection, Opinion Mining, Unsupervised Clustering, Online Debates

1. Introduction

Sentiment analysis, also referred to as opinion mining, is an active research area in natural language processing as well as data mining, that aims to extract and examine opinions, attitudes and emotions expressed in text, with respect to some topic in blog posts, comments and reviews. In addition to sentiment expressed towards products, other online text sources such as opinion polls, debate websites and editorials may contain valuable opinion information articulated around some topics of contention. In this

*Corresponding author

Email addresses: atrabels@ualberta.ca (Amine Trabelsi), zaiane@ualberta.ca (Osmar R. Zaïane)

Table 1: Excerpts of support and opposition opinion to a healthcare bill in the USA.

Support Viewpoint
Many people do not have health care
Provide health care for 30 million people
The government should help old people
Oppose Viewpoint
The government should not be involved
It will produce too much debt
The bill would not help the people

paper, we address the issue of improving the quality of opinion mining from contentious texts, found in surveys' responses, debate websites and editorials. Mining and summarizing these resources is crucial, especially when the opinion is related to a subject that stimulates divergent viewpoints within people (e.g., Healthcare Reform, Same-Sex Marriage, Israel/Palestine conflict). We refer to such subjects as issues of contention. A ***contentious issue*** is "*likely to cause disagreement between people*" (cf. Oxford Dictionaries[1]). Documents such as survey reports, debate site posts and editorials may contain multiple contrastive viewpoints regarding a particular issue of contention. Table 1 presents an example of short-text documents expressing divergent opinions where each is exclusively supporting or opposing a healthcare legislation[2]. Opinion in contentious issues is often expressed implicitly, not necessarily through the usage of usual negative or positive opinion words, like "bad" or "great". This makes its extraction a challenging task. It is usually conveyed through the arguing expression justifying the endorsement of a particular point of view. The act of arguing is "to give reasons why you think that something is right/wrong, true/not true, etc, especially to persuade people that you are right" (cf. Oxford Dictionaries). For example, the arguing expression "many people do not have healthcare", in Table 1, implicitly explains that the reform is intended to fix the problem of uninsured people, and thus, the opinion is probably on the supporting side. On the other hand, the arguing expression "it will produce too much debt" denotes the negative consequence that may result from passing the bill, making it on the opposing side.

The automatic identification and clustering of these kind of arguing expressions, according to the topics they invoke and the viewpoints they convey, is enticing for a variety of application domains. For instance, it can save journalists a substantial amount of work and provide them with drafting elements (viewpoints and associated arguing expressions) about controversial issues. Moreover, a good automatic browsing of divergent arguing expressions in a conflict/issue would help inquisitive people understand the issue itself (e.g., same-sex marriage). Also, it may be used by politicians to monitor the change in argumentation trends, i.e., changes in the main reasons expressed to oppose or support viewpoints. The significant changes may indicate the occurrence of an important event (e.g., a success of a politician's action or speech). Automatic summarization of arguing expressions may benefit survey companies who usually collect large verbatim reports

[1]`http://www.oxfordlearnersdictionaries.com/definition/english/contentious`

[2]extracted from a Gallup Inc. survey `http://www.gallup.com/poll/126521/favor-oppose-obama-healthcare-plan.aspx`

Table 2: Human-made summary of arguing expressions supporting and opposing Obamacare.

Support Viewpoint	***Oppose Viewpoint***
People need health insurance/many uninsured	Will raise cost of insurance/ less affordable
System is broken/needs to be fixed	Does not address real problems
Costs are out of control/help control costs	Need more information on how it works
Moral responsibility to provide/Fair	Against big government involvement (general)
Would make healthcare more affordable	Government should not be involved in healthcare
Don't trust insurance companies	Cost the government too much

about people's opinion regarding an issue of contention. From a text mining perspective, representing a contentious document, as a small set of dimensions, each corresponding to an arguing expression of a different topic and viewpoint, is useful for information retrieval tasks like query answering or dimensionality reduction. In addition, it would enhance the output quality of the opinion summarization task in general.

The rest of this paper is organized as follows. Section 2 states the problem. Section 3 explains the key issues in the context of recent related work. Section 4 provides the technical details of the proposed model, the Joint Topic Viewpoint model (JTV). Section 5 describes the clustering algorithm applied on JTV output and used to obtain a feasible solution. Section 6 provides a description of the experimental set up on three different types of contentious text. Section 7 quantitatively assesses the JTV and constrained clustering adequacy and compares the performance of our solution with state-of-the-art and baseline methods. Section 8 qualitatively evaluate the coherence and the relevance of the final output of our method. Section 9 discusses the future work.

2. Problem Statement

This paper examines the task of mining the underlying topics and the hidden viewpoints of arguing expressions towards the summarization of contentious text. An example of a human-made summary of arguing expressions [1] on, what is commonly known as the Obama healthcare reform, is presented in Table 2. The ultimate research's target is to automatically generate similar snippet-based summaries given a corpus of contentious documents. However, this paper tackles the initial sub-problem of identifying recurrent words and phrases expressing arguing and cluster them according to their topics and viewpoints. This would help solve the general problem. Indeed, the clustered words and phrases can be used as input to query the original documents via information retrieval methods in order to extract relevant fragments or snippets of text related to a particular arguing expression. We use Table 2 examples to define some key concepts which can help us formulate the general problem. Here, the contentious issue yielding the divergent positions is the Obama healthcare. The documents are people's verbatim responses to the question "Why do you favor or oppose a healthcare legislation similar to President Obama's ?".

A ***contention question*** is a question that can generate expressions of two or more divergent viewpoints as a response.

While the previous question explicitly asks for the reasons ("why"), we relax this constraint and consider also usual opinion questions like "Do you favor or oppose Obamacare ?", or "What do you think about Obamacare?".

A ***contentious document*** is a document that contains expressions of one or more divergent viewpoints in response to a contention question.

Table 2 is split into two parts according to the viewpoint: supporting or opposing the healthcare bill. Each row contains one or more phrases, each expressing a reason (or an explanation), e.g., "System is broken" and "needs to be fixed". Though lexically different, these phrases share a common hidden theme (or topic), e.g., healthcare system, and implicitly convey the same hidden viewpoint's semantics, e.g., support the healthcare bill. Thus, we define an ***arguing expression*** as the set of reasons (snippets: words or phrases) sharing a common topic and justifying the same viewpoint regarding a contentious issue.

A ***viewpoint*** (e.g., a column of Table 2) in a contentious document is a stance, in response to a contention question, which is implicitly expressed by a set of arguing expressions (e.g., rows of a column in Table 2).

Thus, the arguing expressions voicing the same viewpoint differ in their topics, but agree in the stance. For example, arguing expressions represented by "system is broken" and "costs are out of control" discuss different topics, i.e., healthcare system and insurance's cost, but both support the healthcare bill. On the other hand, arguing expressions of divergent viewpoints may have similar topic or may not. For instance, "government should help elderly" and "government should not be involved" share the same topic "government's role" while conveying opposed viewpoints.

Our research problem and objectives in terms of the newly introduced concepts are stated as follows. Given a corpus of unlabeled contentious documents $\{doc_1, doc_2, .., doc_D\}$, where each document doc_d expresses one or more viewpoints $\vec{v}^d$ from a set of L possible viewpoints $\{v_1, v_2, .., v_L\}$, and each viewpoint v_l can be conveyed using one or more arguing expressions $\vec{\phi}_l$ from a set of possible arguing expressions discussing K different topics $\{\phi_{1l}, \phi_{2l}, .., \phi_{Kl}\}$, the objective is to perform the following two tasks:

1. automatically extracting coherent words and phrases describing any distinct arguing expression ϕ_{kl};
2. grouping extracted distinct arguing expressions ϕ_{kl} for different topics, $k = 1..K$, into their corresponding viewpoint v_l.

In carrying out the first task, we must meet the main challenge of recognizing arguing expressions having the same topic and viewpoint but which are lexically different, e.g., "provide health care for 30 million people " and " many people do not have healthcare". For this purpose we propose a Joint Topic Viewpoint model (JTV) to account for the dependence structure of topics and viewpoints. For the second task, the challenge is to deal with the situation where an arguing expression, associated with a specific topic, may share more common words and phrases with a divergent arguing expression, discussing the same topic, than with another arguing expression conveying the same viewpoint but discussing a different topic. Recall, the example "government should help elderly" is lexically more similar to "government should not be involved" than to "many people uninsured".

3. Related Work

It is important to note that we do not intend to address argumentation analysis. A large body of early work on argumentation was based on learning deterministic logical concepts [2]. Argumentation theory is the study of how conclusions can be reached from some premises through logical reasoning. In argumentation, one critically examines beliefs to discard wrong claims and build knowledge from supported assertions following the Cartesian view of reasoning. In this work, our targeted text is online text in opinion polls, discussion forums, etc. voicing opinions of laypersons. Apart from long editorials, these text sources are typically short, in which reasoning is not necessarily laid out but claims and point of views are put forward using arguing expressions. There is little or no rationalization or discursive reasoning in online short surveys or micro blogs. Moreover dealing with these types of opinionated real data, unavoidably requires the means to handle the uncertainty (as opposed to determinism) or the ambiguity that arises from incomplete or hidden information (implicit, unsaid or unexpressed topic or a viewpoint). Our objective is not to create a linguistically motivated framework for semantic inference of argumentative structure (e.g., [3]). However, our objective is to design a statistical learning model in order to discover the main arguing expressions and group them by viewpoint. In this section we present a number of the common themes, issues and important concepts in some related work. Potential links to our approach of mining opinion in text of contention are put forward.

3.1. Classifying Stances

An early body of work addresses the challenge of classifying viewpoints in contentious or ideological discourses using supervised techniques [4, 5]. Although the models give good performance, they remain data-dependent and costly to label, making the unsupervised approach more appropriate for the existing huge quantity of online data. A similar trend of studies scrutinizes the discourse aspect of a document in order to identify opposed stances [6, 7]. However, these methods utilize polarity lexicon to detect opinionated text and do not look for arguing expression, which is shown to be useful in recognizing opposed stances [8]. Somasundaran and Wiebe [8] classify ideological stances in online debates using generated arguing clues from the Multi Perspective Question Answering (MPQA) opinion corpus[3]. Our problem is not to classify documents, but to recognize recurrent pattern of arguing phrases instead of arguing clues. Moreover, our approach is independent of any annotated corpora.

3.2. Topic Modeling in Reviews Data

Another emerging body of work applies probabilistic topic models on reviews data to extract appraisal aspects and the corresponding specific sentiment lexicon. These kinds of models are usually referred to as joint sentiment/aspect topic models [9, 10, 11]. Lin and He [12] propose the Joint Sentiment Topic Model (JST) to model the dependency between sentiment and topics. They make the assumption that topics discussed on a review are conditioned on sentiment polarity. Reversely, our JTV model assumes that a viewpoint endorsement (e.g., oppose reform) is conditioned on the discussed topic (e.g., government's role). Moreover, JTV's application is different from that of JST. Most of the joint aspect sentiment topic models are either semi-supervised or weakly supervised using sentiment polarity words (Paradigm lists) to boost their efficiency. In our case, viewpoints are often expressed implicitly and finding specific arguing lexicon for different

[3] `http://mpqa.cs.pitt.edu/`

stances is a challenging task in itself. Indeed, our model is enclosed in another body of work based on a Topic Model framework to mine divergent viewpoints.

3.3. Topic Modeling in Contentious Text

Lin et al. [13] propose a probabilistic graphical model for ideological discourse. This model takes into account lexical variations between authors having different ideological perspectives. The authors empirically show its effectiveness in fitting ideological texts. However, their model assumes that the perspectives expressed in the documents are observed, while, in our work, the viewpoint labels of the contentious documents are hidden.

A recent studies by Mukherjee and Liu [14, 15] examine mining contention from discussion forums data where the interaction between different authors is pivotal. They attempt to jointly discover contention/agreement indicators (CA-Expressions) and topics using three different Joint Topic Expressions models (JTE). The JTEs' output is used to discover points (topics) of contention. The model supposes that people express agreement or disagreement through CA-expressions. However, this is not often the case when people express their viewpoint via other channels than discussion forums like debate sites or editorials. Moreover, agreement or disagreement may also be conveyed implicitly through arguing expressions rejecting or supporting another opinion. JTEs do not model viewpoints and use the supervised Maximum Entropy model to detect CA-expressions.

Qiu and Jiang [16] also incorporate the information on users interactions in threaded debate forums within a topic model. The goal is to model both the posts and the users in a thread and cluster them according to their viewpoints. The topic model is based on three major hypothesis: (1) the topics discussed in divergent viewpoints tend to be different; (2) a user is holding the same viewpoint in all his posts in the thread; and (3) users with the same viewpoints have positive interactions while negative interactions are more probable in the opposite case. In our work, we assume that topics are shared between divergent viewpoints. However, the topics' proportions and their related lexicon are different according to the viewpoint. We focus on capturing the lexical variations between divergent viewpoints, instead of the agreement/disagreement between users. While the users interactions can be very useful for posts classification or clustering, our primary goal is different, i.e., it aims at extracting and clustering meaningful arguing expressions towards the summarization of main contention points in an issue. Moreover, our model tends to be generalizable to different types of contentious text (e.g., surveys responses, editorials) which do not necessarily embrace the same structure of threaded debate forums (i.e., do not contain users information and users interaction).

Fang et al. [17] proposed a Cross-Perspective Topic model (CPT) that takes as input separate collections in the political domain, each related to particular viewpoint (perspective). It finds the shared topics between these different collections and the opinion words corresponding to each topic in a collection. However, CPT does not model the viewpoint variable. Thus, it cannot cluster documents according to their viewpoints. Moreover, the discovered topics are not necessarily of contention. Recently, Gottipati et al. [18] propose a topic model to infer human interpretable text in the domain of issues using Debatepedia[4] as a corpus of evidence. Debatepedia is an online authored encyclopedia to summarize and organize the main arguments of two possible positions. The

[4]`http://dbp.idebate.org`

model takes advantage of the hierarchical structure of arguments in Debatepedia. Our work aims to model unstructured online data, with unrestricted number of positions, in order to, ultimately, help extract a relevant contention summary.

Topic Aspect Model. The closest work to ours is the one presented by Paul et al. [19]. It introduces the problem of contrastive summarization which is very similar to our stated problem in Section 2. They propose the Topic Aspect Model (TAM), represented in Figure 1a, and use the output distributions to compute similarities' scores for sentences. Scored sentences are used in Comparative LexRank [20], a modified Random Walk algorithm, as input to generate the summary.

TAM assumes that any word in the document is associated with a vector of four hidden assignments (l, x, z, y) (See Plate representation of TAM in Figure 1b) . The variable l indicates if a word is a background word or a topical word. z represents a topical assignment. x is a binary variable accounting for the existence of a viewpoint (whether the word expresses any viewpoint or not). It depends on the background/topic value l and the topic assignment z. y represents the viewpoint assignment. Following this scheme, a word can exclusively belong to a background (e.g., think), to a topic (e.g., government), a viewpoint (e.g., good), or both (e.g., involvement). Nevertheless, TAM does not model any dependency between the assignment of a topic and the value of a viewpoint. We consider that this type of dependency should be encoded in our Joint Topic Viewpoint model. We assume that an author chooses the words conveying his viewpoint (e.g., "no government involvement" which express an implicit opposing stance) according to the chosen arguing topic under discussion (e.g., government role).

4. Joint Topic Viewpoint Model

Latent Dirichlet Allocation (LDA) [21] is one of the most popular topic models used to mine text data sets (See Figure 1c). It models a document as a mixture of topics that are discussed in a corpus. Thus, each document is considered as a mixture of the topics, i.e., a membership probabilities to these topics. Topics are different probability distributions over words (the corpus vocabulary). Every word in a document is assigned to a topic. The words in the same sentence or document can be assigned different topics. However, LDA fails to model more complex structures of texts like contention where viewpoints are hidden.

We augment LDA to model a contentious document as a pair of dependent mixtures: a mixture of arguing topics and a mixture of viewpoints for each topic (See Figure 1a). The assumption is that a document discusses the topics in proportions, (e.g., 80% government's role, 20% insurance's cost). Moreover, as explained in Section 2, each one of these topics can be shared by divergent arguing expressions conveying different viewpoints. We suppose that for each discussed topic in the document, the viewpoints are expressed in proportions. For instance, 70% of the document's text discussing the government's role expresses an opposing viewpoint to the reform while 30% of it conveys a supporting viewpoint. Thus, each term in a document is assigned a pair topic-viewpoint label (e.g., "government's role-oppose reform"). A term is a word or a phrase i.e., n-grams (n>1). For each topic-viewpoint pair, the model generates a topic-viewpoint probability distribution over terms. This topic-viewpoint distribution would correspond to what we

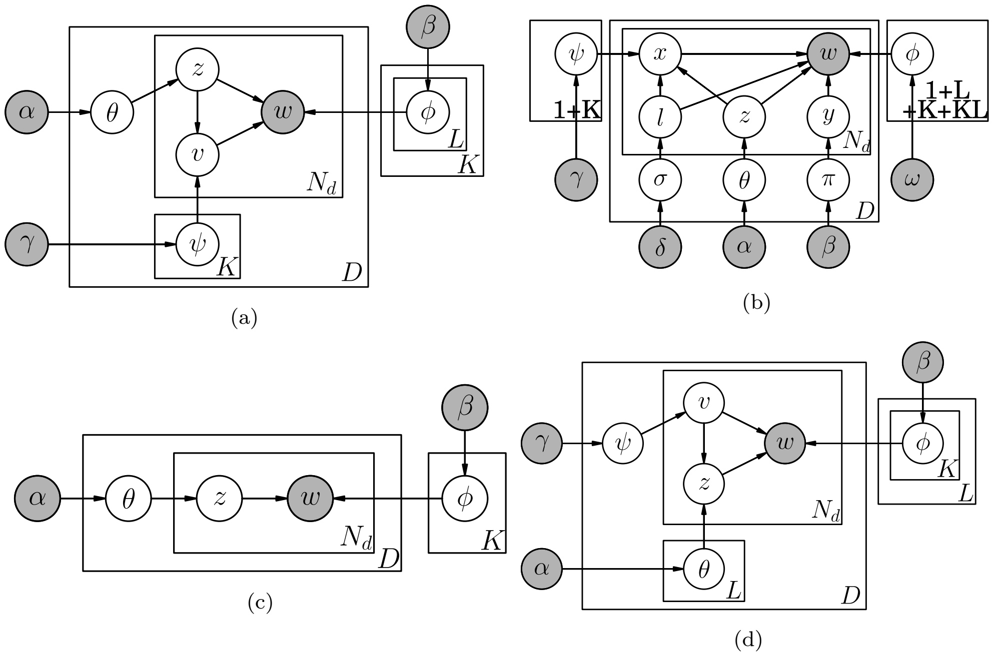


Figure 1: Plate notation of: (a) JTV graphical model; (b)TAM graphical model ; (c) LDA graphical model and (d) JVT graphical model.

define as an arguing expression in Section 2, i.e., a set of terms sharing a common topic and justifying the same viewpoint regarding a contentious issue.

Formally, we assume that a corpus contains D documents $d_{1..D}$, where each document is a term's vector $\vec{w}_d$ of size N_d; each term w_{dn} in a document belongs to the corpus vocabulary of distinct terms of size V. Let K be the total number of topics and L be the total number of viewpoints. Let θ_d denote the probabilities (proportions) of K topics under a document d; ψ_{dk} be the probability distributions (proportions) of L viewpoints for a topic k in the document d (the number of viewpoints L is the same for all topics); and ϕ_{kl} be the multinomial probability distribution over terms associated with a topic k and a viewpoint l.

The generative process (see the JTV graphical model in Fig. 1a) is:

- for each topic k and viewpoint l, draw a multinomial distribution over the vocabulary V: $\phi_{kl} \sim Dir(\beta)$;
- for each document d,
  - draw a topic mixture $\theta_d \sim Dir(\alpha)$
  - for each topic k, draw a viewpoint mixture $\psi_{dk} \sim Dir(\gamma)$
  - for each term w_{dn}
    sample a topic assignment $z_{dn} \sim Mult(\theta_d)$
    sample a viewpoint assignment $v_{dn} \sim Mult(\psi_{dz_{dn}})$
    sample a term $w_{dn} \sim Mult(\phi_{z_{dn}v_{dn}})$

We use fixed symmetric Dirichlet's parameters γ, β and α. They can be interpreted as the prior counts of: terms assigned to viewpoint l and topic k in a document; a particular term w assigned to topic k and viewpoint l within the corpus; terms assigned to a topic k in a document, respectively. In order to learn the hidden JTV's parameters ϕ_{kl}, ψ_{dk} and θ_d, we draw on approximate inference as exact inference is intractable [21]. We use the collapsed Gibbs Sampling [22], a Markov Chain Monte Carlo algorithm. The collapsed Gibbs sampler integrate out all parameters ϕ, ψ and θ in the joint distribution of the model and converge to a stationary posterior distribution over viewpoints' assignments $\vec{v}$ and all topics' assignments $\vec{z}$ in the corpus. It iterates on each current observed token w_i and samples each corresponding v_i and z_i given all the previous sampled assignments in the model $\vec{v}_{\neg i}$, $\vec{z}_{\neg i}$ and observed $\vec{w}_{\neg i}$, where $\vec{v} = \{v_i, \vec{v}_{\neg i}\}$, $\vec{z} = \{z_i, \vec{z}_{\neg i}\}$, and $\vec{w} = \{w_i, \vec{w}_{\neg i}\}$. The derived sampling equation is:

$$p(z_i = k, v_i = l | \vec{z}_{\neg i}, \vec{v}_{\neg i}, w_i = t, \vec{w}_{\neg i}) \propto \frac{n_{kl,\neg i}^{(t)} + \beta}{\sum_{t=1}^{V} n_{kl,\neg i}^{(t)} + V\beta} \times \frac{n_{dk,\neg i}^{(l)} + \gamma}{\sum_{l=1}^{L} n_{dk,\neg i}^{(l)} + L\gamma} \times n_{d,\neg i}^{(k)} + \alpha \quad (1)$$

where $n_{kl,\neg i}^{(t)}$ is the number of times term t was assigned to topic k and the viewpoint l in the corpus; $n_{dk,\neg i}^{(l)}$ is the number of times viewpoint l of topic k was observed in document d; and $n_{d,\neg i}^{(k)}$ is the number of times topic k was observed in document d. All these counts are computed excluding the current token i, which is indicated by the symbol $\neg i$. After the convergence of the Gibbs algorithm, the parameters ϕ, ψ and θ are estimated using the last obtained sample. The probability that a term t belongs to a viewpoint l of topic k is approximated by:

$$\phi_{klt} = \frac{n_{kl}^{(t)} + \beta}{\sum_{t=1}^{V} n_{kl}^{(t)} + V\beta}. \quad (2)$$

The probability of a viewpoint l of a topic k under document d is estimated by:

$$\psi_{dkl} = \frac{n_{dk}^{(l)} + \gamma}{\sum_{l=1}^{L} n_{dk}^{(l)} + L\gamma}. \quad (3)$$

The probability of a topic k under document d is estimated by:

$$\theta_{dk} = \frac{n_{d}^{(k)} + \alpha}{\sum_{k=1}^{K} n_{d}^{(k)} + K\alpha}. \quad (4)$$

5. Constrained Clustering Algorithm for Arguing Expressions

We mentioned in the previous section that an inferred topic-viewpoint distribution ϕ_{kl} can be assimilated to an arguing expression. For convenience, we will use "arguing expression" and "topic-viewpoint" interchangeably to refer to the topic-viewpoint distribution. Indeed, two topic-viewpoint ϕ_{kl} and $\phi_{k'l}$, having different topics k and k', do not necessarily express the same viewpoint, despite the fact that they both have the same index l. The reason stems from the nested structure of the model, where the generation of the viewpoint assignments for a particular topic k is completely independent from that of topic k'. In other words, the model does not trace and match the viewpoint labeling along different topics. Nevertheless, the JTV can still help overcome this problem. According to the JTV's structure, a topic-viewpoint ϕ_{kl}, is probably more similar in distribution to a divergent topic-viewpoint $\phi_{kl'}$, related to the same topic k, than to any other topic-viewpoint $\phi_{k'*}$, corresponding to a different topic k' (we verify this assumption in Section 7.1.2). Therefore, we can formulate the problem of clustering arguing expressions as a constrained clustering problem [23]. The goal is to group the similar topics-viewpoints ϕ_{kl}s into L clusters (number of viewpoints), given the constraint that the L ϕ_{kl}s of the same topic k should not belong to the same cluster (cannot-link constraints). Thus, each cluster C_i where $i = 1..L$ will contain exactly K topics-viewpoints.

Algorithm 1 Topic-Viewpoint Clustering

Require: JTV's output:topic-viewpoint distributions ϕ_{kl}s, number of topics K, number of viewpoints L

1: Initialize the set C with a set of empty clusters; Choose the topic-viewpoint distributions $\phi_{k^\dagger 1}...\phi_{k^\dagger L}$ of the most frequent topic $k^\dagger$ according to JTV as the initial cluster centers.
2: **for** each topic k ($k = 1...K$) **do**
3:   F (clusters to fill) is a copy of set C
4:   A is a set of L topic-viewpoints ϕ_{kl} to assign (having the same topic k)
5:   **while** F is not empty **do**
6:     **for** each ϕ_{kl} in A **do**
7:       find the closest C_i in F
8:       add ϕ_{kl} to potential cluster assignment set S_i (corresponding to cluster C_i)
9:     **end for**
10:     **for** each cluster C_i **do**
11:       **if** the corresponding S_i is not empty **then**
12:         find ϕ^*_{kl} in S_i with the minimum distance from C_i's center and assign it to C_i.
13:         Update C
14:         empty S_i
15:         remove ϕ^*_{kl} from A/remove C_i from F
16:       **end if**
17:     **end for**
18:   **end while**
19: **end for**
20: Update each cluster C_i's center by averaging all $\phi^{(i)}$ that have been assigned to it.
21: Repeat 2 to 20 until convergence
22: **return** set of clusters C

We suggest a slightly modified version of the constrained k-means clustering (COP-KMEANS) [24]. It is presented in Algorithm 1. Unlike COP-KMEANS, we do not

Table 3: Statistics on the six used data sets

	OC		AW		GM1		GM2		IP1		IP2	
View	for	Ag	allow	not	illegal	not	hurt	no	pal	is	pal	is
#doc	434	508	213	136	44	54	149	301	149	149	148	148
tot. #tokens	14594		44482		10666		47915		209481		247059	
Avg. lg. doc.	15.49		127.45		108.83		106.47		702.95		834.65	

consider any must-link constraint but only the above mentioned cannot-link constraints. The centers of clusters are initialized with the topic-viewpoint distributions of the most frequent topic $k^{\dagger}$ according to the output of JTV. The idea is that it is more probable to find at least one most frequent topic-viewpoint pair for a viewpoint l in the most frequent topic $k^{\dagger}$. The cannot-link constraints are implicitly coded in Algorithm 1. Indeed, we constrain the set of L topic-viewpoint ϕ_{kl}s of the same topic k (line 2 to 18) to be in a one-to-one matching with the set C of L clusters (lines 5 to 18). Iteratively, the best match, producing a minimal distance between unassigned topic-viewpoints (of the same topic) and the remaining available clusters, is first established (lines 10 to 16). The distance between a topic-viewpoint distribution ϕ_{kl} and another distribution ϕ_* is measured using the symmetric Jensen-Shannon Distance (D_{JS}) [25] which is based on the Kullback-Leibler Divergence (D_{KL}) [26]:

$$D_{JS}(\phi_{kl}||\phi_*) = \frac{1}{2}[D_{KL}(\phi_{kl}||M) + D_{KL}(\phi_*||M)], \tag{5}$$

with $M = \frac{1}{2}(\phi_{kl} + \phi_*)$ an average variable and

$$D_{KL}(\phi_{kl}||M) = \sum_{t=1}^{V} \phi_{klt}[\log_2 \phi_{klt} - \log_2 p(M = t)], \tag{6}$$

where V is the size of the distinct vocabulary terms and ϕ_{klt} is defined in equation 2.

6. Experimental Set up

In order to evaluate the performances of the JTV model, we utilize three types of multiple contrastive viewpoint text data: (1) short-text data where people on average express their viewpoint briefly with few words like survey's verbatim response or social media posts; (2) mid-range text where people develop their opinion further using few sentences, usually showcasing illustrative examples justifying their stances; (3) long text data, mainly editorials where opinion is expressed in structured and verbose manner. Throughout the evaluation procedure, analysis is performed on six different data sets, corresponding to different contention issues. Table 3 describes the used data sets.

ObamaCare (OC)[5] consists of short verbatim responses concerning the "Obamacare" bill. The survey was conducted by Gallup®from March 4-7, 2010. People were asked why they would oppose or support a bill similar to Obamacare. Table 2 is a human-made summary of this corpus.

[5] http://www.gallup.com/poll/126521/favor-oppose-obama-healthcare-plan.aspx

Assault Weapons (AW)[6]: includes posts extracted from "debate.com". The contention question is "Should assault weapons be allowed in the United States as means of allowing individuals to defend themselves?". The viewpoints are either "should be allowed" or "should not be allowed".

Gay Marriage 1 (GM1)[7]: contains posts from "debate.com" related to the contention question "Should gay marriage be illegal?". The posts' stance are either "should be illegal" or "should be legal".

Gay Marriage 2 (GM2)[8]: contains posts in "createdebate.com" responding to the contention question "How can gay marriage hurt anyone?". Users indicate the stance of their posts (i.e., "hurts everyone?/ does hurt" or "doesn't hurt").

Israel-Palestine (IP) 1 and 2[9]: are two datasets extracted from BitterLemons web site. Israel-Palestine 1 contains articles of two permanent editors, a Palestinian and an Israeli, about the same issues. Articles are published weekly from 2001 to 2005. They discuss several contention issues, e.g., "the American role in the region" and "the Palestinian election". Israel-Palestine 2 contains also weekly articles about the same issues from different Israeli and Palestinian guest authors invited by the editors to convey their views sometimes in form of interviews. Note that each issue, in these data sets' articles, corresponds to a different contention question. Although this does not correspond to our input assumption (i.e., all documents discuss the same contention issue), we are exploring this corpus to measure the scalability of our method for long editorial documents. Moreover, this is a well-known data set used by most of the previous related work in contention [5, 27, 19].

Paul et al. [19] stress the importance of negation features in detecting contrastive viewpoints. Thus, we performed a simple treatment of merging any negation indicators, like "nothing", "no one", "never", etc., found in text with the following occurring word to form a single token. Moreover, we merge the negation "not" with any auxiliary verb (e.g., is, was, could, will) preceding it. Then, we removed the stop-words.

Throughout the experiments below, the JTV's hyperparameters are set to fixed values. The γ is set, according to Steyvers and Griffiths's [28] hyperparameters settings, to $50/L$, where L is the number of viewpoints. β and α are adjusted manually, to give reasonable results, and are both set to 0.01. Along the experiments, we try a different number of topics K. The number of viewpoints L is equal to 2. The number of the Gibbs Sampling iterations is 1000. The TAM model [19] (Section 3.3) and LDA [22] are run as a means of comparison during the evaluation. TAM parameters are set to their default values with same number of topics and viewpoints as JTV. LDA is run with a number of topics equal to twice the number of JTV's topics K, $\beta = 0.01$ and $\alpha = 50/2K$.

7. Quantitative Evaluation

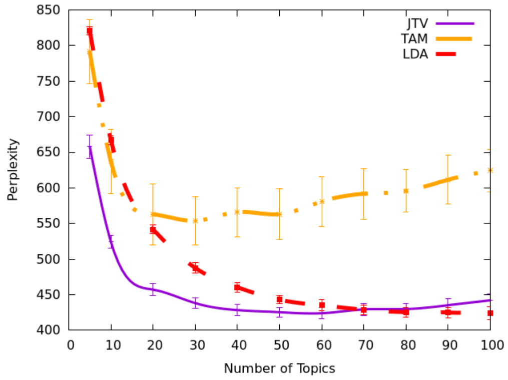
We proceed to a two-fold quantitative analysis of our methods. The first fold of evaluations concerns the assessment of the topic-modeling output of the JTV (Section

[6]`http://www.debate.org/opinions/should-assault-weapons-be-allowed-in-the-united-states-as-means-of-allowing-individuals-to-defend-themselves`

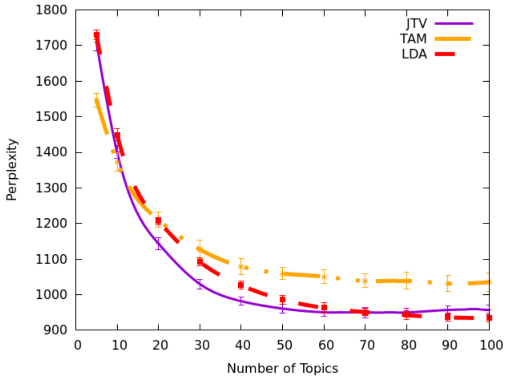
[7]`http://www.debate.org/opinions/should-gay-marriage-be-illegal`

[8]`http://www.createdebate.com/debate/show/How_can_gay_marriage_hurt_any_one`

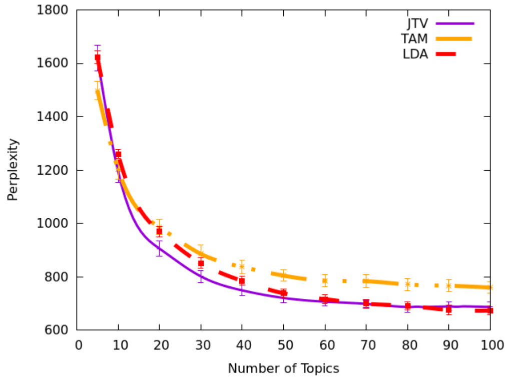
[9]`http://www.bitterlemons.net/`



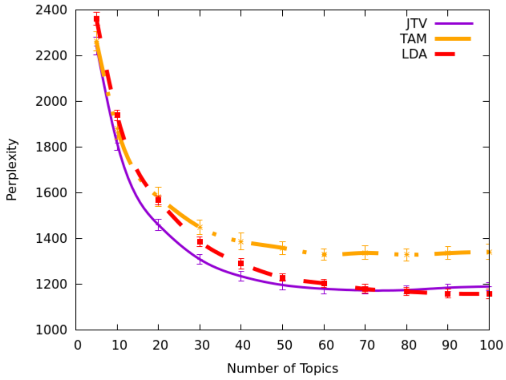
(a) OC



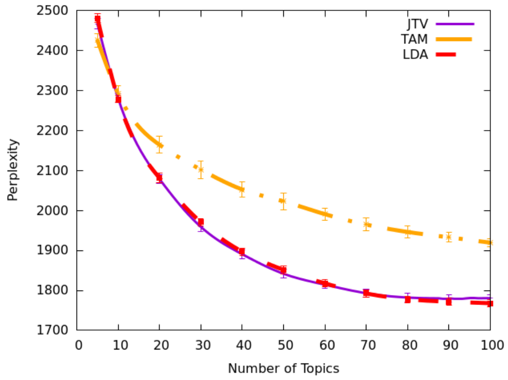
(b) AW



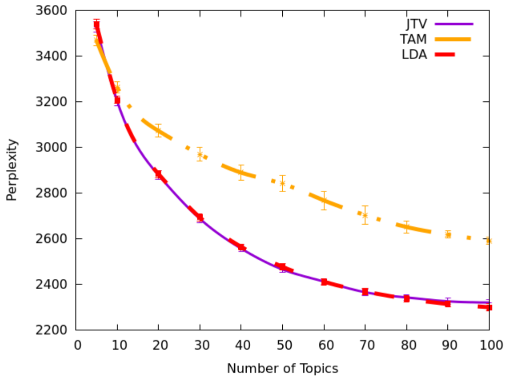
(c) GM1



(d) GM2



(e) IP1



(f) IP2

Figure 2: JTV, LDA and TAM's perplexity plots for six different datasets (lower is better)

4). The second fold of evaluations assesses the constrained clustering task (Section 5). This task uses the JTV's topic-viewpoint distributions as input and tries to cluster them according to their common hidden viewpoint.

7.1. Topic Modeling Evaluation

In order to evaluate the quality of our Joint Topic Model's output, we perform two tasks. The first is the model adequacy where we judge how well our JTV model fits six different data sets. The second is the model generating capacity where we assess how well it is able to generate distinct topic-viewpoint pairs. For the two tasks we benchmark our model against TAM, which incorporates the topic-viewpoint dimension, as well as against the LDA model. The number of topics given as input to LDA is equal to the number of topic-viewpoint pairs. For the evaluation procedure we use two metrics.

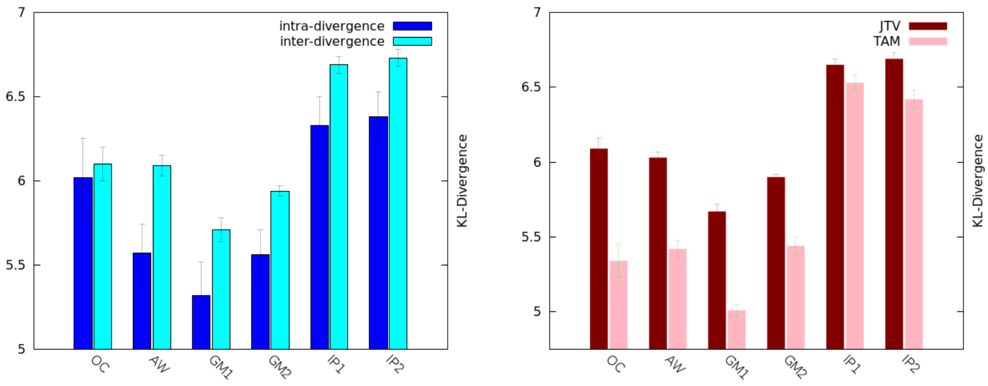
7.1.1. Held-Out Perplexity

We use the perplexity criterion to measure the ability of the learned topic model to fit a new held-out data. Perplexity assesses the generalization performance and, subsequently, provides a comparing framework of learned topic models. The lower the perplexity, the less "perplexed" is the model by unseen data and the better the generalization. It algebraically corresponds to the inverse geometrical mean of the test corpus' terms likelihoods given the learned model parameters [25]. We compute the perplexity under estimated parameters of JTV and compare it to those of TAM and LDA for our six unigrams data sets (Section 6). Figure 2 exhibits, for each corpus, the perplexity plot as function of the number of topics K for JTV, TAM and LDA. For a proper comparison the number of topics of LDA is set to $2K$. Note that for each K, we run the model 50 times. The drawn perplexity corresponds to the average perplexity on the 50 runs where each run computes one-fold perplexity from a 10-fold cross-validation. The figures show evidence that the JTV outperforms TAM for all data sets, used in the experimentation. We can also observe that the JTV's perplexity tend to reach its minimal values for a smaller number of topics than LDA for short and medium length text. For large text, JTV and LDA perplexities are very similar.

7.1.2. Kullback-Leibler Divergence

Kullback-Leibler (KL) Divergence is used to measure the degree of separation between two probability distributions (see Equation 6)[10]. We utilize it for two purposes. The first purpose is to empirically validate the assumption on which the clustering algorithm in Section 5 is based. The assumption states that, according to JTV's structure, a topic-viewpoint ϕ_{kl} is more similar in distribution to a topic-viewpoint $\phi_{kl'}$, related to the same topic k, than to any other topic-viewpoint $\phi_{k'*}$, corresponding to a different topic k'. Thus, two measures of *intra* and *inter-divergence* are computed. The *intra-divergence* is an average KL-Divergence between all topic-viewpoint distributions that are associated with a same topic. The *inter-divergence* is an average KL-Divergence between all pairs of topic-viewpoint distributions belonging to different topics. Figure 3a displays the histograms of JTV's intra and inter divergence values for the six data sets. These quantities are averages on 20 runs of the model for an input number of topics $K = 5$,

[10] Here D_{KL} is computed using the natural logarithm instead of the binary logarithm.



(a) (b)

Figure 3: Histograms of: (a) average topic-viewpoint intra/inter divergences of JTV; (b) average of overall topic-viewpoint divergences of JTV and TAM for six datasets ($K = 5$).

which gives the best differences between the two measures. We observe that a higher divergence is recorded between topic-viewpoints of different topics than between those of a same topic. This is verified for all the data sets considered in our experimentation. The differences between the intra and inter divergences are significant ($p - value < 0.01$) over unpaired t-test (except for ObamaCare). The second purpose of using KL-Divergence is to assess the distinctiveness of generated topic-viewpoint dimensions by JTV and TAM. This is an indicator of a good aggregation of arguing expressions. For a proper comparison, we do not assess LDA's distinctiveness as this latter does not model the hidden viewpoint variable. We compute an *overall-divergence* quantity, which is an average KL-Divergence between all pairs of topic-viewpoint distributions, for JTV and TAM and compare them. Figure 3b illustrates the results for all datasets. Quantities are averages on 20 runs of the models. Both models are run with a number of topics $K = 5$, which gives the best divergences for TAM. Comparing JTV and TAM, we notice that the overall-divergence of JTV's topic-viewpoint is significantly ($p - value < 0.01$) higher for all data sets. This result reveals a better quality of our JTV extracting process of arguing expressions (the first task stated in Section 2).

7.2. Constrained Clustering Evaluation

In this section, we evaluate our clustering algorithm presented in Section 5. The objective of the clustering is to group the similar arguing expressions assimilated to the topic-viewpoint distributions ϕ_{kl}s, provided by the JTV, into $L = 2$ clusters corresponding to the viewpoints. Here, unlike the previous section, we assess the relevance of the topics-viewpoints grouping instead of their intrinsic quality. We take advantage of the documents labels during this evaluation. Indeed, the relevance of arguing expressions grouping according to their viewpoints is measured by the correct clustering percentage of the documents (CCP). A document is clustered given the output of the constrained clustering algorithm. In fact, as explained in Section 4, each word in a document is assigned a topic label k and a viewpoint label l by JTV. Each pair of assignments $\{l, k\}$, and subsequently each word in a document, is assigned to cluster C_i, $i = 1..L$ where

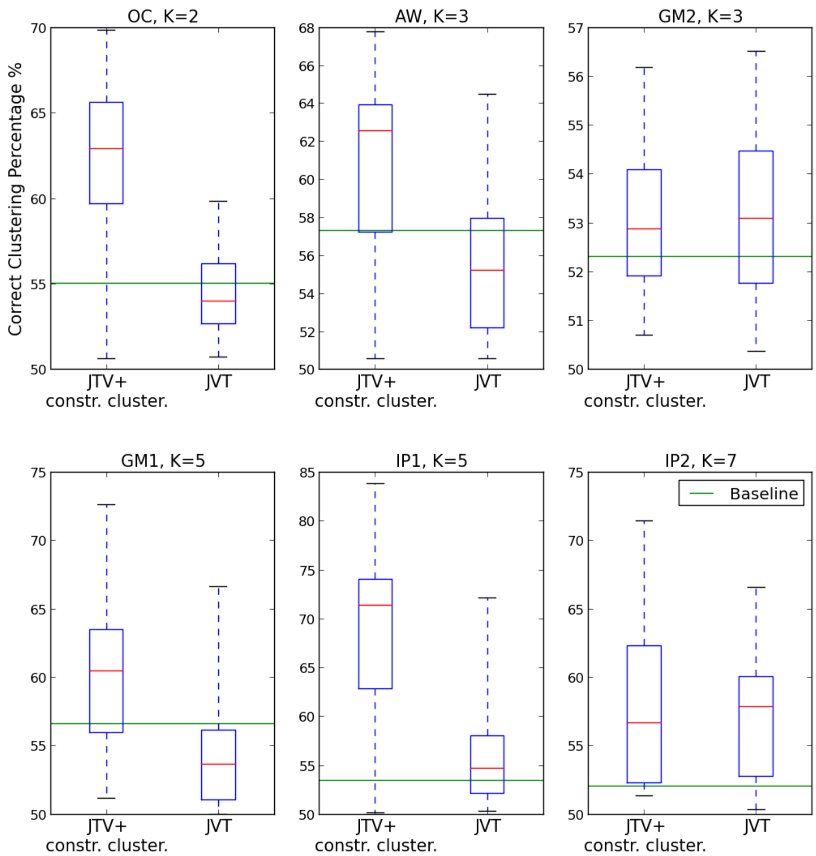


Figure 4: Boxplots of Correct Clustering Percentage (CCP) of the combination JTV + the Constrained Clustering algorithm, the JVT and the baseline method for the six different datasets. The deterministic percentage of the baseline is shown with the green horizontal line.

$L = 2$, by the constrained clustering algorithm (Algorithm 1). Thus, a document can be assigned the majority label C_i within its contained words. In case of parity, a document is discarded. We compare each document's assignment to clusters C_1 or C_2, when the number of viewpoints L is 2, to its original viewpoint label in the ground truth. We choose the matching between the cluster label and the correct viewpoint label that provides the best correct clustering percentage. Although the correct clustering percentage of the documents is used to measure the relevance of our arguing expressions' clustering algorithm, the objective of our work is not to group the documents but to accurately group the arguing expressions (topic-viewpoints pairs) into their viewpoints. We compare the obtained results with those of a simple lexicon-based baseline document clustering method, as well as a topic modeling-based method.

Baseline Method. The baseline consists of clustering the documents using a polarity lexicon, the subjectivity lexicon in the Multi-Perspective Question Answering (MPQA) opinion corpus[11] [29, 30], into two, positive and negative, classes. The positive and negative classes are, in this case, assimilated to two different viewpoints. The subjectivity lexicon contains a list of 8222 words or clues. The majority of the lexicon was collected from MPQA's English news documents, extracted from U.S and International sources between June 2001 and May 2002. Many of the 10 major discussed topics are controversial, e.g., U.S. holding prisoners in Guantanamo Bay, reaction to U.S. State Department report on human rights, 2002 presidential election in Zimbabwe, Israeli settlements in Gaza and West Bank, relations between Taiwan and China and presidential coup in Venezuela. Each word in the lexicon is either labeled as strongly subjective, i.e. often used as a subjective (opinionated) word in most contexts, or weakly subjective, i.e. only have certain subjective usages. We select a subset of the lexicon which contain the words labeled as having a positive or negative prior polarity. In order to classify a document, we , first, extract the words having a prior polarity (positive or negative) in the lexicon. Second, we assign a score of 1 and -1 to weakly subjective positive and negative clues, respectively. We assign a greater score of 2 and -2 to strong subjective positive and negative clues, respectively. Finally, we sum up the scores of extracted words. A document with positive or negative score is clustered in a positive or negative cluster, respectively. A document with a score of zero is discarded. We choose the matching between the positive or negative label and the correct viewpoint label that provides the best correct clustering percentage.

Joint Viewpoint Topic Model. The Joint Viewpoint Topic Model (JVT) (See Figure 1d) is a modified version of the JTV graphical model, where the topic variable z is dependent on the viewpoint variable v, instead of the opposite in JTV. This scheme results in viewpoint-topic distributions automatically clustered according to the viewpoint. JVT graphical model is similar to the Joint Sentiment Topic model (JST), by Lin et al. , where the sentiment variable is replaced with the viewpoint variable. However, JVT does not use any semi-supervision like JST, i.e., a list of polarity words to initialize the model parameters. The comparison with JVT, which automatically cluster the viewpoint-topic dimensions into viewpoints, will allow the analysis of the contribution of the proposed constrained clustering algorithm when used with JTV. The parameters are set as the

[11]http://mpqa.cs.pitt.edu/

following: α is equal to $50/K$, where K is the number of topics; β and γ are both set to 0.01. The CCP is computed out of the JVT's label assignments of viewpoints for each word in a document. Similarly to JVT+constrained clustering algorithm, a document is assigned to its majority label. If there is parity, it is discarded. The best matching between JVT's viewpoints labels and correct labels is hold.

Figure 4 presents six different boxplots, each corresponding to one of our six datasets (Table 3). For each dataset, we perform a uniform under-sampling in order to have a balanced number of opposed viewpoints documents. Each plot contains two boxes, corresponding to the distribution of correct clustering percentage (CCP) over 20 runs, of our combination JTV+constrained clustering algorithm and the JVT (both methods are not deterministic). For each dataset plot, the reported CCPs are the best values obtained for a particular number of topics K, which is set for both topic models JTV and JVT. Different number of topics were tried, $K = 1..10$. The lower and upper edges of the boxes represent the lower and upper quartiles, respectively. The lower and upper whiskers' ends denote the minimum and maximum reached CCP with a particular model. The red line inside each box corresponds to the median value of CCP. The plots also contain a horizontal green line representing the correct clustering percentage of the deterministic lexicon-based baseline method.

We notice that, for all six datasets, the combination JTV+constrained clustering algorithm produces a greater median CCP than the baseline CCP value, despite the fact that JTV+constrained clustering is an unsupervised method, while the baseline is lexicon-based. On the other hand, our method has a better median CCP value than JVT in 4 out of 6 datasets: OC, AW, GM1 and IP1. On the two remaining datasets, GM2 and IP2, the performances of JTV+constrained clustering and JVT are comparable. Separate Experiments on GM2, which we do not report in this paper, show a week accuracy of documents classification when using a supervised algorithm like the support vector machine. This may indicate the difficulty of classifying/clustering the documents of this particular dataset. IP2 dataset contains several interview documents that constitute a different structure from IP1 (editorials) or other debate site datasets. Interviews questions, which often do not denote any stance, are included in the dataset. This can explain the comparable percentage of JTV+constrained and JVT. JVT performs poorly on the remaining survey (OC), debate site (AW,GM1) and editorial (IP1) datasets: the median value is similar or less than that of the baseline CCP.

8. Qualitative Evaluation

We perform a simultaneous qualitative analysis of the generated topic-viewpoint pairs (i.e., arguing expressions) by the JTV model and their clustering (Section 5) according to the viewpoint they convey via Algorithm 1. The analysis is illustrated using the ObamaCare data set. Table 4 presents an example of the result output produced by the clustering component which uses the inferred topic-viewpoint pairs as input. The number of topics and the number of viewpoints (clusters) are set to $K = 5$ and $L = 2$ respectively. Each one of these clusters is represented by a collection of topic-viewpoint pairs automatically generated and assigned to it. Each topic-viewpoint in a given cluster (e.g., Topic 1-Viewpoint 1) is represented by the set of top terms or keywords. The terms are sorted in descending order (from left to right) according to their probabilities. We use these keywords for each topic-viewpoint pair to query the original data. The

Table 4: An example of a final output of the Constrained Clustering algorithm using the JTV's generated topics-viewpoints from Obamacare data set as input.

<table>
<tr><td colspan="8">Viewpoint 1</td></tr>
<tr><td>Topic 1</td><td>keywords</td><td>health</td><td>coverage</td><td>medicine</td><td>affordable</td><td>access</td><td>preexisting</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">broadening healthcare coverage and making it more affordable and addresses preexisting conditions</td></tr>
<tr><td>Topic 2</td><td>keywords</td><td>people</td><td>pay</td><td>insurance</td><td>uninsured</td><td>quality</td><td>dont_have</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">there are several people who donothave healthcare (...) the cost of the care that the uninsured receive in the emergency room is higher than say preventive care that they would otherwise receive if they had insurance</td></tr>
<tr><td>Topic 3</td><td>keywords</td><td>healthcare</td><td>system</td><td>country</td><td>world</td><td>free</td><td>provide</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">The healthcare system in our country is an abomination</td></tr>
<tr><td>Topic 4</td><td>keywords</td><td>people</td><td>cant_afford</td><td>change</td><td>children</td><td>dont_have</td><td>poor</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">Because a lot of people donthave healthcare and cantafford it</td></tr>
<tr><td>Topic 5</td><td>keywords</td><td>insurance</td><td>health</td><td>companies</td><td>dont_have</td><td>prices</td><td>reason</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">(...) even with health insurance you would never be covered completely and you will have health insurance companies accepting or rejecting a claim</td></tr>
<tr><td colspan="8">Viewpoint 2</td></tr>
<tr><td>Topic 1</td><td>keywords</td><td>healthcare</td><td>work</td><td>medicine</td><td>bill</td><td>dont_know</td><td>plan</td></tr>
<tr><td>Oppose</td><td>excerpt</td><td colspan="6">going to turn into another healthcare plan obama needs to put people back to work before they get healthcare</td></tr>
<tr><td>Topic 2</td><td>keywords</td><td>good</td><td>economy</td><td>dont_think</td><td>run</td><td>time</td><td>social</td></tr>
<tr><td>Support</td><td>excerpt</td><td colspan="6">I think social justice very good for the economy</td></tr>
<tr><td>Topic 3</td><td>keywords</td><td>money</td><td>expensive</td><td>make</td><td>doctor</td><td>debt</td><td>save</td></tr>
<tr><td>Oppose</td><td>excerpt</td><td colspan="6">It's ridiculously expensive, it's not going to save our everyday consumer any money (...) put us further and futher in debt</td></tr>
<tr><td>Topic 4</td><td>keywords</td><td>cost</td><td>government</td><td>control</td><td>increase</td><td>involved</td><td>private</td></tr>
<tr><td>Oppose</td><td>excerpt</td><td colspan="6">(...)puts it in the hands of the government instead of the hands of the private sector and it increases the cost to everybody</td></tr>
<tr><td>Topic 5</td><td>keywords</td><td>dont_think</td><td>dont_want</td><td>dollars</td><td>socialized</td><td>abortion</td><td>problem</td></tr>
<tr><td>Oppose</td><td>excerpt</td><td colspan="6">I don't want my tax dollars paying for abortion</td></tr>
</table>

search algorithm is implemented using an inverted index on the documents and their contained sentences. The output is the set of sentences from the document containing the maximum number of the query terms. The document which contain matching terms with higher probabilities has higher priority. Excerpts from the results are displayed in Table 4 for each topic-viewpoint pair. The result is not the arguing snippet summary which is not in the scope of this paper anyway. However, it could lay the ground to a generative short summary task of returned documents, which we would call the arguing expressions snippet summary. In order to assess the viewpoint coherence of clustered arguing expressions, we display the original label, "support" or "oppose", of the document from which the response sentences are extracted.

In Table 4, we, first, observe that most topic-viewpoint pairs corresponding to the same viewpoint, are conveying the same stance. Indeed, for each of the two viewpoints, most of the extracted sentences belong to documents having the same label. For all the topics of Viewpoint 1, the sentences belong to documents which are originally labeled as supporting the reform. For 4 out 5 topics in Viewpoint 2, sentences are labeled as opposing the reform. Thus, each viewpoint contains coherent topics denoting the same implicit stance. Second, the majority labels, in each viewpoint, are divergent which confirms that the modeling was able to distinguish the arguing expressions of the two opposed stances in that case. Third, most of the topics of the arguing expressions (or topic-viewpoint) in Table 4 are similar to those in the ground truth summary of the corpus (Table 2). For instance, Topic 5-Viewpoint 1 corresponds to "Don't trust insurance companies". Topic 4-Viewpoint 1 is similar to "people need health insurance / many uninsured". Topic 4-Viewpoint 2 can be assimilated to "Against big government involvement" or "government should not be involved in healthcare". Similarly, other matchings with the original summary exist in the remaining topic-viewpoint dimension.

Some topic-viewpoint may have incoherence, like Topic 2-Viewpoint 2, where a query with the terms "good", "economy" and "social" results in the extraction of a sentence with an original label of support. This is different from the label of other topic-viewpoint pairs in the same viewpoint. The reason may be the use of similar co-occurring lexicon to express opposing arguing viewpoint. For example, querying the original data with "run and social", words from Topic 2-Viewpoint 2, returns sentences extracted from a document with an opposing label: "*they can't seem to run anything correctly (...) everything's going broke, post office, social security*". These topicality incoherences exist even when we reduce or increase the number of topics. They need to be addressed in our future work.

9. Conclusion and Future Work

We suggested a fine grained probabilistic framework for improving the quality of opinion mining from different types of contention texts. We proposed a Joint Topic Viewpoint model (JTV) for the unsupervised detection of arguing expressions. Unlike common approaches, the proposed model focuses on arguing expressions that are implicitly described in unstructured text according to the latent topics they discuss and the implicit viewpoints they voice. We also implemented a constrained clustering algorithm which gets as input the learned topic-viewpoint pairs from JTV and group them according to their voiced viewpoint. The qualitative and quantitative assessments of the model's output show a good capacity of JTV in handling different contentious issues when compared to similar models. Moreover, analysis of the experimental results shows the effectiveness of the proposed model to automatically detect recurrent and relevant patterns of arguing expressions.

JTV assumes that each topic is discussed with different proportions according to the endorsed viewpoint. Some topics may be specific to only one particular viewpoint. In this case, the corresponding generated topic-viewpoint pairs can be redundant or contain incoherent topical information. This would later mislead the arguing expression clustering task. Future work should relax this assumption in order to enhance the topicality and viewpoint coherence of extracted topic-viewpoint pairs, as well as the arguing phrases. Moreover, automatically finding the optimal numbers of topics and viewpoint remains an open problem. Extention of JTV based on Nonparametric Bayesian models, e.g., Hierarchical Dirichlet Processes [31], can be considered.

In Table 4, we present an example of the final output of the proposed method. The keywords of each arguing expression are used to query the original data set and automatically retrieve the sentences. Although we show the relevance of extracted sentences, and their similarity with the topics of original summary in Table 2, they cannot form a good snippets summary that conveys the meaning of an arguing expression in few words. Future study needs to focus on the generation of these snippets given the retrieved phrases or sentences. Moreover, the coherence of generated arguing expressions terms, and the summaries, needs to be assessed via a human-oriented evaluation. The reference summaries as a ground truth of the used contentious corpora, or the issues themselves, have to be created by human experts for the automatic summary evaluation.

References

[1] J. M. Jones, In u.s., 45% favor, 48% oppose obama healthcare plan (Mar. 2010). URL `http://www.gallup.com/poll/126521/favor-oppose-obama-healthcare-plan.aspx`

[2] F. H. van Eemereen, Crucial Concepts in Argumentation Theory, Amsterdam University Press, 2001.

[3] E. Cabrio, S. Villata, A natural language bipolar argumentation approach to support users in online debate interactions, Argument & Computation 4 (3) (2013) 209–230.

[4] S.-M. Kim, E. H. Hovy, Crystal: Analyzing predictive opinions on the web., in: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007, pp. 1056–1064.

[5] W.-H. Lin, T. Wilson, J. Wiebe, A. Hauptmann, Which side are you on?: Identifying perspectives at the document and sentence levels, in: Proceedings of the Tenth Conference on Computational Natural Language Learning, 2006, pp. 109–116.

[6] S. Park, K. Lee, J. Song, Contrasting opposing views of news articles on contentious issues, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, 2011, pp. 340–349.

[7] M. Thomas, B. Pang, L. Lee, Get out the vote: Determining support or opposition from congressional floor-debate transcripts, in: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, 2006, pp. 327–335.

[8] S. Somasundaran, J. Wiebe, Recognizing stances in ideological on-line debates, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, 2010, pp. 116–124.

[9] Y. Jo, A. H. Oh, Aspect and sentiment unification model for online review analysis, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, 2011, pp. 815–824.

[10] I. Titov, R. McDonald, Modeling online reviews with multi-grain topic models, in: Proceedings of the 17th International Conference on World Wide Web, 2008, pp. 111–120.

[11] W. X. Zhao, J. Jiang, H. Yan, X. Li, Jointly modeling aspects and opinions with a maxent-lda hybrid, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010, pp. 56–65.

[12] C. Lin, Y. He, Joint sentiment/topic model for sentiment analysis, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 375–384.

[13] W.-H. Lin, E. Xing, A. Hauptmann, A joint topic and perspective model for ideological discourse, in: W. Daelemans, B. Goethals, K. Morik (Eds.), Machine Learning and Knowledge Discovery in Databases, Vol. 5212, Springer Berlin Heidelberg, 2008, pp. 17–32.

[14] A. Mukherjee, B. Liu, Mining contentions from discussions and debates, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012, pp. 841–849.

[15] A. Mukherjee, B. Liu, Discovering user interactions in ideological discussions, in: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013, pp. 671–681.

[16] M. Qiu, J. Jiang, A latent variable model for viewpoint discovery from threaded forum posts, in: Proceedings of NAACL-HLT, 2013, pp. 1031–1040.

[17] Y. Fang, L. Si, N. Somasundaram, Z. Yu, Mining contrastive opinions on political texts using cross-perspective topic model, in: Proceedings of the fifth ACM international conference on Web search and data mining, 2012, pp. 63–72.

[18] S. Gottipati, M. Qiu, Y. Sim, J. Jiang, N. A. Smith, Learning topics and positions from debatepedia, in: Proceedings of Conference on Empirical Methods in Natural Language Processing, 2013.

[19] M. J. Paul, C. Zhai, R. Girju, Summarizing contrastive viewpoints in opinionated text, in: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010, pp. 66–76.

[20] G. Erkan, D. R. Radev, Lexrank: Graph-based lexical centrality as salience in text summarization, J. Artif. Intell. Res.(JAIR) 22 (1) (2004) 457–479.

[21] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, Journal of Machine Learning Research 3 (2003) 993–1022.

[22] T. L. Griffiths, M. Steyvers, Finding scientific topics, Proceedings of the National Academy of Sciences of the United States of America 101 (1) (2004) 5228–5235.

[23] S. Basu, I. Davidson, K. Wagstaff, Constrained Clustering: Advances in Algorithms, Theory, and Applications, 1st Edition, Chapman & Hall/CRC, 2008.

[24] K. Wagstaff, C. Cardie, S. Rogers, S. Schrödl, Constrained k-means clustering with background knowledge, in: Proceedings of the Eighteenth International Conference on Machine Learning, 2001, pp. 577–584.

[25] G. Heinrich, Parameter estimation for text analysis, Tech. rep., Fraunhofer IGD (September 2009).

[26] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[27] M. J. Paul, R. Girju, A two-dimensional topic-aspect model for discovering multi-faceted topics., in: Proceedings of AAAI, 2010.

[28] M. Steyvers, T. Griffiths, Probabilistic topic models, Handbook of Latent Semantic Analysis 427 (7) (2007) 424–440.

[29] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing, EMNLP '03, Association for Computational Linguistics, Stroudsburg, PA, USA, 2003, pp. 105–112.

[30] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, HLT '05, Association for Computational Linguistics, Stroudsburg, PA, USA, 2005, pp. 347–354.

[31] Y. W. Teh, M. I. Jordan, M. J. Beal, D. M. Blei, Hierarchical dirichlet processes, Journal of the American Statistical Association 101 (476) (2006) 1566–1581.