Unsupervised Model for Topic Viewpoint Discovery in Online Debates Leveraging Author Interactions

Amine Trabelsi and Osmar R. Zaïane
Department of Computing Science, University of Alberta
Edmonton, Alberta Canada
{atrabels, zaiane}@ualberta.ca

Abstract

Online debate forums provide a valuable resource for textual discussions about controversial social and political issues. Discovering the viewpoints and their discourse or arguments from such resources is important for policy and decision makers. In order to detect the stance, most of the existing methods rely on expensively obtained human annotations and propose supervised solutions. In this work, we introduce a purely unsupervised Author Interaction Topic Viewpoint model (AITV) for viewpoint identification at the post and the discourse levels. The model favors "heterophily" over "homophily" when encoding the nature of the authors' interactions in online debates. It assumes that the difference in viewpoints breeds interactions, unlike similar studies based on social network analysis, which hypothesize that similar viewpoints encourage interactions. We evaluate the model's viewpoint identification and clustering accuracies at the author and post levels. Experiments are held on six corpora about four different controversial issues, extracted from two online debate forums. AITV's results show a better performance in terms of viewpoint identification at the post level than the state-of-the-art supervised methods in terms of stance prediction, even though it is unsupervised. It also outperforms a recently proposed topic model for viewpoint discovery in social networks and achieves close results to a weakly guided unsupervised method in terms of author level viewpoint identification. Our results highlight the importance of encoding "heterophily" for purely unsupervised viewpoint identification in the context of online debates. We also carry out a brief qualitative evaluation of the discourse modeling in terms of Topic-Viewpoint word clusters. AITV shows encouraging results suggesting an accurate discovery of the viewpoints and topics' discourses.

Introduction

Research on people's viewpoints, ideologies, and antagonistic relationships is gaining interest thanks to the emergence of social media and online forums as accessible tools to express opinion on different political and social issues. Online debate forums, specifically, provide a valuable resource for textual discussions about contentious issues. Contentious issues are controversial topics or divisive entities, e.g. legalization of abortion, same-sex marriage and Donald Trump, that usually engender opposing stances or viewpoints (e.g. supporting or opposing same-sex marriage). Forum users usually write posts to defend their standpoint using persuasion, reasons or arguments. Such posts correspond to what we describe as contentious documents (Trabelsi and Zaiane 2014a). Decision makers, politicians or a lay person seeking information to develop an opinion or to make a decision related to a contentious issue need to go through many of the existing posts on the subject. They need an automatic tool to help them overcome the overload of data and provide a contrasting overview of the main viewpoints and reasons given by opposed sides. However, reaching this objective supposes the ability of the tool to accurately identify the viewpoints at the post and/or author levels, as well as the capacity to detect the relevant discourse used to express distinct and recurrent arguing or reasoning themes. In this work, given online forum posts about a contentious issue, we study the problem of unsupervised identification and clustering of the viewpoints at the post level, as well as, the discovery of the contrastive discourse pertaining to each of the viewpoints.

Recent research on stance detection suggests that applying sentiment analysis techniques on contentious documents is not sufficient to produce an effective solution to the problem (Hasan and Ng 2013; Walker et al. 2012b). Indeed, Mohammad, Sobhani, and Kiritchenko(2017) show that both positive and negative lexicons are used, in contentious text, to express the same stance. Moreover, the stance can be implicitly conveyed through reasons and arguments, and not necessarily expressed through polarity sentiment words. Furthermore, challenges to accurate viewpoint detection can arise because of the dialogic nature of online debate (Hasan and Ng 2014; Boltužić and Šnajder 2014). The unstructured and colloquial language that is used makes it "noisy", i.e., containing non-argumentative portions and irrelevant dialogs, which misleads the model. The similarity in words' usage between authors holding divergent viewpoints leads to clustering errors when the model is entirely based on textual features. This is often the case when a post rephrases the opposing side's premise while attacking it or when asking a rhetorical question like "what makes a fetus not human ?". Fetus not human is usually a discourse pertaining to those who support Abortion, but here the same vocabulary is used to express the opposite stance. It has been shown that complementary features like the nature of the authors' interactions at post level (e.g., rebuttal, not rebuttal) can enhance pure text-based approaches in viewpoint distinguishing (Walker et al. 2012b).

In this work, we propose a purely unsupervised Author Interaction Topic Viewpoint model (AITV) for viewpoint discovery at the post and the discourse levels. AITV jointly models the textual content and the interactions between the authors in terms of replies. The model favors "heterophily" over "homophily" when encoding the nature of the authors' interactions in online debates. In this context, "heterophily" means that the difference in viewpoints breeds interactions, unlike similar studies based on social network analysis, which hypothesize that similar viewpoints encourage interactions (Thonet et al. 2017). Thus, "heterophily", here, does not mean the tendency to construct friendship groups with diverse people but the tendency to reply to opposed viewpoint author. In that regard, our assumption is similar to that of the supervision-based methods of Walker et al. (2012b) and Hasan and Ng (2013).

AITV is able to produce: (1) viewpoint assignments for each post; (2) Topic-Viewpoint word distributions denoting "arguing expressions" (Trabelsi and Zaiane 2014b) for each topic and viewpoint. Experiments are held on six corpora about four different controversial issues, extracted from two online debate forums: *4Forums.com* and *CreateDebate.com*. Given the viewpoints' assignments for each post, we evaluate the model's viewpoint identification at the post level first. Viewpoints' assignments for each post are later aggregated to evaluate the author level clustering. AITV's results show a better performance in terms of viewpoint identification at the post level than the state-of-the-art supervised methods in terms of stance prediction, even though it is unsupervised. It also outperforms the recently proposed topic model for viewpoint discovery in social networks (Thonet et al. 2017) and achieves close results to a weakly guided unsupervised method in terms of author level viewpoint identification and clustering. We also carry out a brief qualitative evaluation of the discourse modeling in terms of Topic-Viewpoint word dimensions. We use one corpus, the Abortion data set, as a case study. AITV shows promising characteristics that would allow to avoid some of the discourse challenges mentioned previously and simultaneously accurately distinguish the viewpoints and topics. Our contributions consist of:

- an unsupervised model to detect the viewpoints of the posts which leverages the content and the reply information about the authors (who is replying to whom) and which assumes "heterophily";
- quantitative and qualitative evaluations against supervised state-of-the-art and recent unsupervised methods that denote an accurate learning of the viewpoints at the post and the discourse levels;
- a discussion about author level viewpoint clustering and the limitations of the approach.

Related Work

The studies on viewpoint discovery or stance prediction differ mainly in terms of the type of the social media data that they use (e.g., Twitter, Online Debates), the features that they exploit (e.g., text, authors interactions, disagreement), the targeted task (e.g., post or author level stance prediction, viewpoints' discourse discovery) and the applied learning methods (e.g., supervised or unsupervised). It is important to mention that the Natural Language Processing community usually employs the word stance while the text mining community uses the word viewpoint often to express a political or ideological stand over an issue. We use both words interchangeably in this paper.

The work on supervised methods for stance prediction has recently seen a growing interest (Walker et al. 2012a). The Semantic Evaluation series 2016 (SemEval-16) propose a shared task for stance detection in Twitter (Mohammad et al. 2016). Sobhani, Inkpen, and Matwin (2015) tackle the stance classification for news comments using arguments features that are extracted using Topic Modeling. Hasan and Ng (2013) identify the stance at the post and the sentence levels of online debates corpora. They construct a rich feature set of linguistic and semantic features, and encourage opposing stance between successive posts. Sridhar et al. (2015) model disagreement and collectively predict stances at the post and the author levels. They try different modeling approaches on online debate corpora. The approach that is based on Probabilistic Soft Logic (PSL) and which models disagreement achieves the overall best performance. In our paper, we later (see Experiments Section) compare our results in terms of post level viewpoint identification to the reported results of this state-of-the-art supervised method, on the same debate corpora that it uses. All these described methods extensively rely on human annotations, which are expensive to obtain, and on supervision which does not guarantee scaling to different domains and types of data.

Another line of work focuses on weakly guided or pure unsupervised methods aiming to detect the contrastive discourse in different viewpoints and/or to identify the viewpoints of the posts and the authors. We mention the studies on Topic-Viewpoint modeling, which are based on Latent Dirichelet Allocation (LDA) (Blei, Ng, and Jordan 2003). They are applied on different sources of contentious documents: opinion polls, editorials (Paul, Zhai, and Girju 2010), online forums (Trabelsi and Zaïane 2016), Twitter (Joshi, Bhattacharyya, and Carman 2016), and parliamentary debates (Vilares and He 2017). They hypothesize the existence of underlying topic and viewpoint variables that influence the author's word choice when writing about a controversial issue. The viewpoint variable is also called stance, perspective or argument variable in different studies. The objective is mainly to extract relevant Topic-Viewpoint distributions of words that express the different viewpoints separately, along with their respective discussed topics. Thus, most of the Topic-Viewpoint models do not necessarily attempt to cluster the posts or their authors. However, Qiu and Jiang (2013) exploit the authors' interactions in threaded discussion forums to discover stances of posts and cluster authors with different viewpoints. Similarly, our work leverages the interactions between the authors in online forum debates to determine the opposed viewpoints of the posts and the authors. Conversely, our method jointly models the Topic-Viewpoint distribution to uncover, the viewpoints' discourse. The Topic-Viewpoint word distributions are not modeled in Qiu and Jiang (2013)'s work. Moreover, the polarity of the interactions between the authors, positive or negative, is guided and determined using lexicon-based methods. In our work, we don't exploit any external or specific sentiment lexicon to determine the type of interactions between the authors, which makes our approach purely unsupervised, independent of any external knowledge.

Recently, Thonet et al. (2017) propose different extended versions of Social Network-LDA (SN-LDA) (Sachan et al. 2014) that model the viewpoint discovery in social media: the Social Network Viewpoint Discovery Models (SNVDMs). Similar to our work, SNVDMs jointly model topic, viewpoint and discourse. One of their main objectives is to accurately determine the author's viewpoint. Conversely, they assume that the "homophily" pehenomenon is governing the authors' interactions. SNVDMs are experimented on political Twitter data, and consider a network of replies and retweets interactions. The SNVDM-GPU, the version based on *Generalized Polya Urn* sampling, is performing the best among all degenerate versions. It aims to overcome the sparsity of interactions.

Another recent work (Dong et al. 2017) focuses on a predicting the author's stance and providing insights about the viewpoints' discourse. The authors propose a weakly guided Stance-based Text Generative Model with Link Regularization (STML) which leverages the text content as well as the authors' interactions in news comments and online debates. The weak guidance consists of estimating the signs of interactions, i.e., agreement or disagreement, using heuristics rules like the number of a discussion's turns, the presence of agreement or disagreement signals.

In our paper, similar to the recent research mentioned above, we jointly utilize content and interactions in viewpoint's discourse discovery and posts/authors clustering. Our method is, however, purely unsupervised, i.e., does not require external knowledge or weak guidance to infer the nature of interactions between the authors. Moreover, it does not assume "homophily" but "heterophily" when modeling the interactions in online debate.

Author Interaction Topic Viewpoint Model

As mentioned in the previous sections, the Author Interaction Topic Viewpoint (AITV) Model is a generative Topic-Viewpoint model. Topic-Viewpoint models are extensions of LDA (Blei, Ng, and Jordan 2003). They are mainly data-driven approaches which reduce the documents into topic-viewpoint dimensions. A Topic-Viewpoint pair k-l is a probability distribution over unigram words. The unigrams with top probabilities characterize the used vocabulary when talking about a specific topic k while expressing a particular viewpoint l at the same time.

AITV takes as input the posts or documents, and the information about author-reply interactions in an online debate forum. The objective is to: (1) assign a viewpoint to each post and; (2) assign a topic-viewpoint label to each occurrence of the unigram words. This would help to cluster

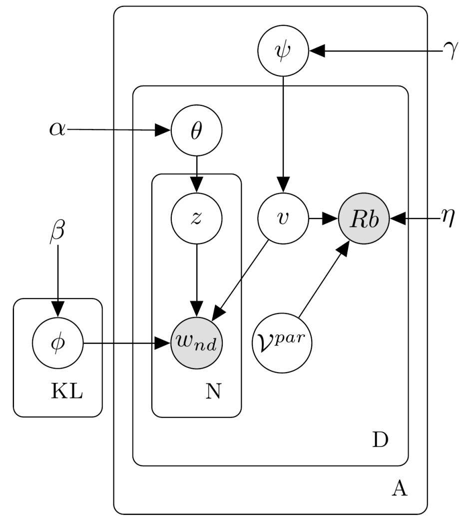

Figure 1: Plate Notation of AITV model

them into Topic-Viewpoint classes. Prior to the topic modeling step, we pre-process the online debate posts. We remove identical portions of text in replying posts. These can be assimilated to references or citations of previous posts text. We remove stop and rare words. We consider working with the stemmed version of the words.

Generative Process

AITV assumes that A authors participate in a forum debate about a particular issue. Each author a writes D_a posts. Each post d_a contains N_{da} words. Each term w_{nd} in a document belongs to the corpus vocabulary of distinct terms of size W. In addition, we assume that we have the information about whether a post replies to a previous post or not. Let K be the total number of topics and L be the total number of viewpoints, in our case set to 2. Let θ_{da} denote the probability distribution of K topics under a post d_a; ψ_a be the probability distributions of L viewpoints for an author a; ϕ_{kl} be the multinomial probability distribution over words associated with a topic k and a viewpoint l. The generative process of a post according to the AITV model (see Figure 1) is described below.

An author a chooses a viewpoint v_{da} from the distribution ψ_a. For each word w_{nd} in the post, the author draws a topic z_{nd} from θ_{da}, then, samples each word w_{nd} from the topic-viewpoint distribution corresponding to topic z_{nd} and viewpoint v_{da}, $\phi_{z_{nd}v_{da}}$.

Note that, in what follows, we refer to a current post with index id and to a current word with index i. When the current post is a reply to a previous post by a different author, it may contain a rebuttal or it may not. If the reply attacks the previous author then the rebuttal variable Rb_{id} is set to 1 else if it supports, the rebuttal takes 0. We define the **parent posts** of a current post as all the posts of the author who the current post is replying to. Similarly, the **child posts** of a current post are all the posts replying to the author of the current post. We assume that the probability of a rebuttal

$Rb_{id} = 1$ depends on the degree of opposition between the viewpoint v_{id} of the current post and the viewpoints $\mathcal{V}_{id}^{par}$ of its parent posts as the following:

$$p(Rb_{id} = 1|v_{id}, \mathcal{V}_{id}^{par}) = \frac{\sum_{l'}^{\mathcal{V}_{id}^{par}} \mathbf{I}(v_{id} \neq l') + \eta}{|\mathcal{V}_{id}^{par}| + 2\eta}, \quad (1)$$

where $\mathbf{I}$(condition) equals 1 if the condition is true and η is a smoothing parameter. This modeling of authors interactions is similar to the users interactions setting presented in (Qiu and Jiang 2013).

Parameters Inference

For the inference of the model's parameters, we use the collapsed Gibbs sampling. For all our parameters, we set fixed symmetric Dirichlet priors. According to Figure 1, the Rb variable is observed. However, the true value of the rebuttal variable is unknown to us. We fix it to 1 to keep the framework fully unsupervised, instead of estimating the reply disagreement using methods based on lexicon polarity like in (Qiu and Jiang 2013). Setting $Rb = 1$ means that all replies of any post are rebuttals attacking all of the parent posts excluding the case when the author replies to his own post. This correspond to our "heterophily" assumption. It comes from the observation that the majority of the replies, in the debate forums framework, are intended to attack the previous proposition (Hasan and Ng 2013). This setting will affect the viewpoint sampling of the current post. The intuition is that, if an author is replying to a previous post, the algorithm is encouraged to sample a viewpoint which opposes the majority viewpoint of parent posts (Equation 1). Similarly, if the current post has some child posts, the algorithm is encouraged to sample a viewpoint opposing the children's prevalent stance. If both parent and child posts exist, the algorithm is encouraged to oppose both, creating some sort of adversarial environment when the prevalent viewpoints of parents and children are opposed. The derived sample equation of current post's viewpoint v_{id} given all the previous sampled assignments in the model $\vec{v}_{\neg id}$ is:

$$\begin{aligned} p(v_{id} = l|\vec{v}_{\neg id}, \vec{w}, \vec{Rb}) &\propto \\ n_{a,\neg id}^{(l)} + \gamma \times &\frac{\prod_{t}^{W_{id}} \prod_{j=0}^{n_{id}^{(t)}-1} n_{l,\neg id}^{(t)} + j + \beta}{\prod_{j=0}^{n_{id}-1} n_{l,\neg id}^{(\cdot)} + W\beta + j} \\ &\times p(Rb_{id} = 1|v_{id}, \mathcal{V}_{id}^{par}) \\ &\times \prod_{c|v_{id} \in \mathcal{V}_{c}^{par}} p(Rb_{c} = 1|v_{c}, \mathcal{V}_{c}^{par}). \quad (2) \end{aligned}$$

The count $n_{a,\neg id}^{(l)}$ is the number of times viewpoint l is assigned to author a's posts excluding the assignment of current post, indicated by $\neg id$; $n_{l,\neg id}^{(t)}$ is the number of times term t is assigned to viewpoint l in the corpus excluding assignments in current post; $n_{l,\neg id}^{(\cdot)}$ is the total number of words assigned to l; W_{id} is the set of vocabulary of words in post id; $n_{id}^{(t)}$ is the number of time word t occurs in the post. The third term of the multiplication in Equation 2 corresponds to Equation 1 and is applicable when the current post is a reply. The fourth term of the multiplication takes effect when the current post has child posts. It is a product over each child c according to Equation 1. It computes how much would the children's rebuttal be probable if the value of v_{id} is l. It is important to mention that during the implementation of the viewpoint sampling, we used few tricks that helped improving the model in terms of effectiveness and efficiency. First, we only consider as children the posts that are replying to the current post, instead of all the posts replying to the author of the current post. This enhances the efficiency for large datasets with keeping at least the same effectiveness of the original setting. Second, in order to make the Gibbs Sampling less variable to the random initializations, we set an automatic initialization process that helped stabilizing the model. The automatic initialization consists of offsetting terms 1 and 2 in Equation 2 for the initial 100 iterations. Thus, we only leverage the interactions, and not the text content. Third, following (Hasan and Ng 2014), we unify all the posts' viewpoint of a given author by assigning the majority label among them. This is done few iterations before stopping the Gibbs sampling.

Given the assignment of a viewpoint $v_{id} = l$, we also jointly sample the topic for each word i in post id, according to the following:

$$p(z_i = k|w_i = t, \vec{z}_{\neg i}, \vec{w}_{\neg i}, \vec{v}) \propto n_{id,\neg i}^{(k)} + \alpha \times \frac{n_{kl,\neg i}^{(t)} + \beta}{n_{kl,\neg i}^{(\cdot)} + W\beta}, \quad (3)$$

Here $n_{id,\neg i}^{(k)}$ is the number of times topic k is observed in document id, excluding the current word i; $n_{kl,\neg i}^{(t)}$ corresponds to the number of times the word t is assigned to topic-viewpoint kl excluding the current occurence; $n_{kl,\neg i}^{(\cdot)}$ is a summation of $n_{kl,\neg i}^{(t)}$ over all words.

After the convergence of the Gibbs algorithm, each post is assigned a viewpoint. Thus, we can cluster the post according to their assignments. Although the modeling suggests that an author may have different viewpoints, the viewpoint's unification trick mentioned above ensures that an author will have a unique viewpoint by the end of the sampling. Thus, the authors also can be clustered. Each word is assigned a topic and a viewpoint label. We exploit these labels to first create clusters, where each cluster corresponds to a topic-viewpoint value kl. It contains all the unigrams that are assigned to kl at least one time. Second, we rank the words inside each cluster according to their assignment frequencies.

Datasets

We evaluate the proposed model on six datasets about four different controversial issues, extracted from 4Forums.com (Abbott et al. 2016) and CreateDebate.com (Hasan and Ng 2014). Table 1 presents the datasets and their key statistics.

	4Forums		CreateDebate			
	Abortion	Gay Marriage	Gun Control	Abortion	Gay Rights	Obama
nb. posts	7795	6782	3653	1876	1363	962
nb. authors	333	294	274	506	368	277
% majority label posts	56.03	65.54	67.80	55.34	62.10	54.76
% reply posts	99.38	99.32	98.87	76.81	76.45	59.46
% replies btw. opposing stance posts	77.6	72.1	63.59	81.3	87.07	84.44

Table 1: Statistics on the six datasets used in experiments belonging to two online debate forums: 4Forums and CreateDebate.

	4Forums		CreateDebate			
	Abortion	Gay Marriage	Gun Control	Abortion	Gay Rights	Obama
AITV	92.0 ± 1.9	90.6 ± 0.3	70.5 ± 11.6	72.6 ± 10.1	79.3 ± 2.8	67.8 ± 9.9
PSL (Sridhar et al. 2015)	77.0 ± 8.9	80.5 ± 8.5	65.4 ± 8.3	66.8 ± 12.2	72.7 ± 8.9	63.5 ± 16.3

Table 2: Averages and standard deviations of post level viewpoint identification accuracy in percentage (AITV) and stance prediction accuracy in percentage (PSL)

The 4Forums datasets contain the ground truth stance labels at the author level, while the CreateDebate have annotated labels at the post level. In order to perform extrinsic clustering evaluation at both the post and author levels, we apply the author label for all of the corresponding posts when dealing with 4Forums datasets. For CreateDebate, we assign to each author the majority label of his/her corresponding posts (Sridhar et al. 2015). For all the datasets, Table 1 reports the percentage of rebuttals as the percentage of replies between authors of opposed stance labels among all the replies.

Experiments and Analysis

We conduct experiments in order to evaluate AITV's performance on *4Forums* and *CreateDebate* in terms of: (1) viewpoint identification at the post level, (2) viewpoint clustering at the author level, (3) text clustering and detection of Topic-Viewpoint word distribution.

Experiments Set Up

All the reported results of AITV in this section correspond to aggregation measures on 10 runs or repeats. The number of Topics K is 30 unless specified otherwise. The number of Viewpoint L is always set to 2. AITV hyperparameters are set as follows: $\alpha = 0.1$; $\beta = 1$; $\gamma = 1$; $\eta = 0.01$. The number of the Gibbs Sampling iterations is 1500. The words occurring less than 20 times are considered rare words and are removed.

Post Level Viewpoint Identification

Given AITV's output, which consists of post level viewpoint assignments, we compute an extrinsic viewpoint identification accuracy measure, given the fact that all of the used datasets contain ground-truth viewpoint labels. We choose the better alignment of output viewpoint labels with the ground truth, support/oppose class labels, and compute the percentage of posts that are "correctly clustered" as the viewpoint identification accuracy. We compare AITV's viewpoint identification results, on all corpora, against the state-of-the-art supervised method (Sridhar et al. 2015) (see Section Related Work). In Table 2, we report the average stance prediction accuracy of the best overall method in Sridhar et al.(2015)'s work. The method is based on PSL (Probabilistic Soft Logic). Its results are estimated on 5 repeats of 5-fold cross-validation. AITV's reported values are averaged over 10 repeats.

Table 2 shows that AITV clearly outperforms PSL on each of the datasets. This is achieved although it is a purely unsupervised method. We also notice that the best performances are recorded on the largest and highest connected datasets (see % reply posts in Table 1). Indeed, Abortion and Gay Marriage datasets of 4Forums reach 90%+ accuracies with low variances. The patterns in terms of the best and lowest accuracies over all the datasets are the same for both of the reported methods. We also observe that the datasets containing greater percentages of replies between posts of opposing stance, i.e. rebuttals, are not necessarily the ones for which AITV performs the best. This suggests that the adversarial setting of viewpoint sampling for AITV, with the help of the high number of connections, can properly distinguish communities. Thus, its performance is not just the consequence or the result of using the dataset that corresponds the most to the "heterophily" assumption.

We compare AITV's performance against its degenerate version "AITV-Rebuttal Known". The first objective of this experiment is to compare AITV to a close version to the weakly-guided work of Qiu and Jiang (2013) [1] (see also Related Work Section). The second objective is to evaluate the performance of AITV which supposes that the rebuttal information is unknown against a degenerate version that uses the ground truth about rebuttals. Finally, we want to compare against a version that does not implement the tricks discussed when sampling the viewpoints in the Section detailing AITV model. The AITV Rebuttal Known (AITV-RK) version, like Qiu and Jiang(2013), models background words and does not implement the three sampling tricks consisting of considering only the immediate child posts, the

[1]At the time of writing, the implementation of Qiu and Jiang (2013)'s work is not available publicly.

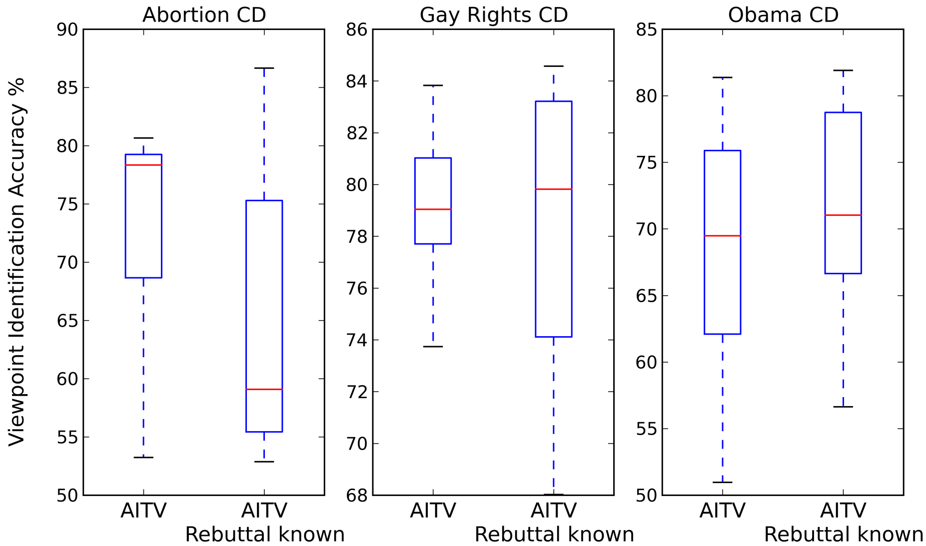

Figure 2: Boxplots of the post level viewpoint identification accuracies for AITV and AITV-RK, for CreatDebate.

	4Forums			CreateDebate		
	Abortion	Gay Marriage	Gun Control	Abortion	Gay Rights	Obama
AITV	70.4 ± 2.1	70.3 ± 1.4	57.5 ± 7.6	55.2 ± 3.3	60.4 ± 3.2	56.8 ± 4.1
SNVDM-GPU	52.2 ± 1.4	52.3 ± 1.8	54.1 ± 2.7	52.0 ± 1.8	52.8 ± 2.3	52.2 ± 1.7
STML (Dong et al. 2017)	75.6	68.6	66.3	-	-	-
PSL (Sridhar et al. 2015)	65.8 ± 4.4	77.1 ± 4.4	67.1 ± 5.4	67.4 ± 7.5	74.0 ± 5.3	63.0 ± 8.3

Table 3: Averages and standard deviations of author level viewpoint identification accuracy in percentage (AITV, SNVDM-GPU) and stance prediction accuracy in percentage (PSL, STML)

automatic initialization and the unification of the author's viewpoints. Qiu and Jiang(2013) determine the rebuttal between the authors using lexicon-based methods. AITV-RK goes further and uses the ground truth values of rebuttals which only exist for the CreateDebate datasets. Figure 2 presents the box-plots of the post level viewpoint identification accuracies for AITV and AITV-RK over 10 runs, for CreateDebate. We observe that when the rebuttal is known the difference is not significant in terms of median values. In fact, AITV has even a better median on Abortion issue. We also observe a slightly lower variance for AITV on the higher connected datasets. This may be due to the automatic initialization based on authors' interactions which helps in reducing the variance of the non-deterministic outputs, due to the Gibbs Sampling.

Author level Viewpoint Identification and Clustering

In this section, we compare author level viewpoint identification and clustering performances against another recently proposed Topic-Viewpoint unsupervised method on social network analysis, the SNVDM-GPU (Thonet et al. 2017) (See Section Related Work). SNVDM-GPU supposes "homophily" in reply and retweets interactions in Twitter. It only outputs author level viewpoint assignment. We apply it on our six datasets. SNVDM is run 10 times and default parameters are used with acquaintance $\tau = 10$. Also, we compare AITV to the recently introduced weakly unsupervised method STML (Dong et al. 2017) (See Section Related Work). The code of STML cannot be made available during the time of this paper's writing. Therefore, we only report the values on the CreateDebate datasets which are presented in the original paper. Table 3 contains the average viewpoint identification accuracies for AITV and SNVDM-GPU and the average stance prediction accuracies for STML and PSL. AITV outperforms its rival unsupervised method SNVDM, specifically for the datasets containing many interactions. It has also close to comparable performance with the weakly guided STML on Abortion and Gay Marriage. However, AITV's performance in this task remains far from that of the supervised PSL, except for Abortion on 4Forums. We notice a big drop in accuracies between the post level and the author level for AITV. We suspect that AITV is able to accurately detect the viewpoints for highly interactive authors, who reply a lot or get many replies, and thus account for a big portion of the total posts in the online debate. However, it has low accuracy when authors are non interactive. We further develop this point in the Discussion Section.

We evaluate the two unspervised Topic-Viewpoint clustering methods AITV and SNVDM with the BCubed F-

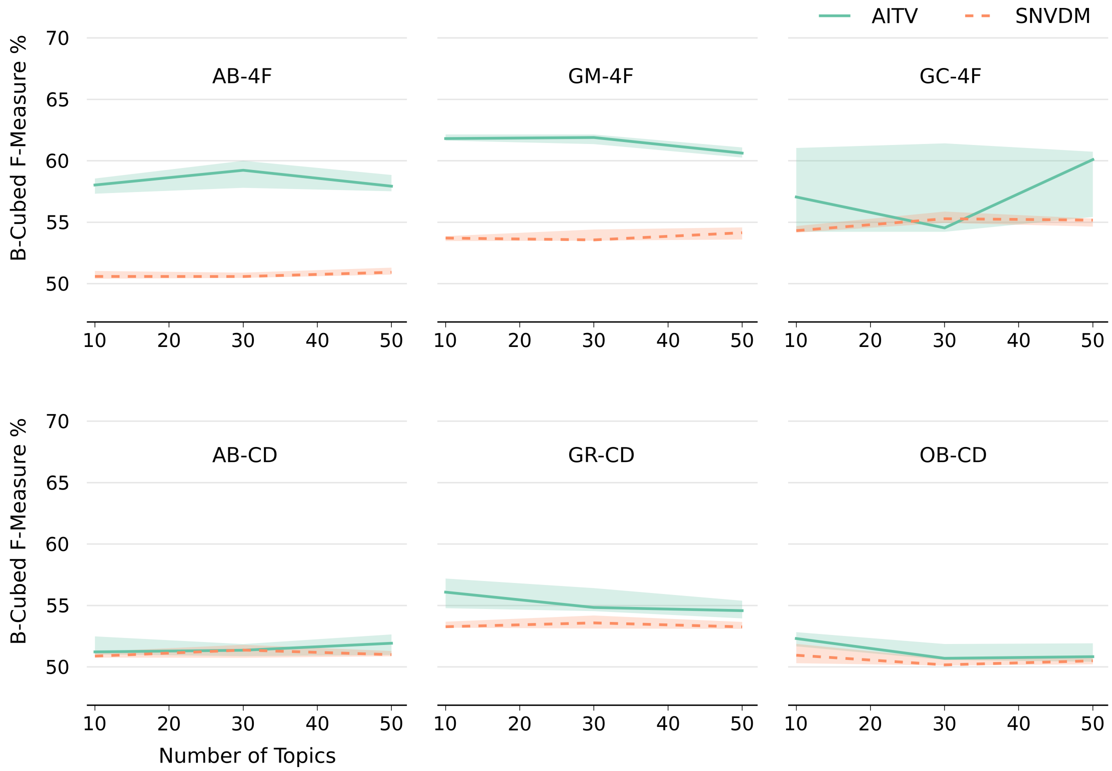

Figure 3: AITV and SNVDM-GPU median and quartile values of the Bcubed F-Measure for author level viewpoint clustering.

Measure. The BCubed F-Measure, which is based on B-Cubed recall and B-Cubed precision, satisfies the four essential criteria needed for a clustering quality measure (Han, Pei, and Kamber 2011). We run both models on a different number of topics: 10, 30 and 50 in order to check potential variations in performance. Figure 3 plots the AITV and SNVDM-GPU median and quartile values of the BCubed F-Measure for author level viewpoint clustering. The plot confirms the results found when evaluating the viewpoint accuracy about the better overall performance of AITV at the author level clustering comparing to SNVDM-GPU. It also emphasizes the problem of the high variance in AITV's performance for Gun Control dataset, which has the lowest percentage of rebuttals among its counterparts (see Table 1). Both models are pretty constant for all the other datasets and for different number of topics.

Topic-Viewpoints Words Clustering

We qualitatively evaluate the AITV's Topic-Viewpoint clustering of the unigram words. We consider the Abortion dataset as a case study. Topic-Viewpoint clusters are usually represented by the top frequent words. We assimilate those clusters to a representation of reasons or argument expressions about a specific topic of argumentation from a particular viewpoint (Trabelsi and Zaiane 2014b). The problem with using unigrams is that inferring the topic of the cluster is often not a straight forward task (See examples in Table 4). Moreover, in the context of controversial issues, the used vocabulary for different viewpoints may be very similar. This is one of the challenges described in the Introduction Section. For instance, we can observe that the top words of examples 1 and 5 in Table 4, which are related to opposed viewpoints, contain 4 common words out of 5. In order to perform a better evaluation of the discourse output of our model, we choose to use the unigrams pertaining to each Topic-Viewpoint and to query back the original datasets to retrieve a representative sentence. The sentence must belong to a post that is assigned the corresponding viewpoint according to AITV. The third column of Table 4 contains the result of this procedure for some selected Topic-Viewpoint clusters generated by AITV. We can observe that the sentences, corresponding to examples 1 and 5, shed light upon the nature of the viewpoint of the cluster. Although, clearly both sentences are discussing the topic of women's rights, the viewpoint of example 5 is claiming that right while sentence 1 is questioning it in the context of Abortion. A similar pattern can be seen in examples 2 and 6. However, the topic of argumentation is changing here to the fetus's rights. We can observe the change of the topics within the same viewpoint and the similarity of the themes at the inter-viewpoints level. This suggests that our AITV has been successful in distinguishing between topics and viewpoints discourses. We can also observe that the example sentence 4 corresponds to a rhetorical question. This may give insight on how to overcome the rhetorical discourse challenge.

	View 1: Oppose Legalization of Abortion	
	Topic-View Top 5 words	Sentence
1	not abort child woman don	Taking away the womans right to destroy her child is not about taking away her choice.
2	human fetus right dna live	The fetus is a living, human being, who has every right in the world.
3	human life begin cell person	IMO, life begins when a unique cell is created by the combination of a human egg and a human sperm.
4	kill not babi abort mother	If the court or parental unit is not allowed to interfere with abortion plans, does the mother have the right to kill the child?
		View 2: Support Legalization of Abortion
	Topic-View Top 5 words	Sentence
5	not abort child women pregnanc	It is my opinion that women should have the opportunity to stop a pregnancy they do not want, and not be forced to have a child.
6	right not woman fetus abort	The fetus has no rights to violate, but even if it did it's right to live would not allow it to use the woman's body against her will.
7	exist mental not fetus bodi	Before a fetus has a mental existence, it is just a growing human body - a thing, not a person.
8	kill abort not murder peopl	Therefore your Abortion is not murder.

Table 4: Clustered Viewpoints by AITV in terms of Topic-Viewpoint discourse dimensions (Top 5 words), along with the corresponding sentences, retrieved using the top words as query.

	4Forums		CreateDebate	
	wrongly clustered	correct. clustered	wrong. clustered	correct. clustered
% Non Interacting authors	4.0	2.12	42.95	29.30
% Interactions with same view per author	64.97	32.49	28.15	17.97
Median number of interactions per author	2.6	6.35	1.35	1.93

Table 5: Interactions statistics on wrongly and correctly clustered authors by AITV, averaged on the datasets of the two forums.

Conclusion and Discussion

In this paper, we present AITV, a purely unsupervised model, which jointly leverages the content and the interactions between the authors in online debates in order to detect the viewpoints at the post and author levels. The model also attempts to jointly discover the discourse of the viewpoints and their sub-topics. The quantitative and qualitative evaluations are held against one supervised state-of-the-art method and two recent unsupervised approaches. The results denote an accurate learning of the viewpoints at the post and the discourse levels. However, although the good performance of AITV against the recently proposed SNVDM at the author level clustering, it does not outperform neither the weakly-guided, nor the supervised method on this task. Moreover, we notice a significant drop in AITV's performance comparing to the post level task.

We discuss here some of the potential reasons pertaining to this drop. We average some interaction statistics, over the two forums' datasets, about the authors that were mis-clustered and correctly clustered by AITV, in Table 5. We consider any received or sent out reply as an interaction involving the author. We make two observations. The first is that mis-clustered authors on average interacted more often with the posts that have the same viewpoints, than the correctly clustered authors. This is valid for both forums. Moreover, this percentage is almost 65% for mis-clustered authors of 4Forums. These correspond to authors leaning towards "homophily". The second observation is that the percentages of low interactive authors and those with no interactions are also higher within the mis-clustered than within the correctly clustered over both forums. However, CreateDebate has significantly more non interactive authors than 4Forums. These represent on average 42.95% of mis-clustered authors comparing to 29.30% for correctly clustered. The mis-clustered authors of 4Forums interact rarely on average compared to the correct ones. Overcoming these limitations should be part of future work on unsupervised viewpoint identification. Future work may also include the application of similar models to AITV on Twitter mention networks. Indeed, Conover et al. (2011) observe that the users of opposed ideologies interacts at a much higher rate in the mention network comparing to retweet network. Automatic summarization of contentious issue, given the encouraging results of AITV, may also be explored.

References

Abbott, R.; Ecker, B.; Anand, P.; and Walker, M. A. 2016. Internet argument corpus 2.0: An sql schema for dialogic social media and the corpora to go with it. In *LREC*.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3:993–1022.

Boltužić, F., and Šnajder, J. 2014. Back up your stance: Recognizing arguments in online discussions. In *Proceedings of the First Workshop on Argumentation Mining*, 49–58. Baltimore, Maryland: Association for Computational Linguistics.

Conover, M.; Ratkiewicz, J.; Francisco, M.; Goncalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on twitter. In *International AAAI Conference on Web and Social Media*.

Dong, R.; Sun, Y.; Wang, L.; Gu, Y.; and Zhong, Y. 2017. Weakly-guided user stance prediction via joint modeling of content and social interaction. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, 1249–1258. New York, NY, USA: ACM.

Han, J.; Pei, J.; and Kamber, M. 2011. *Data mining: concepts and techniques*. Elsevier.

Hasan, K. S., and Ng, V. 2013. Stance classification of ideological debates: Data, models, features, and constraints. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 1348–1356. Nagoya, Japan: Asian Federation of Natural Language Processing.

Hasan, K. S., and Ng, V. 2014. Why are you taking this stance? identifying and classifying reasons in ideological debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 751–762. Doha, Qatar: Association for Computational Linguistics.

Joshi, A.; Bhattacharyya, P.; and Carman, M. 2016. Political issue extraction model: A novel hierarchical topic model that uses tweets by political and non-political authors. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 82–90. San Diego, California: Association for Computational Linguistics.

Mohammad, S.; Kiritchenko, S.; Sobhani, P.; Zhu, X.; and Cherry, C. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, 31–41. San Diego, California: Association for Computational Linguistics.

Mohammad, S. M.; Sobhani, P.; and Kiritchenko, S. 2017. Stance and sentiment in tweets. *ACM Trans. Internet Technol.* 17(3):26:1–26:23.

Paul, M.; Zhai, C.; and Girju, R. 2010. Summarizing contrastive viewpoints in opinionated text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, 66–76. Cambridge, MA: Association for Computational Linguistics.

Qiu, M., and Jiang, J. 2013. A latent variable model for viewpoint discovery from threaded forum posts. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1031–1040. Atlanta, Georgia: Association for Computational Linguistics.

Sachan, M.; Dubey, A.; Srivastava, S.; Xing, E. P.; and Hovy, E. 2014. Spatial compactness meets topical consistency: Jointly modeling links and content for community detection. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM '14, 503–512. New York, NY, USA: ACM.

Sobhani, P.; Inkpen, D.; and Matwin, S. 2015. From argumentation mining to stance classification. In *Proceedings of the 2nd Workshop on Argumentation Mining*, 67–77. Denver, CO: Association for Computational Linguistics.

Sridhar, D.; Foulds, J.; Huang, B.; Getoor, L.; and Walker, M. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 116–125. Beijing, China: Association for Computational Linguistics.

Thonet, T.; Cabanac, G.; Boughanem, M.; and Pinel-Sauvagnat, K. 2017. Users are known by the company they keep: Topic models for viewpoint discovery in social networks. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 87–96.

Trabelsi, A., and Zaiane, O. R. 2014a. Finding arguing expressions of divergent viewpoints in online debates. In *Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM)*, 35–43. Gothenburg, Sweden: Association for Computational Linguistics.

Trabelsi, A., and Zaiane, O. R. 2014b. Mining contentious documents using an unsupervised topic model based approach. In *Proceedings of the 2014 IEEE International Conference on Data Mining*, 550–559.

Trabelsi, A., and Zaïane, O. R. 2016. Mining contentious documents. *Knowledge and Information Systems* 48(3):537–560.

Vilares, D., and He, Y. 2017. Detecting perspectives in political debates. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1573–1582. Copenhagen, Denmark: Association for Computational Linguistics.

Walker, M. A.; Anand, P.; Abbott, R.; Tree, J. E. F.; Martell, C.; and King, J. 2012a. That is your evidence?: Classifying stance in online political debate. *Decision Support Systems* 53(4):719 – 729.

Walker, M. A.; Anand, P.; Abbott, R.; and Grant, R. 2012b. Stance classification using dialogic properties of persuasion. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL HLT '12, 592–596. Stroudsburg, PA, USA: Association for Computational Linguistics.