Staying Informed: Supervised and Semi-Supervised Multi-view Topical Analysis of Ideological Perspective

Amr Ahmed
School of Computer Science
Carnegie Mellon University
amahmed@cs.cmu.edu

Eric P. Xing
School of Computer Science
Carnegie Mellon University
epxing@cs.cmu.edu

Abstract

With the proliferation of user-generated articles over the web, it becomes imperative to develop automated methods that are aware of the ideological-bias implicit in a document collection. While there exist methods that can classify the ideological bias of a given document, little has been done toward understanding the nature of this bias on a topical-level. In this paper we address the problem of modeling ideological perspective on a topical level using a factored topic model. We develop efficient inference algorithms using Collapsed Gibbs sampling for posterior inference, and give various evaluations and illustrations of the utility of our model on various document collections with promising results. Finally we give a Metropolis-Hasting inference algorithm for a semi-supervised extension with decent results.

1 Introduction

With the avalanche of user-generated articles over the web, it is quite important to develop models that can recognize the ideological bias behind a given document, summarize where this bias is manifested on a topical level, and provide the user with alternate views that would help him/her staying informed about different perspectives. In this paper, we follow the notion of ideology as defines by Van Dijk in (Dijk, 1998) as "a set of general abstract beliefs commonly shared by a group of people." In other words, an ideology is a set of ideas that directs one's goals, expectations, and actions. For instance, *freedom of choice* is a general aim that directs the actions of "liberals", whereas *conservation of values* is the parallel for "conservatives".

We can attribute the lexical variations of the word content of a document to three factors:

- **Writer Ideological Belief.** A liberal writer might use words like freedom and choice regardless of the topical content of the document. These words define the abstract notion of belief held by the writer and its frequency in the document largely depends on the writer's style.

- **Topical Content.** This constitutes the main source of the lexical variations in a given document. For instance, a document about abortion is more likely to have facts related to abortion, health, marriage and relationships.

- **Topic-Ideology Interaction.** When a liberal thinker writes about abortion, his/her abstract beliefs are materialized into a set of concrete opinions and stances, therefore, we might find words like: pro-choice and feminism. On the contrary, a conservative writer might stress issues like pro-life, God and faith.

Given a collection of ideologically-labeled documents, our goal is to develop a computer model that *factors* the document collection into a representation that reflects the aforementioned three sources of lexical variations. This representation can then be used for:

- **Visualization.** By visualizing the abstract notion of belief in each ideology, and the way each ideology approaches and views mainstream topics, the user can view and contrast each ideology side-by-side and build the right mental landscape that acts as the basis for his/her future decision making.

1140

- **Classification or Ideology Identification.** Given a document, we would like to tell the user from which side it was written, and explain the ideological bias in the document at a topical level.

- **Staying Informed: Getting alternative views**[1]. Given a document written from perspective A, we would like the model to provide the user with other documents that represent alternative views about the same topic addressed in the original document.

In this paper, we approach this problem using Topic Models (Blei et al., 2003). We introduce a factored topic model that we call multi-view Latent Dirichlet Allocation or mview-LDA for short. Our model views the word content of each document as the result of the interaction between the document's idealogical and topical dimensions. The rest of this paper is organized as follows. First, in Section 2, we review related work, and then present our model in Section 3. Then in Section 4, we detail a collapsed Gibbs sampling algorithm for posterior inference. Sections 5 and 6 give details about the dataset used in the evaluation and illustrate the capabilities of our model using both qualitative and quantitative measures. Section 7 describes and evaluates the efficacy of a semi-supervised extension, and finally in Section 8 we conclude and list several directions for future research.

2 Related Work

Ideological text is inherently subjective, thus our work is related to the growing area of subjectivity analysis(Wiebe et al., 2004; Riloff et al., 2003). The goal of this area of research is to learn to discriminate between subjective and objective text. In contrast,in modeling ideology, we aim toward contrasting two or more ideological perspectives each of which is subjective in nature. Further more, subjective text can be classified into sentiments which gave rise to a surge of work in automatic opinion mining (Wiebe et al., 2004; Yu and Hatzivassiloglou, 2003; Pang et al., 2002; Turney and Littman, 2003; Popescu and Etzioni, 2005) as well as sentiment

analysis and product review mining (Nasukawa and Yi, 2003; Hu and Liu, 2004; Pang and Lee, 2008; Branavan et al., 2008; Titov and McDonald, 2008; Titov and McDonald, 2008; Mei et al., 2007; Ling et al., 2008). The research goal of sentiment analysis and classification is to identify language used to convey positive and negative opinions, which differs from contrasting two ideological perspectives. While ideology can be expressed in the form of a sentiment toward a given topic,like abortion, ideological perspectives are reflected in many ways other than sentiments as we will illustrate later in the paper. Perhaps more related to this paper is the work of (Fortuna et al., 2008; Lin et al., 2008) whose goal is to detect bias in news articles via discriminative and generative approaches, respectively. However, this work still addresses ideology at an abstract level as opposed to our approach of modeling ideology at a topical level. Finally, independently, (Paul and Girju, 2009) gives a construction similar to ours however for a different task [2].
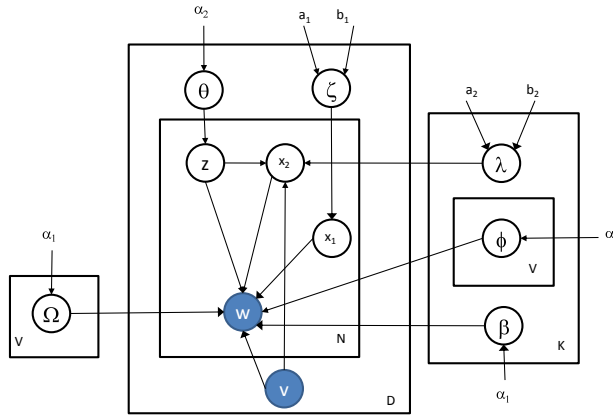
3 Multi-View Topic Models

In this section we introduce multi-view topic models, or mview-LDA for short. Our model, mview-LDA, views each document as the result of the interaction between its topical and idealogical dimensions. The model seeks to explain lexical variabilities in the document by attributing this variabilities to one of those dimensions or to their interactions. Topic models, like LDA, define a generative process for a document collection based on a set of parameters. LDA employs a semantic entity known as *topic* to drive the generation of the document in question. Each topic is represented by a topic-specific word distribution which is modeled as a multinomial distribution over words, denoted by $\text{Multi}(\beta)$. The generative process of LDA proceeds as follows:

1. Draw topic proportions $\theta_d|\alpha \sim \text{Dir}(\alpha)$.
2. For each word
   (a) Draw a topic $z_n|\theta_d \sim \text{Mult}(\theta_d)$.
   (b) Draw a word $w_n|z_n, \beta \sim \text{Mult}(\beta_{z_n})$.

In step 1 each document d samples a topic-mixing vector θ_d from a Dirichlet prior. The component $\theta_{d,k}$

---

[1] In this paper, we use the words ideology, view, perspective interchangeably to denote the same concept

[2] In fact, we only get to know about this related work after our paper was accepted

Variable	Meaning
w	word
v	document's ideology
z	topic
x_1, x_2	word switches, one per word (see text)
θ	document-specific distribution over topics
ξ	document's expected usage of the ideology's background topic
Ω	ideology's background-topic
β	ideology-independent topic distribution
ϕ	ideology-specific topic distribution
λ	topic bias across ideology

Figure 1: A plate diagram of the graphical model.

of this vector defines how likely topic k will appear in document d. For each word in the document w_n, a topic indicator z_n is sampled from θ_d, and then the word itself is sampled from a topic-specific word distribution specified by this indicator. Thus LDA can capture and represent lexical variabilities via the components of θ_d which represents the topical content of the document. In the next section we will explain how our new model mview-LDA can capture other sources of lexical variabilities beyond topical content.

3.1 Multi-View LDA

As we noted earlier, LDA captures lexical variabilities due to topical content via θ_d and the set of topics $\beta_{1:K}$. In mview-LDA each document d is tagged with the ideological view it represents via the observed variable v_d which takes values in the discrete range: $\{1, 2, \cdots, V\}$ as shown in Fig. 1. For simplicity, lets first assume that $V = 2$. The topics $\beta_{1:K}$ retain the same meaning: a set of K multinomial distributions each of which represents a given theme or factual topic. In addition, we utilize an ideology-specific topic Ω_v which is again a multinomial distribution over the same vocabulary. Ω_v models the abstract belief shared by all the documents written from view v. In other words, if v denotes the liberal perspective, then Ω_v gives high probability to words like progressive, choice, etc. Moreover, we defined a set of $K \times V$ topics that we refer to as ideology-specific topics. For example, topic $\phi_{v,k}$ represents how ideology v addresses topic k. The generative process of a document d with ideological view v_d

proceeds as follows:

1. Draw $\xi_d \sim \text{Beta}(a_1, b_1)$

2. Draw topic proportions $\theta_d | \alpha \sim \text{Dir}(\alpha_2)$.

3. For each word w_n

    (a) Draw $x_{n,1} \sim \text{Bernoulli}(\xi_d)$

    (b) If$(x_{n,1} = 1)$

        i. Draw $w_n | x_{n,1} = 1 \sim \text{Multi}(\Omega_{v_d})$

    (c) If$(x_{n,1} = 0)$

        i. Draw $z_n | \theta_d \sim \text{Mult}(\theta_d)$.

        ii. Draw $x_{n,2} | v_d, z_n \sim \text{Bernoulli}(\lambda_{z_n})$

        iii. If$(x_{n,2} = 1)$

            A. Draw $w_n | z_n, \beta \sim \text{Multi}(\beta_{z_n})$.

        iv. If$(x_{n,2} = 0)$

            A. Draw $w_n | v_d, z_n \sim \text{Multi}(\phi_{v_d, z_n})$.

In step 1, we draw a document-specific biased coin,ξ_d. The bias of this coin determines the proportions of words in the document that are generated from its ideology background topic Ω_{v_d}. As in LDA, we draw the document-specific topic proportion θ_d from a Dirichlet prior. θ_d thus controls the lexical variabilities due to topical content inside the document.

To generate a word w_n, we first generate a coin flip $x_{n,1}$ from the coin ξ_d. If it turns head, then we proceed to generate this word from the ideology-specific topic associated with the document's ideological view v_d. In this case, the word is drawn independently of the topical content of the document, and thus accounts for the lexical variation due to the ideology associated with the document. The proportion of such words is document-specific by design

and depends on the writer's style to a large degree. If $x_{n,1}$ turns to be tail,we proceed to the next step and draw a topic-indicator z_n. Now, we have two choices: either to generate this word directly from the ideology-independent portion of the topic β_{z_n}, or to draw the word from the ideology-specific portion ϕ_{v_d,z_n}. The choice here is **not** document specific, but rather depends on the interaction between the ideology and the specific topic in question. If the ideology associated with the document holds a strong opinion or view with regard to this topic, then we expect that most of the time we will take the second choice, and generate w_n from ϕ_{v_d,z_n}; and vice versa. This decision is controlled by the Bernoulli variable λ_{z_n}. Therefore, in step $c.ii$, we first generate a coin flip $x_{n,2}$ from λ_{z_n}. Based on $x_{n,2}$ we either generate the word from the ideology-independent portion of the topic β_{z_n}, and this constitutes how the model accounts for lexical variation due to the topical content of the document, or generate the word from the ideology-specific portion of the topic ϕ_{v_d,z_n}, and this specifies how the model accounts for lexical variation due to the interaction between the topical and ideological dimensions of the document.

Finally, it is worth mentioning that the decision to model λ_{z_n}[3] at the topic-ideology level rather than at the document level, as we have done with ξ_d, stems from our goal to capture ideology-specific behavior on a corpus level rather than capturing document-specific writing style. However, it is worth mentioning that if one truly seeks to measure the degree of bias associated with a given document,then one can compute the frequency of the event $x_{n,2} = 0$ from posterior samples. In this case, λ_{z_n} acts as the prior bias only. Moreover, computing the frequency of the event $x_{n,2} = 0$ and $z_n = k$ gives the document's bias toward topic k per se.

Finally, it is worth mentioning that all multinomial topics in the model: β, Ω, ϕ are generated once for the whole collection from a symmetric Dirichlet prior, similarly, all bias variables, $\lambda_{1:K}$ are sampled from a Beta distribution also once at the beginning of the generative process.

---

[3]In an earlier version of the work we modeled λ on a per-ideology basis, however, we found that using a single shared λ results in more robust results

4 Posterior Inference Via Collapsed Gibbs Sampling

The main tasks can be summarized as follows:

- **Learning**: Given a collection of documents, find a point estimate of the model parameters (i.e. $\beta, \Omega, \phi, \lambda$,etc.).

- **Inference**: Given a new document, and a point estimate of the model parameters, find the posterior distribution of the latent variables associated with the document at hand: $(\theta_d, \{x_{n,1}\}, \{z_n\}, \{x_{n,2}\})$.

Under a hierarchical Bayesian setting, like the approach we took in this paper, both of these tasks can be handled via posterior inference. Under the generative process, and hyperparmaters choices, outlined in section 3, we seek to compute:

$$P(\mathbf{d}_{1:D}, \beta_{1:K}, \mathbf{\Omega}_{1:V}, \phi_{\mathbf{1:V,1:K}}, \lambda_{1:K}|\alpha, a, b, \mathbf{w}, \mathbf{v}),$$

where $\mathbf{d}$ is a shorthand for the hidden variables $(\theta_d, \xi_d, \mathbf{z}, \mathbf{x_1}, \mathbf{x_2})$ in document d. The above posterior probability is unfortunately intractable,and we approximate it via a collapsed Gibbs sampling procedure (Griffiths and Steyvers, 2004; Gelman et al., 2003) by integrating out, i.e. collapsing, the following hidden variables: the topic-mixing vectors θ_d and the ideology bias ξ_d for each document, as well as all the multinomial topic distributions: (β, Ω and ϕ) in addition to the ideology-topic biases given by the set of λ random variables.

Therefore, the state of the sampler at each iteration contains only the following topic indicators and coin flips for each document:$(\mathbf{z}, \mathbf{x_1}, \mathbf{x_2})$. We alternate sampling each of these variables conditioned on its Markov blanket until convergence. At convergence, we can calculate expected values for all the parameters that were integrated out, especially for the topic distributions, for each document's latent representation (mixing-vector) and for all coin biases. To ease the calculation of the Gibbs sampling update equations we keep a set of sufficient statistics (SS) in the form of co-occurrence counts and sum matrices of the form C_{eq}^{EQ} to denote the number of times instance e appeared with instance q. For example, C_{wk}^{WK} gives the number of times word w was sampled from the ideology-independent portion of

topic k. Moreover, we follow the standard practice of using the subscript $-i$ to denote the same quantity it is added to without the contribution of item i. For example,$C^{WK}_{wk,-i}$ is the same as C^{WK}_{wk} without the contribution of word w_i. For simplicity, we might drop dependencies on the document whenever the meaning is implicit form the context.

For word w_n in document d, instead of sampling $z_n, x_{n,1}, x_{n,2}$ independently, we sample them as a block as follows:

$$P(x_{n,1} = 1 | w_n = w, v_d = v) \propto$$
$$(C^{DX_1}_{d1,-n} + a_1) \times \frac{C^{VW}_{vw,-n} + \alpha_1}{\sum_{w'}(C^{VW}_{vw',-n} + \alpha_1)}$$

$$P(x_{n,1} = 0, x_{2,n} = 1, z_n = k | w_n = w, v_d = v)$$
$$\propto (C^{DX_1}_{d0,-n} + b_1) \times \frac{C^{KX_2}_{k1,-n} + a_2}{C^{KX_2}_{k1,-n} + C^{KX_2}_{k0,-n} + a_2 + b_2}$$
$$\times \frac{C^{KW}_{kw,-n} + \alpha_1}{\sum_{w'}(C^{KW}_{kw',-n} + \alpha_1)} \times \frac{C^{DK}_{dk,-n} + \alpha_2}{\sum_{k'}(C^{DK}_{dk',-n} + \alpha_2)}$$

$$P(x_{n,1} = 0, x_{2,n} = 0, z_n = k | w_n = w, v_d = v)$$
$$\propto (C^{DX_1}_{d0,-n} + b_1) \times \frac{C^{KX_2}_{k0,-n} + b_2}{C^{KX_2}_{k1,-n} + C^{KX_2}_{k0,-n} + a_2 + b_2}$$
$$\times \frac{C^{VKW}_{vkw,-n} + \alpha_1}{\sum_{w'}(C^{VKW}_{vkw',-n} + \alpha_1)} \times \frac{C^{DK}_{dk,-n} + \alpha_2}{\sum_{k'}(C^{DK}_{dk',-n} + \alpha_2)}$$

The above three equations can be normalized to form a $2 * K + 1$ multinomial distribution: one component for generating a word from the ideology topic, K components for generating the word from the ideology-independent portion of topic $k = 1, \cdots, K$, and finally K components for generating the word from the ideology-specific portion of topic $k = 1, \cdots, K$. Each of these $2 * K + 1$ cases corresponds to a unique assignment of the variables $z_n, x_{n,1}, x_{n,2}$. Therefore, our Gibbs sampler just repeatedly draws sample from this $2*K+1$-components multinomial distribution until convergence. Upon convergence, we compute point estimates for all the collapsed variables by a simple marginalization of the appropriate count matrices. During **inference**, we hold the corpus-level count matrices fixed, and keep sampling from the above

$2 * K + 1$-component multinomial while only changing the document-level count matrices: C^{DK}, C^{DX_1} until convergence. Upon convergence, we compute estimates for ξ_d and θ_d by normalizing C^{DK} and C^{DX_1} (or possibly averaging this quantity across posterior samples). As we mentioned in Section 3, to compute the ideology-bias in addressing a given topic say k in a given document, say d, we can simply compute the expected value of the event $x_{n,2} = 0$ and $z_n = k$ across posterior samples.

5 Data Sets

We evaluated our model over three datasets: the bitterlemons croups and a two political blog-data set. Below we give details of each dataset.

5.1 The Bitterlemons dataset

The bitterlemons corpus consists of the articles published on the website http://bitterlemons.org/. The website is set up to contribute to mutual understanding between Palestinians and Israelis through the open exchange of ideas. Every week, an issue about the Israeli-Palestinian conflict is selected for discussion, and a Palestinian editor and an Israeli editor contribute one article each addressing the issue. In addition, the Israeli and Palestinian editors invite one Israeli and one Palestinian to express their views on the issue. The data was collected and pre-proceed as describes in (Lin et al., 2008). Overall, the dataset contains 297 documents written from the Israeli's point of view, and 297 documents written from the Palestinian's point of view. On average each document contains around 740 words. After trimming words appearing less than 5 times, we ended up with a vocabulary size of 4100 words. We split the dataset randomly and used 80% of the documents for training and the rest for testing.

5.2 The Political Blog Datasets

The first dataset refereed to as *Blog-1* is a subset of the data collected and processed in (Yano et al., 2009). The authors in (Yano et al., 2009) collected blog posts from blog sites focusing on American politics during the period November 2007 to October 2008. We selected three blog sites from this dataset: the Right Wing News (right-ideology) ; the Carpetbagger, and Daily Kos as representatives

	Israeli View	US role	Palestinian View
	arafat state leader roadmap george election month iraq week peace june realistic yasir senior involvement clinton november post mandate terrorism	bush US president american sharon administration prime settlement pressure policy washington ariel new middle	unit state american george powell minister colin visit internal policy statement express pro previous package work transfer european administration receive

Roadmap process

palestinian israeli peace year political process state end right government need conflict way security	process force terrorism unit road demand provide confidence element interim discussion want union succee point build positive recognize present timetable	roadmap phase violence security ceasefire state plan international step implement authority final quartet issue map effort	roadmap end settlement implementation obligation stop expansion commitment consolidate fulfill unit illegal present previou assassination meet forward negative calm	palestinian israeli Peace political occupation process end security conflict way government people time year force negotiation

Israeli Background topic

Arab Involvement

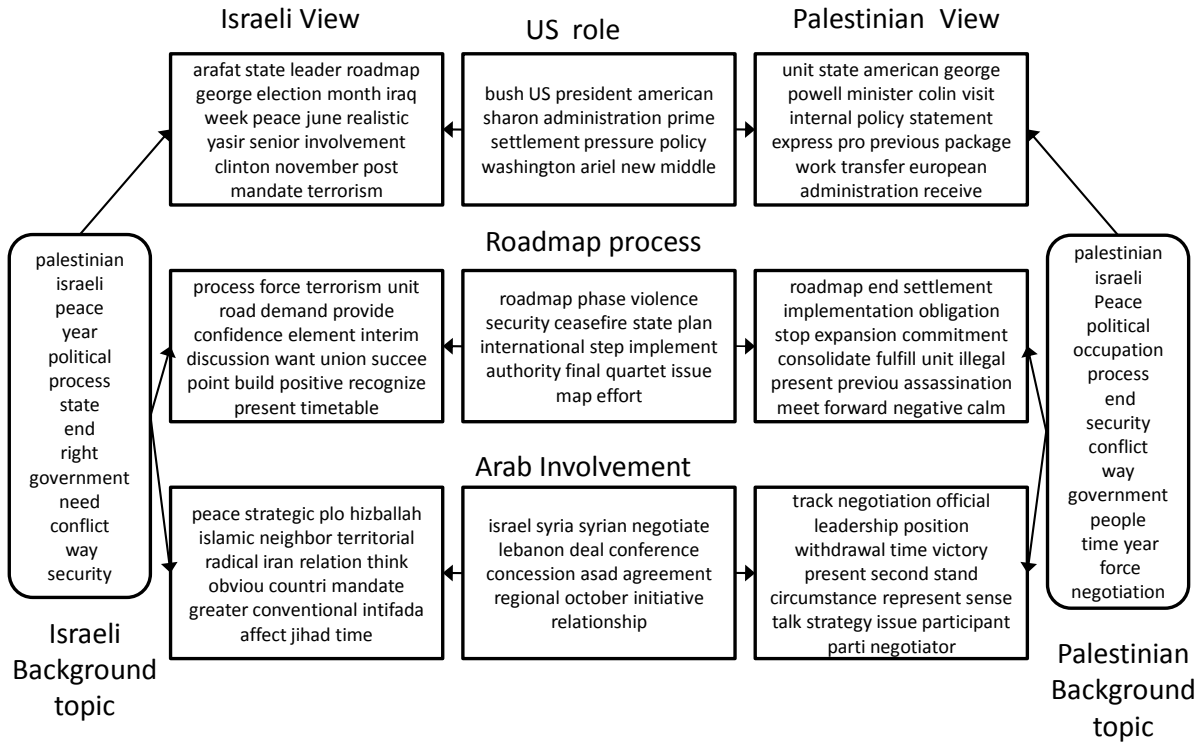| peace strategic plo hizballah islamic neighbor territorial radical iran relation think obviou countri mandate greater conventional intifada affect jihad time	israel syria syrian negotiate lebanon deal conference concession asad agreement regional october initiative relationship	track negotiation official leadership position withdrawal time victory present second stand circumstance represent sense talk strategy issue participant parti negotiator

Palestinian Background topic

Figure 2: Illustrating the big picture overview over the bitterlemons dataset using few topics. Each box lists the top words in the corresponding multinomial topic distribution. See text for more details

of the liberal view (left-ideology). After trimming short posts of less than 20 words, we ended up with 2040 posts distributed as 1400 from the left-wing and the rest from the right-wing. On average, each post contains around 100 words and the total size of the vocabulary is 14276 words. For this dataset, we followed the train-test split in (Yano et al., 2009). In this split each blog is represented in both training and test sets. Thus this dataset does not measure the model's ability to generalize to a totally different writing style.

The second dataset refereed to as *Blog-2* is similar to Blog-1 in its topical content and time frame but larger in its blog coverage (Eisenstein and Xing, 2010). Blog-2 spans 6 blogs: three from the left-wing and three from the right-wing. The dataset contains 13246 posts. After removing words that appear less then 20 times, the total vocabulary becomes 13236 with an average of 200 words per post. We used 4 blogs (2 from each view) for training and held two blogs (one from each view) for testing. Thus this dataset measures the model's ability

to generalize to a totally new blog.

6 Experimental Results

In this section we gave various qualitative and quantitative evaluations of our model over the datasets listed in Section 5. For all experiments, we set $\alpha_1 = .01, \alpha_2 = .1, a = 1$ and $b = 1$. We run Gibbs sampling during training for 1000 iterations. During inference, we ran Gibbs sampling for 300 iterations, and took 10 samples, with 50-iterations lag, for evaluations.

6.1 Visualization and Browsing

One advantage of our approach is its ability to create a "big-picture" overview of the interaction between ideology and topics. In figure 2 we show a portion of that diagram over the bitterlemons dataset. First note how the ideology-specific topics in both ideology share the top-three words, which highlights that the two ideologies seek peace even though they still both disagree on other issues. The figure gives example of three topics: the US role, the Roadmap peace process, and the Arab involvement in the conflict

(the name of these topics were hand-crafted). For each topic, we display the top words in the ideology-independent part of the topic (β), along with top words in each ideology's view of the topic (ϕ).

For example, when discussing the roadmap process, the Palestinian view brings the following issues: [the Israeli side should] implement the obligatory points in this agreement, stop expansion of settlements, and move forward to the commitments brought by this process. On the other hand, the Israeli side brings the following points: [Israelis] need to build confidence [with Palestinian], address the role of terrorism on the implementation of the process, and ask for a positive recognition of Israel from the different Palestinian political parties. As we can see, the ideology-specific portion of the topic **needn't** always represent a **sentiment** shared by its members toward a given topic, but it might rather includes extra important dimensions that need to be taken into consideration when addressing this topic.

Another interesting topic addresses the involvement of the neighboring Arab countries in the conflict. From the Israeli point of view, Israel is worried about the existence of hizballah [in lebanon] and its relationship with radical Iran, and how this might affect the Palestinian-uprising (Intifada) and Jihad. From the other side, the Palestinians think that the Arab neighbors need to be involved in the peace process and negotiations as some of these countries like Syria and Lebanon are involved in the conflict.

The user can use the above chart as an entry point to retrieve various documents pertinent to a given topic or to a given view over a specific topic. For instance, if the user asks for a representative sample of the Israeli(Palestinian) view with regard to the roadmap process, the system can first retrieve documents tagged with the Israeli(Palestinian) view and having a high topical value in their latent representation θ over this topic. Finally, the system then sorts these documents by how much bias they show over this topic. As we discussed in Section 4, this can be done by computing the expected value of the event $x_{n,2} = 0$ and $z_n = k$ where k is the topic under consideration.

6.2 Classification

We have also performed a classification task over all the datasets. The Scenario proceeded as follows.

We train a model over the training data with various number of topics. Then given a test document, we predict its ideology using the following equation:

$$v_d = \arg max_{v \in V} P(\mathbf{w_d}|v); \qquad (1)$$

We use three baselines. The first baseline is an SVM classifier trained on the normalized word frequency of each document. We trained SVM using a regularization parameter in the range $\{1, 10, 20, \cdots, 100\}$ and report the best result (i.e. no cross-validation was performed). The other two are supervised LDA models: supervised LDA (sLDA) (Wang et. al., 2009; Blei and McCauliffe, 2007) and discLDA (Lacoste-Julien et al., 2008). discLDA is a conditional model that divides the available number of topics into class-specific topics and shared-topics. Since the code is not publicly available, we followed the same strategy in the original paper and share $0.1K$ topics across ideologies and then divide the rest of the topics between ideologies[4]. However, unlike our model, there are no internal relationships between these two sets of topics. The decision rule employed by discLDA is very similar to the one we used for mview-LDA in Eq (1). For sLDA, we used the publicly available code by the authors.

As shown in Figure 3, our model performs better than the baselines over the three datasets. We should note from this figure that mview-LDA peaks at a small number of topics, however, each topic is represented by three multinomials. Moreover, it is evident from the figure that the experiment over the blog-2 dataset which measures each model's ability to generalizes to a totally unseen new blog is a harder task than generalizing to unseen posts form the same blog. However, our model still performs competitively with the SVM baseline. We believe that separating each topic into an ideology-independent part and ideology-specific part is the key behind this performance, as it is expected that the new blogs would still share much of the ideology-independent parts of the topics and hopefully would use similar (but

---

[4](Lacoste-Julien et al., 2008) gave an optimization algorithm for learning the topic structure (transformation matrix), however since the code is not available, we resorted to one of the fixed splitting strategies mentioned in the paper. We tried other splits but this one gives the best results
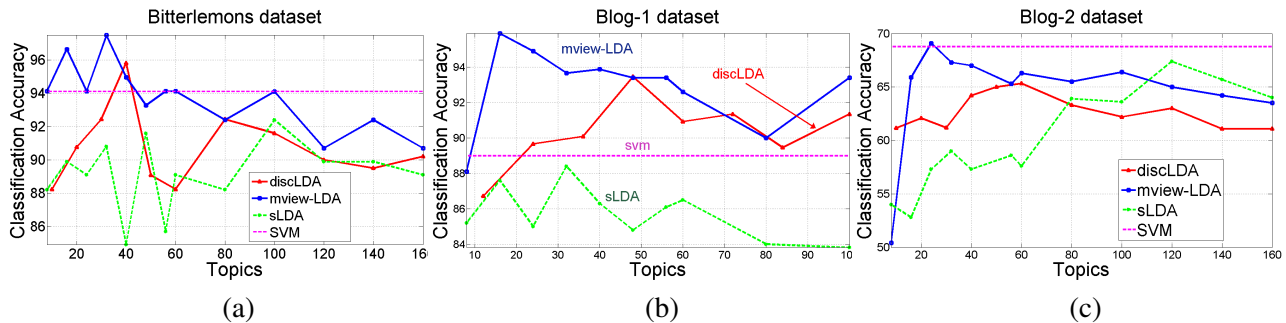
Figure 3: Classification accuracy over the Bitterlemons dataset in (a) and over the two blog datasets in (b) and (c). For SVM we give the best result obtained across a wide range of the SVM's regularization parameter(not the cross-validation result).

no necessarily all) words from the ideology-specific parts of each topic when addressing this topic.

Finally, it should be noted that the bitterlemons dataset is a multi-author dataset and thus the models were tested on some authors that were not seen during training, however, two factors contributed to the good performance by all models over this dataset. The first being the larger size of each document (740 words per document as compared to 200 words per post in blog-2) and the second being the more formal writing style in the bitterlemons dataset.

6.3 An Ablation Study

To understand the contribution of each component of our model, we conducted an ablation study over the bitterlemons dataset. In this experiment we turned-off one feature of our model at a time and measured the classification performance. The results are shown in Figure 4. Full, refers to the full model; No-Ω refers to a model in which the ideology-specific background topic Ω is turned-off; and No-ϕ refers to a model in which the ideology-specific portions of the topics are turned-off. As evident from the figure, ϕ is more important to the model than Ω and the difference in performance between the full model and the No-ϕ model is rather significant. In fact without ϕ the model has little power to discriminate between ideologies beyond the ideology-specific background topic Ω.

6.4 Retrieval: Getting the Other Views

To evaluate the ability of our model in finding alternative views toward a given topic, we conducted the following experiment over the Bitterlemons corpus. In this corpus each document is associated with a meta-topic that highlights the issues addressed in this document like: "A possible Jordanian role",
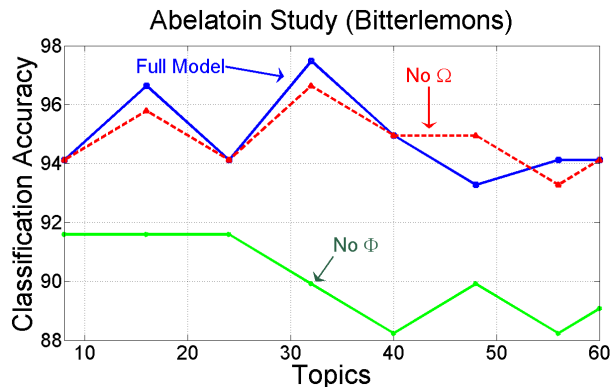

Figure 4: An Ablation study over the bitterlemons dataset.

"Demography and the conflict",etc. There are a total of 148 meta-topics. These topics were not used in fitting our model but we use them in the evaluation as follows. We divided the dataset into 60% for training and 40% for testing. We trained mview-LDA over the training set, and then used the learned model to infer the latent representation of the test documents as well as their ideologies. We then used each document in the training set as a query to retrieve documents from the test set that address the same meta-topic in the query document but from the other-side's perspective. Note that we have access to the view of the query document but not the view of the test document. Moreover, the value of the meta-topic is only used to construct the ground-truth result of each query over the test set. In addition to mview-LDA, we also implemented a strong baseline using SVM+Dirichlet smoothing that we will refer to as LM. In this baseline, we build an SVM classifier over the training set, and use Dirichlet-smoothing to represent each document (in test and training set) as a multinomial-distribution over the vocabulary. Given a query document d, we rank documents in
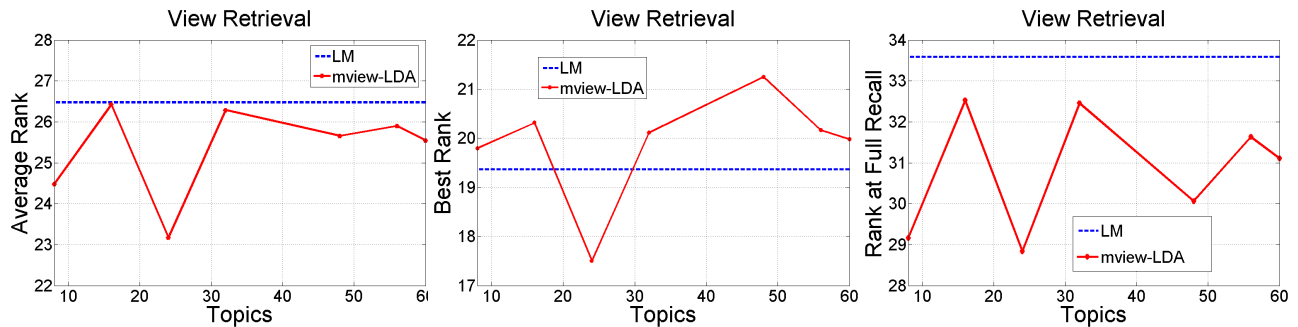
Figure 5: Evaluating the performance of the view-Retrieval task. Figure compare performance between mview-LD vs. an SVM+a smoothed language model approach using three measures: average rank, best rank and rank at full recall. (*Lower better*)

.

the test set by each model as follows:

- mview-LDA: we computed the cosine-distance between $\theta_d^{\text{mv}-\text{LDA}-shared}$ and $\theta_{d'}^{\text{mv}-\text{LDA}-shared}$ weighted by the probability that d' is written from a different view than v_d. The latter quantity can be computed by normalizing $P(v|d')$. Moreover, $\theta_{d,k}^{\text{mv}-\text{LDA}-shared} \propto \sum_n I\big[(x_{n,1} = 0) \text{ and } (x_{n,2} = 1) \text{ and } (z_n = k)\big]$, and n ranges over words in document d. Intuitively, we would like $\theta_d^{\text{mv}-\text{LDA}-shared}$ to reflect variation due to the topical content, but not ideological view of the document.

- LM: For a document d', we apply the SVM classifier to get $P(v|d')$, then we measure similarity by computing the cosine-distance between the smoothed multinomial-distribution of d and d'. We combine these two components as in mview-LDA.

Finally we rank documents in the test set in a descending-order and evaluate the resulting ranking using three measures: the rank at full recall (lowest rank), average rank, and best rank of the ground-truth documents as they appear in the predicted ranking. Figure 5 shows the results across a number of topics. From this figure, it is clear that our model outperforms this baseline over all measures. It should be noted that this is a very hard task since the meta-topics are very fine-grained like: Settlements revisited, The status of the settlements, Is the roadmap still relevant?,The ceasefire and the roadmap: a progress report,etc. We did not attempt to cluster these meta-topics since our goal is just to compare our model against the baseline.

7   A Semi-Supervised Extension

In this section we present and assess the efficacy of a semi-supervised extension of mview-LDA. In this setting, the model is given a set of ideologically-labeled documents and a set of unlabeled documents. One of the key advantages of using a probabilistic graphical model is the ability to deal with hidden variables in a principled way. Thus the only change needed in this case is adding a single step in the sampling algorithm to sample the ideology v of an unlabeled document as follows:

$$P(v_d = v | rest) \quad \propto \quad P(\mathbf{w_d} | v_d = v, \mathbf{z_d}, \mathbf{x_{1,d}}, \mathbf{x_{2,d}})$$

Note that the probability of the indicators $(\mathbf{x_{1,d}}, \mathbf{x_{2,d}}, \mathbf{z_d})$ do not depend on the view of the document and thus got absorbed in the normalization constant, and thus one only needs to measures the likelihood of generating the words in the document under the view v. We divide the words into three groups: $A_d = \{w_n | x_{n,1} = 1\}$ is the set of words generated from the view-background topic, $B_{d,k} = \{w_n | z_n = k, x_{n,1} = 0, x_{n,2} = 1\}$ is the set of words generated from β_k, and $C_{d,k} = \{w_n | z_n = k, x_{n,1} = 0, x_{n,2} = 0\}$ is the set of words generated from $\phi_{k,v}$. The probability of $B_{d,k}$ does not depend on the value of v and thus can be absorbed into the normalization factor. Therefore, we only need to compute the following probability:$P(A_d, C_{d,1:K} | v_d = v, rest) =$

$$\prod_k \int_{\phi_{k,v}} P(C_{d,k} | \phi_{k,v}, rest) p(\phi_{k,v} | rest) d\phi_{k,v}$$

$$\times \quad \int_\Omega P(A_d | \Omega, rest) p(\Omega | rest) d\Omega \qquad (2)$$

All the integrals in (2) reduce to the ratio of two log partition functions. For example, the product of integrals containing $C_{d,k}$ reduce to:

$$\prod_k \frac{\prod_w \Gamma\left(C_{dkw}^{DKW,X_2=0} + C_{vkw,-d}^{VKW} + \alpha_1\right)}{\Gamma\left(\sum_w \left[C_{dkw}^{DKW,X_2=0} + C_{vkw,-d}^{VKW} + \alpha_1\right]\right)}$$
$$\times \frac{\Gamma\left(\sum_w \left[C_{vkw}^{VKW} + \alpha_1\right]\right)}{\prod_w \Gamma\left(C_{vkw,-d}^{VKW} + \alpha_1\right)} \quad (3)$$

Unfortunately, the above scheme does not mix well because the value of the integrals in (2) are very low for any view other than the view of the document in the current state of the sampler. This happens because of the tight coupling between v_d and the indicators $(\mathbf{x_1}, \mathbf{x_2}, \mathbf{z})$. To remedy this problem we used a Metropolis-Hasting step to sample $(v_d, \mathbf{x_1}, \mathbf{x_2}, \mathbf{z})$ jointly. We construct a set of V proposals each of which is indexed by a possible view: $q_v(\mathbf{x_1}, \mathbf{x_2}, \mathbf{z}) = P(\mathbf{x_1}, \mathbf{x_2}, \mathbf{z}|v_d = v, \mathbf{w_d})$. Since we have a collection of proposal distributions, we select one of them at random at each step. To generate a sample from $q_{v*}()$, we run a few iterations of a restricted Gibbs scan over the document d conditioned on fixing $v_d = v*$ and then take the last sample jointly with $v*$ as our proposed new state. With probability $\min(r,1)$, the new state $(v*, \mathbf{x_1}*, \mathbf{x_2}*, \mathbf{z}*)$ is accepted, otherwise the old state is retained. The acceptance ratio, r, is computed as: $r = \frac{p(\mathbf{w_d}|v*,\mathbf{x_1}*,\mathbf{x_2}*,\mathbf{z}*)}{p(\mathbf{w_d}|v,\mathbf{x_1},\mathbf{x_2},\mathbf{z})}$, where the non-* variables represent the current state of the sampler. It is interesting to note that the above acceptance ratio is equivalent to a likelihood ratio test. We compute the marginal probability $P(\mathbf{w_d}|..)$ using the estimated-theta method (Wallach et al., 2009).

We evaluated the semi-supervised extension using the blog-2 dataset as follows. We reveal $R\%$ of the labels in the training set; then we train mview-LDA only over the labeled portion and train the semi-supervised version (ss-mview-LDA) on both the labeled and unlabeled documents. Finally we evaluate the classification performance on the test set. We used $R = \{20, 40, 80\}$. The results are given in Table 1 which shows a decent improvement over the supervised mview-LDA.

R	mview-LDA	ss-mview-LDA
80	65.60	66.41
60	62.31	65.43
20	60.87	63.25

Table 1: Classification performance of the semi-supervised model. R is the ratio of labeled documents.

8 Discussion and Future Work

In this paper, we addressed the problem of modeling ideological perspective at a topical level. We developed a factored topic model that we called multiView-LDA or mview-LDA for short. mview-LDA factors a document collection into three set of topics: ideology-specific, topic-specific, and ideology-topic ones. We showed that the resulting representation can be used to give a bird-eyes' view to where each ideology stands with regard to mainstream topics. Moreover, we illustrated how the latent structure induced by the model can by used to perform bias-detection at the document and topic level, and retrieve documents that represent alternative views.

It is important to mention that our model induces a hierarchical structure over the topics, and thus it is interesting to contrast it with hierarchical topic models like hLDA (Blei et al., 2003) and PAM (Li and McCallum, 2006; Mimno et al., 2007). First, these models are unsupervised in nature, while our model is supervised. Second, the semantic of the hierarchical structure in our model is different than the one induced by those models since documents in our model are constrained to use a specific portion of the topic structure while in those models documents can freely sample words from any topic. Finally, in the future we plan to extend our model to perform joint modeling and summarization of idealogical discourse.

9 Acknowledgment

1149

References

J. Wiebe, T. Wilson, R.Bruce, M. Bell, and M. Martin. Learning subjective language. Computational Linguistics, 30(3), 2004.

H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In Proceedings of EMNLP-2003

B. Pang, L. Lee, and S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP-2002.

P. Turney and M. Littman. Measuring praise and criticism: Inference of semantic orientation from association. ACM TOIS, 21(4):315346, 2003

A. Popescu and O. Etzioni. Extracting product features and opinions from reviews. In Proceedings of HLT/EMNLP-2005, pages 339346, 2005.

T. Nasukawa and J. Yi. Sentiment analysis: Capturing favorability using natural language processing. In Proceedings of K-CAP, 2003.

M. Hu and B. Liu. Mining and summarizing customer reviews. In Proceedings of KDD, 2004.

B. Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(12), 1135, 2008.

S. Branavan, H. Chen, J. Eisenstein and R. Barzilay. Learning Document-Level Semantic Properties from Free-text Annotations, Proceedings of ACL, 2008.

I. Titov and R. McDonald. Modeling Online Reviews with Multi-Grain Topic Models International World Wide Web Conference (WWW), 2008.

I. Titov and R. McDonald. A Joint Model of Text and Aspect Ratings for Sentiment Summarization Association for Computational Linguistics (ACL), 2008.

Q. Mei, X. Ling, M. Wondra, H. Su, ChengXiang Zhai. Topic Sentiment Mixture: Modeling Facets and Opinions in Weblogs, Proceedings of the 16th International World Wide Web Conference (WWW), pages 171-180, 2007.

X. Ling, Q. Mei, C. Zhai, B. Schatz. Mining Multi-Faceted Overviews of Arbitrary Topics in a Text Collection, Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD' 08), pages 497-505, 2008

B. Fortuna , C. Galleguillos, N. Cristianini. Detecting the bias in media with statistical learning methods. In: Text Mining: Theory and Applications. Taylor and Francis Publisher,2008.

W. Lin, E.P. Xing, and A. Hauptmann. A Joint Topic and Perspective Model for Ideological Discourse European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD), 2008.

T. A. Van Dijk. Ideology: A multidisciplinary approach. Sage Publications, 1998.

T. Griffiths, M. Steyvers Finding scientific topics.PNAS, 101:5228-5235, 2004.

A. Gelman, J. Carlin, Hal Stern, and Donald Rubin. Bayesian Data Analysis, Chapman-Hall, 2 edition,2003.

D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. Journal of Machine Learning Research, 3:9931022, January 2003.

T. Yano, W. W. Cohen, and N. A. Smith. Predicting Response to Political Blog Posts with Topic Models. NAACL-HLT 2009, Boulder, CO, MayJune 2009

J. Eisenstein and E.P. Xing.The CMU-2008 Political Blog Corpus. CMU-ML-10-101 Technical Report, 2010.

C. Wang, D. Blei and L. Fei-Fei. Simultaneous image classification and annotation. CVPR, 2010.

D. Blei and J. McAuliffe. Supervised topic models. NIPS, 2007.

S. Lacoste-Julien, F. Sha, and M. Jordan. DiscLDA: Discriminative Learning for Dimensionality Reduction and Classification. Neural Information Processing Systems Conference (NIPS08), Vancouver, British Columbia, December 2008.

E. Riloff, J. Wiebe, and T. Wilson. Learning subjective nouns using extraction pattern bootstrapping. In Proceedings of CoNLL-2003.

D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In Neural Information Processing Systems (NIPS)16, 2003.

D. Mimno, W. Li and A. McCallum. Mixtures of Hierarchical Topics with Pachinko Allocation. In International Conference of Machine Learning, ICML, 2007.

W. Li, and A. McCallum. Pachinko Allocation: DAG-structured Mixture Models of Topic Correlations. In International Conference of Machine Learning, ICML, 2006.

M. Paul and R. Girju. Cross-cultural Analysis of Blogs and Forums with Mixed-Collection Topic Models. EMNLP 2009.

H. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation Methods for Topic Models. ICML 2009.