Modeling Polarizing Topics: When Do Different Political Communities Respond Differently to the Same News?

Ramnath Balasubramanyan
Language Technologies Institute
Carnegie Mellon University
rbalasub@cs.cmu.edu

William W. Cohen
Machine Learning Department
Carnegie Mellon University
wcohen@cs.cmu.edu

Doug Pierce and **David P. Redlawsk**
Political Science Department
Rutgers University
drpierce@eden.rutgers.edu
redlawsk@rutgers.edu

Abstract

Political discourse in the United States is getting increasingly polarized. This polarization frequently causes different communities to react very differently to the same news events. Political blogs as a form of social media provide an unique insight into this phenomenon. We present a multi-target, semi-supervised latent variable model, MCR-LDA to model this process by analyzing political blogs posts and their comment sections from different political communities jointly to predict the degree of polarization that news topics cause. Inspecting the model after inference reveals topics and the degree to which it triggers polarization. In this approach, community responses to news topics are observed using sentiment polarity and comment volume which serves as a proxy for the level of interest in the topic. In this context, we also present computational methods to assign sentiment polarity to the comments which serve as targets for latent variable models that predict the polarity based on the topics in the blog content. Our results show that the joint modeling of communities with different political beliefs using MCR-LDA does not sacrifice accuracy in sentiment polarity prediction when compared to approaches that are tailored to specific communities and additionally provides a view of the polarization in responses from the different communities.

Introduction

Recent work in political psychology has made it clear that political decision-making is strongly influenced by emotion. For instance, (Lodge and Taber 2000) propose a theory of "motivated reasoning", in which political information is processed in a way that is determined, in part, by a quickly-computed emotional react to that information. Strong experimental evidence for motivated reasoning (sometimes called "hot cognition") exists (Huang and Price 2001); (Redlawsk 2002); (Redlawsk 2006); (Isbell, Ottati, and Bruns 2006). However, despite some recent proposals (Kim, Taber, and Lodge 2008) it is unclear how to computationally model a person's emotional reaction to news, and how to collect the data necessary to fit such a model. One problem is that emotional reactions are different for different people and communities - a fact exploited in the use of political "code words" intended to invoke a reaction in only a particular

subset of the electorate (a technique sometimes called "dog whistle politics").

In this paper, we evaluate the use of machine learning methods to predict how *members of a different political communities will emotionally respond to the same news story*. The models we use also identify the topics of discussion and the difference in responses that they evoke. Community responses in these models are measured using sentiment polarity and comment volume (as in (Yano, Cohen, and Smith 2009)). More specifically, we use a dataset of widely read ("A-list") political blogs, and attempt to predict the aggregate sentiment in the comment section of blogs and the volume in terms of the number of comments, as a function of the textual content of the blog posting. In contrast to work done traditionally in sentiment analysis which focuses on determining the sentiment expressed in text, in this work, we focus on the task of predicting the sentiment that a block of text will evoke in readers, expressed in the comment section, as a response to the blog post. This task is related to, but distinct from, several other studies that have been made using comments and discussions in political communities, or analysis of sentiment in comments - (O'Connor et al. 2010), (Tumasjan et al. 2010).

Below we discuss the methods used to address the various parts of this task. First, we evaluate two methods to automatically determine the comment polarity: SentiWord-Net (Baccianella and Sebastiani 2010) a general purpose resource that assigns sentiment scores to entries in WordNet, and an automated corpus-specific technique based on pointwise mutual information(Balasubramanyan et al. 2011). The quality of the polarity assessments by these techniques are made by comparing them to hand annotated assessments on a small number of blog posts. Second, we consider two community-specific methods for predicting comment polarity from post content: support vector machine classification, and sLDA, a topic-modeling-based approach. Next, we demonstrate that emotional reactions are indeed community-specific and then propose a new model Multi Community Response LDA (MCR-LDA) which is a multi-target, semi-supervised LDA that is better suited to model responses from multiple communities. This model is used to identify topics that evoke different reactions from communities that lie on different points on the political spectrum. This reaction is measured in terms of the sentiment polarity expressed

in the comments and also in terms of the volume of comments it triggers. Finally, we present our conclusions.

Data

In this study, we use a collection of blog posts from five blogs: Carpetbagger(CB)[1], Daily Kos(DK)[2], Matthew Yglesias(MY)[3], Red State(RS)[4], and Right Wing News(RWN)[5], that focus on American politics made available by (Yano, Cohen, and Smith 2009). The posts were collected during November 2007 to October 2008, which preceded the US presidential elections held in November 2008. The blogs included in the dataset vary in political ideology with blogs like Daily Kos that are Democrat-leaning and blogs like Red State tending to be much more conservative. Since we are interested in studying the responses to blog posts, the corpus only contains posts where there have been at least one comment in the six days after the post was published. It is important to note that only the text in the blog posts and comments are used in this study. All non-textual information like pictures, hyperlinks, videos etc. are discarded. In terms of text processing, for each blog, a vocabulary is created consisting of all terms that occur at least 5 times in the blog. Stopwords are eliminated using a standard stopword list. Each blog post is then represented as a bag of words from the post. Table shows statistics of the datasets.

We also use a newer version of the Daily Kos and Red State blogs to study polarization in a more current context. This version of the dataset was collected in the December 2010 to November 2011 time frame and is processed in the same way as described above. The topics of discussion in the newer dataset therefore pertains more to the issues surrounding the mid-term elections in 2010 and the primaries for the 2012 presidential elections.

Data: blog vocabulary V, standard sentiment word lists P and N
Result: Blog specific sentiment word lists
  for w in V **do**
    $avg\_pos\_PMI \leftarrow \frac{\sum_{s \in P} PMI(w,s)}{|P|}$
    $avg\_neg\_PMI \leftarrow \frac{\sum_{s \in N} PMI(w,s)}{|N|}$
    $polarity \leftarrow avg\_pos\_PMI - avg\_neg\_PMI$
  end for
  $sorted\_V \leftarrow V$ sorted by $polarity$
  $positive\_words \leftarrow$ top N of $sorted\_V$
  $negative\_words \leftarrow$ bottom N of $sorted\_V$
  return $positive\_words, negative\_words$

Algorithm 1: Using PMI to construct blog specific sentiment word lists

---

Comments Sentiment Polarity Detection

The first step in understanding the nature of posts that evoke emotional responses is to get a measure of the polarity in the sentiment expressed in the comments section of a blog post. The role of this stage in the system can be seen in Figure 3. The measure indicates the polarity of the response that the issues in the blog post and its treatment, evokes in a community.

The simplest approach to detecting sentiment polarity is to use a sentiment lexicon such as SentiWordNet (Baccianella and Sebastiani 2010) which associates a large number of words in WordNet with a positive, negative and objective score (summing up to 1). To detect the polarity all the comments for a blog post in the comment section are aggregated and for the words in the comments that are found in SentiWordNet, the net positive and negative scores based on the dominant word sense in SentiWordNet are computed. The sentiment in the comment section is deemed to be positive if the net positive score exceeds the negative score and negative otherwise. Therefore, each blog post is now associated with a binary response variable indicating the polarity of the sentiment expressed in the comments.
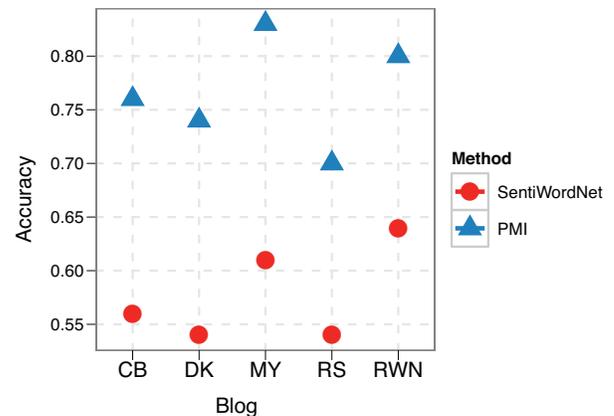

Figure 1: Evaluation of automatic comment polarity detection

Using Pointwise Mutual Information

Lack of coverage is a problem with the sentiment lexicon approach. Additionally, sentiment lexicons are usually generally purpose and do not take into consideration the particular idiosyncrasies of dialect in specialized domains and communities. To address these issues we use a Pointwise Mutual Information (PMI)(Turney 2002) based approach to build a community specific sentiment lexicon using the general purpose lexicon as a seed. The positive and negative seed lists are constructed by choosing the 100 topmost positive and negative words from SentiWordNet and manually eliminating words from this list that do not pertain to sentiment in our context. The seed lists are then used to construct a larger set of positive and negative words by computing the PMI of the words in the seed lists with every other word in

| Blog	Political alignment	#Posts	Vocabulary size	Average #words per post	Average #comments per post	Avg. #words per comment section
Carpetbagger (CB)	liberal	1201	4998	170	31	1306
Daily Kos (DK)	liberal	2597	6400	103	198	3883
Matthew Yglesias (MY)	liberal	1813	4010	69	35	1420
Red State (RS)	conservative	2357	8029	158	28	806
Right Wing Nation (RWN)	conservative	1184	6205	185	33	1015

Table 1: Dataset statistics
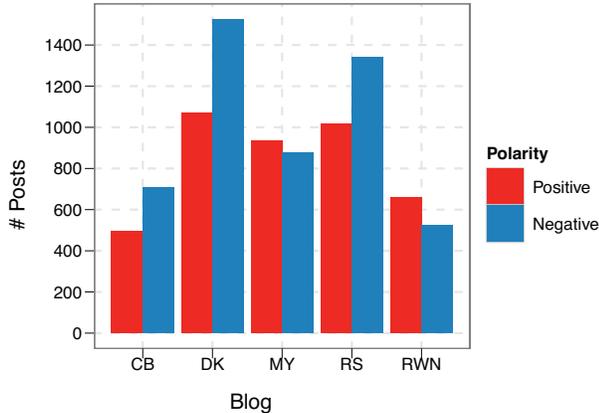
Figure 2: Number of positive and negative comment sections

the vocabulary. Algorithm 1 shows the sentiment word list construction procedure. It is important to note that different lists are constructed separately for each blog/community resulting in sentiment polarity lists that are particular to the community and ideology associated with each blog. Words in the vocabulary are ranked by the difference in the average of the PMI with positive and negative seed words. The top 1000 words in the resultant sorted list are treated as positive words and the bottom 1000 words as negative words. The comment section of every post is tagged with a positive or negative polarity as was described previously by computing the total positive and negative word counts.

It should be noted that while this method provides better estimates of comment sentiment polarity (Figure 1), it involves some manual effort in constructing a seed set that the general purpose SentiWordNet method does not require. The distribution between positive and negative comment sections in each blog obtained using the PMI method is shown in Figure 2.

Evaluation of Comment Sentiment Polarity using Human Labeling

A presumably reliable and accurate but expensive method to label comments is to perform manual labeling. We use the manual labels only for evaluating the accuracy of the general purpose SentiWordnet and PMI techniques. Approximately 30 blog posts from each blog, were labeled with either a positive or negative label based on the sentiment in the comment sections. The guideline in labeling was to determine if the sentiment in the comment section was positive or negative to the subject of the post. The chief intention of this exercise is to determine the quality of the polarity assessments of the SentiWordNet and PMI methods. While it is possible to directly use the assessments and train a classifier, the performance of the classifier will be limited by the very small number of training examples (30 instead of thousands of examples). The accuracy of the two automatic methods to determine comment polarity is shown in Figure 1.

It can be seen that the custom built community specific sentiment word lists obtained using PMI provides a better sentiment label than the general purpose SentiWordNet approach. The better accuracy of the PMI method can be explained by the fact that SentiWordNet is not customized for the political domain which tends to make it noisy for text in political blogs. For instance, the word "bush" in a general context is fairly neutral in terms of sentiment, but in the context of politics, it is laden with strong sentiment. The PMI technique corresponds more closely with the human labels but it requires some human effort in building the initial seed list of positive and negative words.

Within-community response prediction from blog content

We now address the problem of using machine learning techniques to predict the sentiment polarity of community responses in the form of comments based on the blog post contents. Figure 3 shows the setup of the system. The contents of the blog post are used as input to the algorithms described below and the target that they attempt to predict is the sentiment polarity that was detected using the PMI algorithm described in the previous section. As shown in the figure, each blog/community is treated separately and the analyses in this section reflect *within-community* characteristics.

SVM

Firstly, we use *support vector machines* (SVM) to perform classification. We frame the classification task as follows: The input features to the classifier are the words in the blog post i.e. each blog post is treated as a bag of words and the output variable is the binary comment polarity computed in the previous section using the PMI approach. For our experiments, we used the SVMLight package [6] with a simple
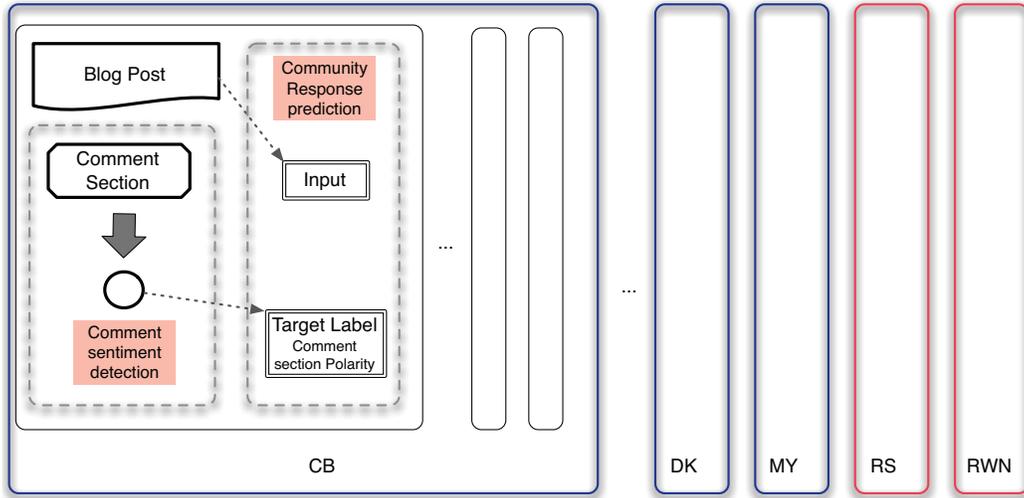
---

[6]http://svmlight.joachims.org/

Figure 3: Within-community response prediction based on blog content

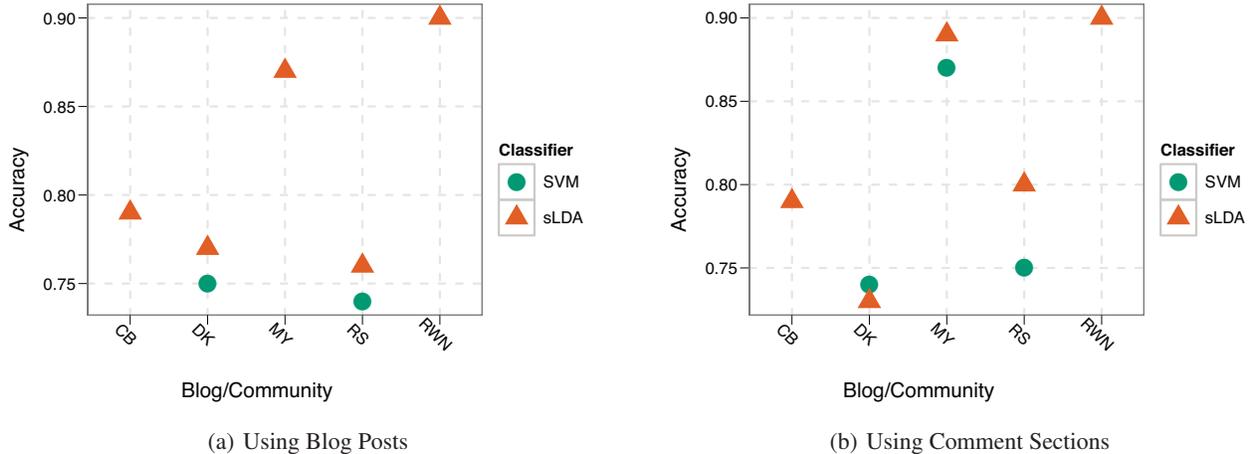(a) Using Blog Posts

(b) Using Comment Sections

Figure 4: Predicting community response - Comment sentiment polarity

linear kernel and evaluated the classifier using 10 fold cross validation.

Figure 4(a) shows the accuracy of the classifier in predicting the community response in terms of sentiment polarity for each blog in the dataset. The errors in classification can be attributed in part to the inherent difficulty of the task due to the noise of the polarity labeling schemes and in part due to the difficulty in obtaining a signal to *predict comment polarity* from the body of the post.

Supervised LDA

Next, we use Supervised Latent Dirichlet Allocation (sLDA) (Blei and McAuliffe 2008) for classification. sLDA is a model that is an extension of Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan 2003) that models each document as having an output variable in addition to the document contents. The output variable in the classification case

is modeled as an output of a logistic regression model that uses the posterior topic distribution of the LDA model as features. In this task, the output variable is +1 or -1 depending on the polarity of the comment section. In the experiments with sLDA, we set the number of topics as 15 after experimenting with a range of topics and use 10-fold cross validation. The number of topics is set lower than it usually is with topic modeling, due to the relatively short length and small number of documents.

Figure 4(a) shows the accuracy of SVM and sLDA in predicting the comment polarity based on the blog posts. It can be seen from the figure that sLDA outperforms SVM in predicting the PMI labels for every blog. sLDA provides the additional advantage of inducing topics from the bodies of the blog posts that serve to characterize the different issues that each blog addresses. The logistic regression parameters also indicate how each topic influences the output variable.
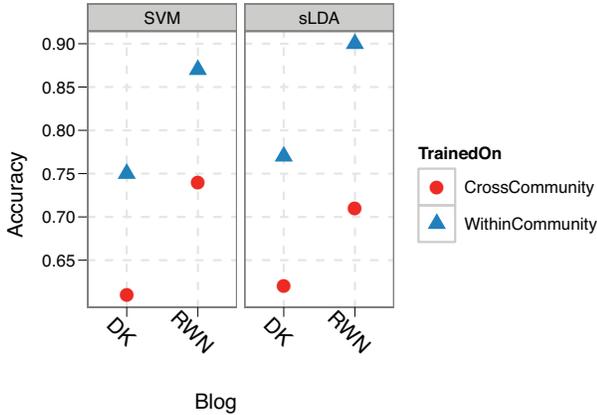
Figure 5: Cross blog results: Accuracy using SVM/sLDA

Target	MSE
RS comment volume	61.73
DK comment volume	932.05

Table 2: Volume prediction performance of MCR-LDA

Using Comments to Predict Comment Polarity

In the previous experiments we used the bodies of the blog posts to predict comment polarity. Multiple factors contribute to making this a difficult task. One major factor is the difficulty of learning potentially noisy labels using automatic methods. More interestingly, we operate under the hypothesis that there is signal about comment polarity in the bodies of the blog posts. To test this hypothesis, we train classifiers on the comment sections themselves to predict comment polarity. This serves to eliminate the effect of our hypothesis and focus on the inherent difficulty in learning the noisy labels. Figure 4(b) shows the results of these experiments. We see that once again, sLDA results are comparable to the accuracies reported by SVM. More importantly, we note that the accuracy in predicting the comment polarity is not significantly higher than the accuracy in predicting the polarity from blog posts which strongly suggests that blog posts have quite a bit of information regarding comment polarity.

Multi-community Response Modeling from Blog Content

Cross Community Experiments

We first investigate the importance of building community-specific models in the approaches described in the previous section. To examine the effect of the nature of the blog on classifier performance, we train models on the blog posts from a conservative blog (RWN) using PMI-determined polarities as targets and test the model by running liberal blog data (from DK) through it. Similarly, we test RWN blog entries by training it on a classifier trained on DK posts. The results of the experiments are in Figure 5. The plot also includes the within-community prediction results from the previous section for easy comparison. We see that the accuracy in predicting polarity degrades when blog posts are tested on a cross-community classifier trained on posts from a blog of opposite political affiliation. These results indicate

that community response modeling using SVM and sLDA is very specifically tailored to the blog that it is trained on.

Multi-community Response Modeling

In the previous section, we discussed prediction of sentiment polarity in the comments based on the blog posts separately for each blog to capture community specific reactions. Now, we turn our attention to identifying community reactions to the same blog posts from readers from different sub-communities with different political beliefs. The reactions to blog posts are measured using

- comment sentiment polarity
- volume of comments

We first construct a corpus that combines blog posts from the new version of Red State and Daily Kos and train a MCR-LDA model (described in the next sub-section) with it. For each blog post, the MCR-LDA model has four targets - the comment polarity and volume of comments from Daily Kos readers and the same from Red State readers. Figure 6 shows a pictorial description of the setup with missing values indicated by dashed lines.

Multi-community Response LDA (MCR-LDA)

In a fully supervised sLDA model, each document's target attribute is modeled using a regression model on the topic proportions. In MCR-LDA, we design a multi-target, semi-supervised LDA where documents can have multiple target variables some of which may have missing values during inference. The regression models for the multiple targets are simultaneously trained during inference. We use a
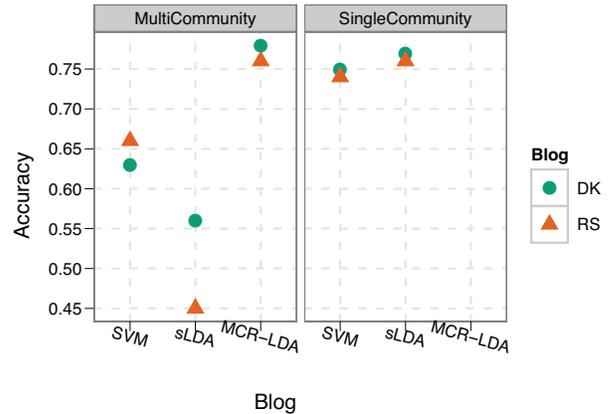

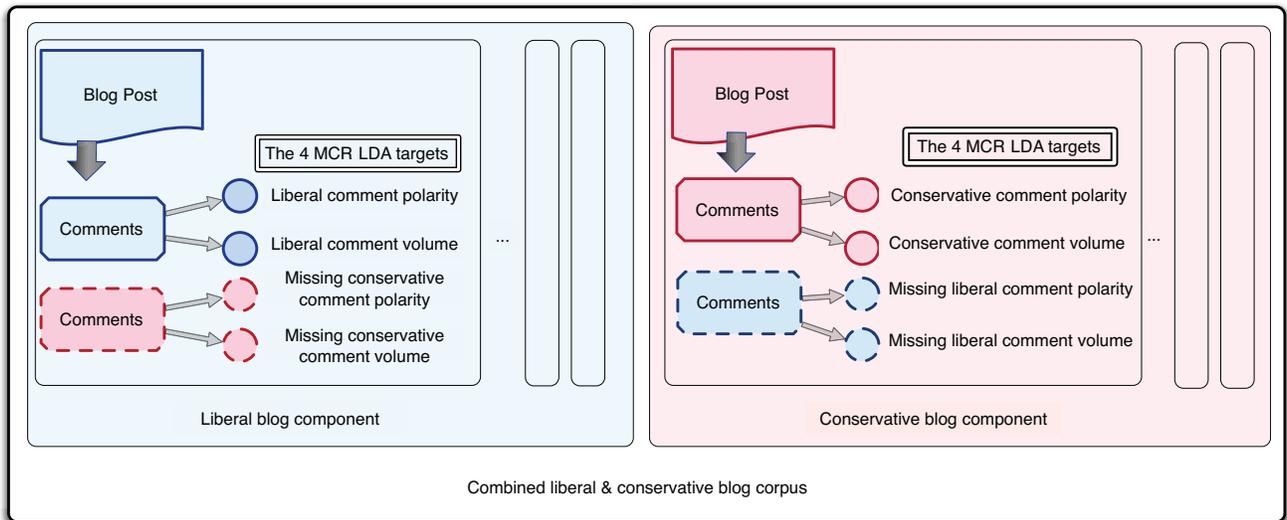Figure 7: Accuracy of MCR-LDA in predicting sentiment polarity

Figure 6: Polarizing topic identification using multi-target semi-supervised sLDA

Gibbs sampling procedure to fit the model to the data during which each iteration involves a round of sampling topics for words and a phase where the regression models for the multiple targets are trained. During the regression training stage of each iteration, documents with missing targets are not considered for training the regression model for that target. Specifically in this paper, the four targets in MCR-LDA are the two communities' comment sentiment polarity and comment volume. In the combined corpus of liberal and conservative blogs, the comment sentiment polarity and volume of liberal readers for blog posts in the conservative blog and vice-versa are missing. The joint modeling however allows topic commonality to be identified in the liberal and conservative blogs and trains four regression models each of which predict how readers from each of the communities will respond to the topic. The topics with the biggest difference in regression co-efficients in predicting the two targets indicate topics that are the most polarized.

Firstly, we evaluate the model by measuring its accuracy in predicting the comment response polarity using 10-fold cross validation as in the previous section.

The MCR-LDA model uses as its training set a combined corpus of DK and RS blog posts and targets as seen in Figure 6. The SVM and sLDA models described in previous sections however used blog-specific models. To see the performance of SVM and sLDA on a combined corpus, we first combine the corpora like we do for MCR-LDA as described in the figure (without the volume targets) and treat the comment response polarity of RS and DK as the target. This approach is sub-optimal since the SVM and sLDA models are not able to distinguish between RS and DK comment polarities. In the left pane of figure 7 we see that the prediction accuracies from the combined corpus models are much lower for SVM and sLDA than the right pane values which are obtained from blog-specific models. MCR-LDA in the left pane which is trained on a combined corpus however

has accuracies that are on par (for Red State) or slightly better (for DK) than SVM and sLDA which are trained in a community-specific fashion. These results indicate that mixing the two blogs together does not cause a drop in sentiment polarity prediction performance.

Next, we evaluate MCR-LDA in its ability to predict the volume of comments that a blog post triggers as a response from the community. The model is evaluated using mean squared error (MSE) between the known comment volume and the predicted comment volume. It should be noted that although there are two volume targets for every blog post, we are able to only evaluate one target for every blog since the true value of the other target is unknown. Table 2 shows the MSE in predicting the DK comment volume and the RS comment volume.

MCR-LDA's big advantage derived from the joint modeling lies in the identification of topics that have widely different liberal and conservative co-efficients in predicting the volume of comments and sentiment polarity. Such topics can be viewed as the most contentious topics between the two communities since it evokes very different reactions from the two communities. Table 3 shows the topics and co-efficients from a run on the new version of the dataset that have the largest difference in co-efficients between the Daily Kos and Red State targets. For each topic, the top words in the multinomial are shown in the second column. The topic labels were assigned by the authors after inspection of the top words in the topic.

The first two topics in the table are topics that have the highest difference in co-efficients with the Daily Kos co-efficients being positive. The first topic is about various aspects of the current administration's energy policy including the war in the Middle East and its effect on oil supply, nuclear and alternative sources of energy, the BP oil spill off the Gulf coast of the US and its impact on the environment. This topic expectedly got positive comments from Daily Kos

Topic	Top words	Red State Co-efficients		Daily Kos Co-efficients	
		Sentiment	Volume	Sentiment	Volume
Positive in Daily Kos, Negative in Red State					
Energy and envi-ronment	oil war obama president energy military administration american world bp government nuclear united national climate spill gas	-0.113	-20.965	0.123	15.331
Union rights and women's rights	court union women rights federal public government justice abortion workers unions supreme legal wisconsin act united people marriage decision employees labor life judge laws walker constitution anti school health department	-0.162	-33.965	0.174	8.543
Positive in Red State, Negative in Daily Kos					
Senate proce-dures	senate bill house vote rep act votes committee reform amendment week rules majority senators time senator debate republicans legislation pass congress floor rule passed ll day business reid conference filibuster	0.146	10.602	-0.361	-70.117
Republican primaries	party republican voters percent republicans romney president obama gop tea poll perry democrats palin campaign presidential conservative election political support candidate candidates bachmann people paul rick time vote mitt polls sarah week conservatives	0.364	202.899	-0.15	-12.948
Non-polarized topics					
Economy, taxes, social security	tax jobs cuts health social percent security government people budget spending care billion economy economic deficit medicare million federal insurance money taxes cut plan pay benefits americans income job program unemployment financial workers debt cost increase rate class	0.005	12.167	0.006	9.345
Mid-term elec-tions	gop democratic state senate race campaign poll gov rep district governor candidate former democrat seat run democrats john house lead party vote scott week dem senator incumbent candidates brown running races news county gubernatorial recall voters nomi-nee tom elections	0.098	-33.76	0.081	-12.82

Table 3: MCR-LDA induced polarizing topics

readers with a moderate volume co-efficient of 15.331. Red State readers, who are presumably more conservative were much more negative in sentiment about the issue. The next topic pertains to a controversy over labor laws in the state of Wisconsin and women's rights issues. It is therefore no surprise that this topic draws positive comments from liberal readers and a more negative leaning sentiment from Red State readers. The next two topics are the topics which are most polarized with a positive sentiment from Red State readers. The first topic deals with Republican efforts to filibuster and slow down proceedings in the legislature. The slow down tactic to arrest the president's intended legislative course was popular with the conservative faction and is reflected in the results here which show a high positive sentiment from Red State readers and a negative sentiment

from Daily Kos readers. The last topic in the table is mainly about the Republican primaries as is evidenced by the presence of many Republican presidential candidates in the top words (Romney, Perry, Palin, Bachmann etc.). As expected, Red State readers display a much more positive sentiment about this issue than Daily Kos readers. The last two topics in the table are the topics that have the closest co-efficients indicating that they are the least contentious topics of discussion. The first topic is about the economy, social security and tax issues which have been issues that has caused some discontent in the nation across the board. The second topic discusses the mid-term elections and more specifically about the state and local elections.

It is interesting to note that the volume co-efficients are correlated with the sentiment co-efficients which indicates

that topics that evoke positive sentiments also tend to attract a much higher volume and similarly negative topics tend to trigger much fewer comments. Overall, the topics and their co-efficients reinforce domain knowledge about political discourse in the US. This aspect is crucial because it suggests that the technique can be used in other domains where domain knowledge is difficult and expensive to obtain. Moreover, it provides a mechanism to reassure us that our notions are backed up by evidence in actual data.

Conclusion

We addressed the task of predicting the community response that is induced in political discourses. To this end, we tackled the tasks of determining the sentiment polarity of comments in blogs and the task of predicting the polarity based on the content of the blog post. Our experiments show that the community specific PMI method provides a more accurate picture of the sentiment in comments than the generic SentiWordNet technique. To tackle mixed-community response modeling, we introduce a new model - MCR-LDA, that identifies topics and the responses they evoke in different sub-communities. The newly proposed model does as well as community-specific models in predicting response sentiment and more importantly was used to jointly analyze two political blogs with very different political ideologies and identify topics that evoke very different sentiments in the two sub-communities.

References

Baccianella, A. E. S., and Sebastiani, F. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. Valletta, Malta: European Language Resources Association (ELRA).

Balasubramanyan, R.; Cohen, W. W.; Pierce, D.; and Redlawsk, D. P. 2011. What pushes their buttons? predicting comment polarity from the content of political blog posts. In *Proceedings of the Workshop on Language in Social Media (LSM 2011)*, 12–19. Portland, Oregon: Association for Computational Linguistics.

Blei, D., and McAuliffe, J. 2008. *Supervised Topic Models*. Cambridge, MA: MIT Press. 121–128.

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

Huang, L.-N., and Price, V. 2001. Motivations, goals, information search, and memory about political candidates. *Political Psychology* 22(4):pp. 665–692.

Isbell, L. M.; Ottati, V. C.; and Bruns, K. C. 2006. Affect and politics: Effects on judgment, processing, and information seeking. In Redlawsk, D., ed., *Feeling Politics: Emotion in Political Information Processing.* New York, USA: Palgrave Macmillan.

Kim, S.-y.; Taber, C. S.; and Lodge, M. 2008. A Computational Model of the Citizen as Motivated Reasoner: Modeling the Dynamics of the 2000 Presidential Election. *SSRN eLibrary*.

Lodge, M., and Taber, C. 2000. *Three Steps toward a Theory of Motivated Political Reasoning*. Cambridge University Press.

O'Connor, B.; Balasubramanyan, R.; Routledge, B. R.; and Smith, N. A. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*.

Redlawsk, D. P. 2002. Hot cognition or cool consideration? testing the effects of motivated reasoning on political decision making. *The Journal of Politics* 64(04):1021–1044.

Redlawsk, D. 2006. Motivated reasoning, affect, and the role of memory in voter decision-making. In Redlawsk, D., ed., *Feeling Politics: Emotion in Political Information Processing.* New York, USA: Palgrave Macmillan.

Tumasjan, A.; Sprenger, T. O.; Sandner, P. G.; and Welpe, I. M. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment. In Cohen, W. W., and Gosling, S., eds., *ICWSM*. The AAAI Press.

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, 417–424. Stroudsburg, PA, USA: Association for Computational Linguistics.

Yano, T.; Cohen, W.; and Smith, N. A. 2009. Predicting response to political blog posts with topic models. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*.