Subgroup Detection in Ideological Discussions

Amjad Abu-Jbara
EECS Department
University of Michigan
Ann Arbor, MI, USA
amjbara@umich.edu

Pradeep Dasigi
Department of Computer Science
Columbia University
New York, NY, USA
pd2359@columbia.edu

Mona Diab
Center for Computational Learning Systems
Columbia University
New York, NY, USA
mdiab@ccls.columbia.edu

Dragomir Radev
EECS Department
University of Michigan
Ann Arbor, MI, USA
radev@umich.edu

Abstract

The rapid and continuous growth of social networking sites has led to the emergence of many communities of communicating groups. Many of these groups discuss ideological and political topics. It is not uncommon that the participants in such discussions split into two or more subgroups. The members of each subgroup share the same opinion toward the discussion topic and are more likely to agree with members of the same subgroup and disagree with members from opposing subgroups. In this paper, we propose an unsupervised approach for automatically detecting discussant subgroups in online communities. We analyze the text exchanged between the participants of a discussion to identify the attitude they carry toward each other and towards the various aspects of the discussion topic. We use attitude predictions to construct an attitude vector for each discussant. We use clustering techniques to cluster these vectors and, hence, determine the subgroup membership of each participant. We compare our methods to text clustering and other baselines, and show that our method achieves promising results.

1 Introduction

Online forums discussing ideological and political topics are common[1]. When people discuss a disputed topic they usually split into subgroups. The members of each subgroup carry the same opinion

toward the discission topic. The member of a subgroup is more likely to show positive attitude to the members of the same subgroup, and negative attitude to the members of opposing subgroups.

For example, let us consider the following two snippets from a debate about the enforcement of a new immigration law in Arizona state in the United States:

(1) Discussant 1: Arizona immigration law is good. Illegal immigration is bad.

(2) Discussant 2: I totally disagree with you. Arizona immigration law is blatant racism, and quite unconstitutional.

In (1), the writer is expressing positive attitude regarding the immigration law and negative attitude regarding illegal immigration. The writer of (2) is expressing negative attitude towards the writer of (1) and negative attitude regarding the immigration law. It is clear from this short dialog that the writer of (1) and the writer of (2) are members of two opposing subgroups. Discussant 1 is supporting the new law, while Discussant 2 is against it.

In this paper, we present an unsupervised approach for determining the subgroup membership of each participant in a discussion. We use linguistic techniques to identify attitude expressions, their polarities, and their targets. The target of attitude could be another discussant or an entity mentioned in the discussion. We use sentiment analysis techniques to identify opinion expressions. We use named en-

---

[1]www.politicalforum.com, www.createdebate.com, www.forandagainst.com, etc

tity recognition and noun phrase chunking to identify the entities mentioned in the discussion. The opinion-target pairs are identified using a number of syntactic and semantic rules.

For each participant in the discussion, we construct a vector of attitude features. We call this vector the *discussant attitude profile*. The attitude profile of a discussant contains an entry for every other discussant and an entry for every entity mentioned in the discission. We use clustering techniques to cluster the attitude vector space. We use the clustering results to determine the subgroup structure of the discussion group and the subgroup membership of each participant.

The rest of this paper is organized as follows. Section 2 examines the previous work. We describe the data used in the paper in Section 2.4. Section 3 presents our approach. Experiments, results and analysis are presented in Section 4. We conclude in Section 5

2 Related Work

2.1 Sentiment Analysis

Our work is related to a huge body of work on sentiment analysis. Previous work has studied sentiment in text at different levels of granularity. The first level is identifying the polarity of individual words. Hatzivassiloglou and McKeown (1997) proposed a method to identify the polarity of adjectives based on conjunctions linking them. Turney and Littman (2003) used pointwise mutual information (PMI) and latent semantic analysis (LSA) to compute the association between a given word and a set of positive/negative seed words. Takamura et al. (2005) proposed using a spin model to predict word polarity. Other studies used WordNet to improve word polarity prediction (Hu and Liu, 2004a; Kamps et al., 2004; Kim and Hovy, 2004; Andreevskaia and Bergler, 2006). Hassan and Radev (2010) used a random walk model built on top of a word relatedness network to predict the semantic orientation of English words. Hassan et al. (2011) proposed a method to extend their random walk model to assist word polarity identification in other languages including Arabic and Hindi.

Other work focused on identifying the subjectivity of words. The goal of this work is to deter-

mine whether a given word is factual or subjective. We use previous work on subjectivity and polarity prediction to identify opinion words in discussions. Some of the work on this problem classifies words as factual or subjective regardless of their context (Wiebe, 2000; Hatzivassiloglou and Wiebe, 2000; Banea et al., 2008). Some other work noticed that the subjectivity of a given word depends on its context. Therefor, several studies proposed using contextual features to determine the subjectivity of a given word within its context (Riloff and Wiebe, 2003; Yu and Hatzivassiloglou, 2003; Nasukawa and Yi, 2003; Popescu and Etzioni, 2005).

The second level of granularity is the sentence level. Hassan et al. (2010) presents a method for identifying sentences that display an attitude from the text writer toward the text recipient. They define attitude as the mental position of one participant with regard to another participant. A very detailed survey that covers techniques and approaches in sentiment analysis and opinion mining could be found in (Pang and Lee, 2008).

2.2 Opinion Target Extraction

Several methods have been proposed to identify the target of an opinion expression. Most of the work have been done in the context of product reviews mining (Hu and Liu, 2004b; Kobayashi et al., 2007; Mei et al., 2007; Stoyanov and Cardie, 2008). In this context, opinion targets usually refer to product features (i.e. product components or attributes, as defined by Liu (2009)). In the work of Hu and Liu (2004b), they treat frequent nouns and noun phrases as product feature candidates. In our work, we extract as targets frequent noun phrases and named entities that are used by two or more different discussants. Scaffidi et al. (2007) propose a language model approach to product feature extraction. They assume that product features are mentioned more often in product reviews than they appear in general English text. However, such statistics may not be reliable when the corpus size is small.

In another related work, Jakob and Gurevych (2010) showed that resolving the anaphoric links in the text significantly improves opinion target extraction. In our work, we use anaphora resolution to improve opinion-target

Participant A posted:	I support Arizona because they have every right to do so. They are just upholding well-established federal law. All states should enact such a law.
Participant B commented on A's post:	I support the law because the federal government is either afraid or indifferent to the issue. Arizona has the right and the responsibility to protect the people of the State of Arizona. If this requires a possible slight inconvenience to any citizen so be it.
Participant C commented on B's post:	That is such a sad thing to say. You do realize that under the 14th Amendment, the very interaction of a police officer asking you to prove your citizenship is Unconstitutional? As soon as you start trading Constitutional rights for "security", then you've lost.

Table 1: Example posts from the Arizona Immigration Law thread

pairing as shown in Section 3 below.

2.3 Community Mining

Previous work also studied community mining in social media sites. Somasundaran and Wiebe (2009) presents an unsupervised opinion analysis method for debate-side classification. They mine the web to learn associations that are indicative of opinion stances in debates and combine this knowledge with discourse information. Anand et al. (2011) present a supervised method for stance classification. They use a number of linguistic and structural features such as unigrams, bigrams, cue words, repeated punctuation, and opinion dependencies to build a stance classification model. This work is limited to dual sided debates and defines the problem as a classification task where the two debate sides are know beforehand. Our work is characterized by handling multi-side debates and by regarding the problem as a clustering problem where the number of sides is not known by the algorithm. This work also utilizes only discussant-to-topic attitude predictions for debate-side classification. Out work utilizes both discussant-to-topic and discussant-to-discussant attitude predictions.

In another work, Kim and Hovy (2007) predict the results of an election by analyzing discussion threads in online forums that discuss the elections. They use a supervised approach that uses unigrams, bigrams, and trigrams as features. In contrast, our work is unsupervised and uses different types information. Moreover, although this work is related to ours at the goal level, it does not involve any opinion analysis.

Another related work classifies the speakers side in a corpus of congressional floor debates, using the speakers final vote on the bill as a labeling for side (Thomas et al., 2006; Bansal et al., 2008;

Yessenalina et al., 2010). This work infers agreement between speakers based on cases where one speaker mentions another by name, and a simple algorithm for determining the polarity of the sentence in which the mention occurs. This work shows that even with the resulting sparsely connected agreement structure, the MinCut algorithm can improve over stance classification based on textual information alone. This work also requires that the debate sides be known by the algorithm and it only identifies discussant-to-discussant attitude. In our experiments below we show that identifying both discussant-to-discussant and discussant-to-topic attitudes achieves better results.

2.4 Data

In this section, we describe the datasets used in this paper. We use three different datasets. The first dataset (*politicalforum*, henceforth) consists of 5,743 posts collected from a political forum[2]. All the posts are in English. The posts cover 12 disputed political and ideological topics. The discussants of each topic were asked to participate in a poll. The poll asked them to determine their stance on the discussion topic by choosing one item from a list of possible arguments. The list of participants who voted for each argument was published with the poll results. Each poll was accompanied by a discussion thread. The people who participated in the poll were allowed to post text to that thread to justify their choices and to argue with other participants. We collected the votes and the discussion thread of each poll. We used the votes to identify the subgroup membership of each participant.

The second dataset (*createdebate*, henceforth) comes from an online debating site [3]. It consists of

---

[2]http://www.politicalforum.com
[3]http://www.createdebate.com

Source	Topic	Question	#Sides	#Posts	#Participants
Politicalforum	Arizona Immigration Law	Do you support Arizona in its decision to enact their Immigration Enforcement law?	2	738	59
	Airport Security	Should we pick muslims out of the line and give additional scrutiny/screening?	4	735	69
	Vote for Obama	Will you vote for Obama in the 2012 Presidential elections?	2	2599	197
Createdebate	Evolution	Has evolution been scientifically proved?	2	194	98
	Social networking sites	It is easier to maintain good relationships in social networking sites such as Facebook.	2	70	31
	Abortion	Should abortion be banned	3	477	70
Wikipedia	Ireland	Misleading description of Irland island partition	3	40	10
	South Africa Goverment	Was the current form of South African government born in May 1910?	3	23	5
	Oil Spill	Obama's response to gulf oil spill	3	30	12

Table 2: Example threads from our three datasets

30 debates containing a total of 2,712 posts. Each debate is about one topic. The description of each debate states two or more positions regarding the debate topic. When a new participant enters the discussion, she explicitly picks a position and posts text to support it, support a post written by another participant who took the same position, or to dispute a post written by another participant who took an opposing position. We collected the discussion thread and the participant positions for each debate.

The third dataset (*wikipedia*, henceforth) comes from the Wikipedia[4] discussion section. When a topic on Wikipedia is disputed, the editors of that topic start a discussion about it. We collected 117 Wikipeida discussion threads. The threads contains a total of 1,867 posts.

The *politicalforum* and *createdebate* datasets are self labeled as described above. To annotate the Wikipedia data, we asked an expert annotator (a professor in sociolinguistics who is not one of the authors) to read each of the Wikipedia discussion threads and determine whether the discussants split into subgroups in which case he was asked to determine the subgroup membership of each discussant.

Table 2 lists few example threads from our three datasets. Table 1 shows a portion of discussion thread between three participants about enforcing a new immigration law in Arizona. This thread appeared in the polictalforum dataset. The text posted by the three participants indicates that A's position

is with enforcing the law, that B agrees with A, and that C disagrees with both. This means that A and B belong to the same opinion subgroup, while belongs to an opposing subgroup.

We randomly selected 6 threads from our datasets (2 from *politicalforum*, 2 from *createdebate*, and 2 from *Wikipedia*) and used them as development set. This set was used to develop our approach.

3 Approach

In this section, we describe a system that takes a discussion thread as input and outputs the subgroup membership of each discussant. Figure 1 illustrates the processing steps performed by our system to detect subgroups. In the following subsections we describe the different stages in the system pipeline.

3.1 Thread Parsing

We start by parsing the thread to identify posts, participants, and the reply structure of the thread (i.e. who replies to whom). In the datasets described in Section 2.4, all this information was explicitly available in the thread. We tokenize the text of each post and split it into sentences using CLAIRLib (Abu-Jbara and Radev, 2011).

3.2 Opinion Word Identification

The next step is to identify the words that express opinion and determine their polarity (positive or negative). Lehrer (1974) defines word polarity as the direction the word deviates to from the norm. We

---
[4]http://www.wikipedia.com

use OpinionFinder (Wilson et al., 2005a) to identify polarized words and their polarities.

The polarity of a word is usally affected by the context in which it appears. For example, the word *fine* is positive when used as an adjective and negative when used as a noun. For another example, a positive word that appears in a negated context becomes negative. OpinionFinder uses a large set of features to identify the contextual polarity of a given polarized word given its isolated polarity and the sentence in which it appears (Wilson et al., 2005b). Snippet (3) below shows the result of applying this step to snippet (1) above (*O* means neutral; *POS* means positive; *NEG* means negative).

(3) Arizona/O Immigration/O law/O good/POS ./O Illegal/O immigration/O bad/NEG ./O

3.3 Target Identification

The goal of this step is to identify the possible targets of opinion. A target could be another discussant or an entity mentioned in the discussion. When the target of opinion is another discussant, either the discussant name is mentioned explicitly or a second person pronoun is used to indicate that the opinion is targeting the recipient of the post. For example, in snippet (2) above the second person pronoun *you* indicates that the opinion word *disagree* is targeting *Discussant 1*, the recipient of the post.

The target of opinion can also be an entity mentioned in the discussion. We use two methods to identify such entities. The first method uses shallow parsing to identify noun groups (NG). We use the Edinburgh Language Technology Text Tokenization Toolkit (LT-TTT) (Grover et al., 2000) for this purpose. We consider as an entity any noun group that is mentioned by at least two different discussants. We replace each identified entity with a unique placeholder ($ENTITY_{ID}$). For example, the noun group *Arizona immigration law* is mentioned by *Discussant 1* and *Discussant 2* in snippets 1 and 2 above respectively. Therefore, we replace it with a placehold as illustrated in snippets (4) and (5) below.

(4) Discussant 1: $ENTITY_1$ is good. Illegal im-

NER	NP Chunking
Barack Obama	the Republican nominee
Middle East	the maverick economists
Bush	conservative ideologues
Bob McDonell	the Nobel Prize
Iraq	Federal Government

Table 3: Some of the entities identified using NER and NP Chunking in a discussion thread about the US 2012 elections

migration is bad.

(5) Discussant 2: I totally disagree with you. $ENTITY_1$ is blatant racism, and quite unconstitutional.

We only consider as entities noun groups that contain two words or more. We impose this requirement because individual nouns are very common and regarding all of them as entities will introduce significant noise.

In addition to this shallow parsing method, we also use named entity recognition (NER) to identify more entities. We use the Stanford Named Entity Recognizer (Finkel et al., 2005) for this purpose. It recognizes three types of entities: person, location, and organization. We impose no restrictions on the entities identified using this method. Again, we replace each distinct entity with a unique placeholder. The final set of entities identified in a thread is the union of the entities identified by the two aforementioned methods. Table 3

Finally, a challenge that always arises when performing text mining tasks at this level of granularity is that entities are usually expressed by anaphorical pronouns. Previous work has shown that For example, the following snippet contains an explicit mention of the entity *Obama* in the first sentence, and then uses a pronoun to refer to the same entity in the second sentence. The opinion word *unbeatable* appears in the second sentence and is syntactically related to the pronoun *He*. In the next subsection, it will become clear why knowing which entity does the pronoun *He* refers to is essential for opinion-target pairing.
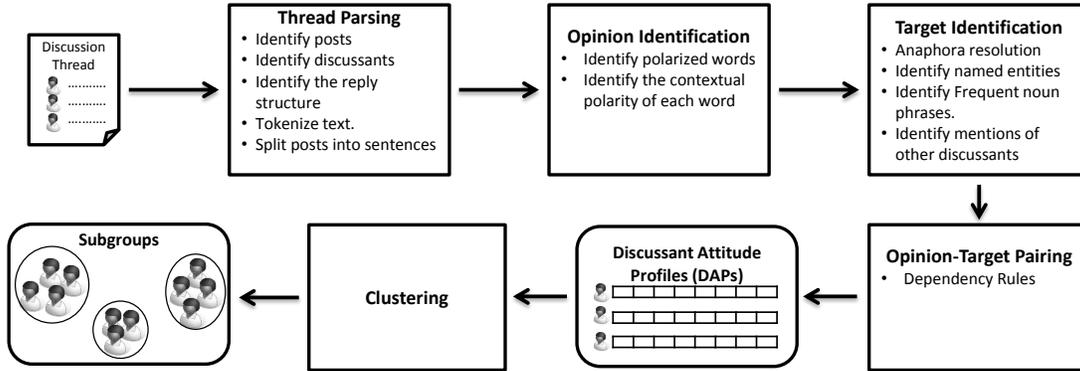
(6) It doesn't matter whether you vote for Obama.

Figure 1: An overview of the subgroups detection system

He is unbeatable.

Jakob and Gurevych (2010) showed experimentally that resolving the anaphoric links in the text significantly improves opinion target extraction. We use the Beautiful Anaphora Resolution Toolkit (BART) (Versley et al., 2008) to resolve all the anaphoric links within the text of each post separately. The result of applying this step to snippet (6) is:

(6) It doesn't matter whether you vote for Obama. Obama is unbeatable.

Now, both mentions of *Obama* will be recognized by the Stanford NER system and will be identified as one entity.

3.4 Opinion-Target Pairing

At this point, we have all the opinion words and the potential targets identified separately. The next step is to determine which opinion word is targeting which target. We propose a rule based approach for opinion-target pairing. Our rules are based on the dependency relations that connect the words in a sentence. We use the Stanford Parser (Klein and Manning, 2003) to generate the dependency parse tree of each sentence in the thread. An opinion word and a target form a pair if they stratify at least one of our dependency rules. Table 4 illustrates some

of these rules [5]. The rules basically examine the types of the dependencies on the shortest path that connect the opinion word and the target in the dependency parse tree. It has been shown in previous work on relation extraction that the shortest dependency path between any two entities captures the information required to assert a relationship between them (Bunescu and Mooney, 2005).

If a sentence S in a post written by participant P_i contains an opinion word OP_j and a target TR_k, and if the opinion-target pair satisfies one of our dependency rules, we say that P_i expresses an attitude towards TR_k. The polarity of the attitude is determined by the polarity of OP_j. We represent this as $P_i \xrightarrow{+} TR_k$ if OP_j is positive and $P_i \xrightarrow{-} TR_k$ if OP_j is negative.

It is likely that the same participant P_i express sentiment toward the same target TR_k multiple times in different sentences in different posts. We keep track of the counts of all the instances of positive/negative attitude P_i expresses toward TR_k. We represent this as $P_i \xrightarrow[n-]{m+} TR_k$ where m (n) is the number of times P_i expressed positive (negative) attitude toward TR_k.

3.5 Discussant Attitude Profile

We propose a representation of discussantsáttitudes towards the identified targets in the discussion thread. As stated above, a target could be another discussant or an entity mentioned in the discussion.

---

[5]The code will be made publicly available at the time of publication

ID	Rule	In Words	Example
R1	$OP \rightarrow nsubj \rightarrow TR$	The target *TR* is the nominal subject of the opinion word *OP*	ENTITY1$_{TR}$ is good$_{OP}$.
R2	$OP \rightarrow dobj \rightarrow TR$	The target *T* is a direct object of the opinion *OP*	I hate$_{OP}$ ENTITY2$_{TR}$
R3	$OP \rightarrow prep\_* \rightarrow TR$	The target *TR* is the object of a preposition that modifies the opinion word *OP*	I totally disagree$_{OP}$ with you$_{TR}$.
R4	$TR \rightarrow amod \rightarrow OP$	The opinion is an adjectival modifier of the target	The bad$_{OP}$ ENTITY3$_{TR}$ is spreading lies
R5	$OP \rightarrow nsubjpass \rightarrow TR$	The target *TR* is the nominal subject of the passive opinion word *OP*	ENTITY4$_{TR}$ is hated$_{OP}$ by everybody.
R6	$OP \rightarrow prep\_* \rightarrow poss \rightarrow TR$	The opinion word *OP* connected through a $prep\_*$ relation as in $R2$ to something possessed by the target *TR*	The main flaw$_{OP}$ in your$_{TR}$ analysis is that it's based on wrong assumptions.
R7	$OP \rightarrow dobj \rightarrow poss \rightarrow TR$	The target *TR* possesses something that is the direct object of the opinion word *OP*	I like$_{OP}$ ENTITY5$_{TR}$'s brilliant ideas.
R8	$OP \rightarrow csubj \rightarrow nsubj \rightarrow TR$	The opinon word *OP* is a causal subject of a phrase that has the target *TR* as its nominal subject	What ENTITY6$_{TR}$ announced was misleading$_{OP}$.

Table 4: Examples of the dependency rules used for opinion-target pairing.

Our representation is a vector containing numerical values. The values correspond to the counts of positive/negative attitudes expressed by the discussant toward each of the targets. We call this vector the *discussant attitude profile (DAP)*. We construct a DAP for every discussant. Given a discussion thread with d discussants and e entity targets, each attitude profile vector has $n = (d + e) * 3$ dimensions. In other words, each target (discussant or entity) has three corresponding values in the DAP: 1) the number of times the discussant expressed positive attitude toward the target, 2) the number of times the discussant expressed a negative attitude towards the target, and 3) the number of times the the discussant interacted with or mentioned the target. It has to be noted that these values are not symmetric since the discussions explicitly denote the source and the target of each post.

3.6 Clustering

At this point, we have an attitude profile (or vector) constructed for each discussant. Our goal is to use these attitude profiles to determine the subgroup membership of each discussant. We can achieve this goal by noticing that the attitude profiles of discussants who share the same opinion are more likely to be similar to each other than to the attitude profiles of discussants with opposing opinions. This suggests that clustering the attitude vector space will achieve the goal and split the discussants into subgroups according to their opinion.

4 Evaluation

In this section, we present several levels of evaluation of our system. First, we compare our system to baseline systems. Second, we study how the choice of the clustering algorithm impacts the results. Third, we study the impact of each component in our system on the performance. All the results reported in this section that show difference in the performance are statistically significant at the 0.05 level (as indicated by a 2-tailed paired t-test). Before describing the experiments and presenting the results, we first describe the evaluation metrics we use.

4.0.1 Evaluation Metrics

We use two evaluation metrics to evaluate subgroups detection accuracy: Purity and Entropy. To compute Purity (Manning et al., 2008), each cluster is assigned the class of the majority vote within the cluster, and then the accuracy of this assignment is measured by dividing the number of correctly assigned members by the total number of instances. It can be formally defined as:

$$purity(\Omega, C) = \frac{1}{N} \sum_k \max_j |\omega_k \cap c_j| \quad (1)$$

where $\Omega = \{\omega_1, \omega_2, ..., \omega_k\}$ is the set of clusters and $C = \{c_1, c_2, ..., c_J\}$ is the set of classes. ω_k is interpreted as the set of documents in ω_k and c_j as

the set of documents in c_j. The purity increases as the quality of clustering improves.

The second metric is Entropy. The Entropy of a cluster reflects how the members of the k distinct subgroups are distributed within each resulting cluster; the global quality measure is computed by averaging the entropy of all clusters:

$$Entropy = -\sum^{j} \frac{n_j}{n} \sum^{i} P(i,j) \times log_2 P(i,j)$$
$$(2)$$

where $P(i,j)$ is the probability of finding an element from the category i in the cluster j, n_j is the number of items in cluster j, and n the total number of items in the distribution. In contrast to purity, the entropy decreases as the quality of clustering improves.

4.1 Comparison to Baseline Systems

We compare our system (DAPC) that was described in Section 3 to two baseline methods. The first baseline (GC) uses graph clustering to partition a network based on the interaction frequency between participants. We build a graph where each node represents a participant. Edges link participants if they exchange posts, and edge weights are based on the number of interactions. We tried two methods for clustering the resulting graph: spectral partitioning (Luxburg, 2007) and a hierarchical agglomeration algorithm which works by greedily optimizing the modularity for graphs (Clauset et al., 2004).

The second baseline (TC) is based on the premise that the member of the same subgroup are more likely to use vocabulary drawn from the same language model. We collect all the text posted by each participant and create a tf-idf representations of the text in a high dimensional vector space. We then cluster the vector space to identify subgroups. We use k-means (MacQueen, 1967) as our clustering algorithm in this experiment (comparison of various clustering algorithms is presented in the next subsection). The distances between vectors are Eculidean distances. Table 5 shows that our system performs significantly better the baselines on the three datasets in terms of both the purity (P) and the entropy (E) (notice that lower entropy values indicate better clustering). The values reported are the

Method	Createdebate		Politicalforum		Wikipedia	
	P	E	P	E	P	E
GC - Spectral	0.50	0.85	0.50	0.88	0.49	0.89
GC - Hierarchical	0.48	0.86	0.47	0.89	0.49	0.87
TC - kmeans	0.51	0.84	0.49	0.88	0.52	0.85
DAPC - kmeans	**0.64**	**0.68**	**0.61**	**0.80**	**0.66**	**0.55**

Table 5: Comparison to baseline systems

Method	Createdebate		Politicalforum		Wikipedia	
	P	E	P	E	P	E
DAPC - EM	0.63	0.71	0.61	0.82	0.63	0.61
DAPC - FF	0.63	0.70	0.60	0.83	0.64	0.59
DAPC - kmeans	0.64	0.68	0.61	0.80	0.66	0.55

Table 6: Comparison of different clustering algorithms

average results of the threads of each dataset. We believe that the baselines performed poorly because the interaction frequency and the text similarity are not key factors in identifying subgroup structures. Many people would respond to people they disagree with more, while others would mainly respond to people they agree with most of the time. Also, people in opposing subgroups tend to use very similar text when discussing the same topic and hence text clustering does not work as well.

4.2 Choice of the clustering algorithm

We experimented with three different clustering algorithms: expectation maximization (EM), and k-means (MacQueen, 1967), and FarthestFirst (FF) (Hochbaum and Shmoys, 1985; Dasgupta, 2002). As we did in the previous subsection, we use Eculidean distance to measure the distance between vectors All the system (DAP) components are included as described in Section 3. The purity and entropy values using each algorithm are shown in Table 6. Although k-means seems to be performing slightly better than other algorithms, the differences in the results are not significant. This indicates that the choice of the clustering algorithm does not have a noticeable impact on the results. We also experimented with using Manhattan distance and cosine similarity instead of Euclidean distance to measure the distance between attitude vectors. We noticed that the choice of the distance does not have significant impact on the results as well.

4.3 Component Evaluation

In this subsection, we evaluate the impact of the different components in the pipeline on the system performance. We do that by removing each component from the pipeline and measuring the change in performance. We perform the following experiments: 1) We run the full system with all its components included (DAPC). 2) We run the system and include only discussant-to-discussant attitude features in the attitude vectors (DAPC-DD). 3) We include only discussant-to-entity attitude features in the attitude vectors (DAPC-DE). 4) We include only sentiment features in the attitude vector; i.e. we exclude the interaction count features (DAPC-SE). 5) We include only interaction count features to the attitude vector; i.e. we exclude sentiment features (DAPC-INT). 6) We skip the anaphora resolution step in the entity identification component (DAPC-NO_AR). 7) We only use named entity recognition to identify entity targets; i.e. we exclude the entities identified through noun phrasing chunking (DAPC-NER). 8) Finally, we only noun phrase chunking to identify entity targets (DAPC-NP). In all these experiments k-means is used for clustering and the number of clusters is set as explained in the previous subsection.

The results show that all the components in the system contribute to better performance of the system. We notice from the results that the performance of the system drops significantly if sentiment features are not included. This is result corroborates our hypothesis that interaction features are not sufficient factors for detecting rift in discussion groups. Including interaction features improve the performance (although not by a big difference) because they help differentiate between the case where participants A and B never interacted with each other and the case where they interact several time but never posted text that indicate difference in opinion between them. We also notice that the performance drops significantly in DAPC-DD and DAPC-DD which also supports our hypotheses that both the sentiment discussants show toward one another and the sentiment they show toward the aspects of the discussed topic are important for the task. Although using both named entity recognition (NER) and noun phrase chunking achieves better results, it

Method	Createdebate		Politicalforum		Wikipedia	
	P	E	P	E	P	E
DAPC	**0.64**	**0.68**	**0.61**	**0.80**	**0.66**	**0.55**
DAPC-DD	0.59	0.77	0.57	0.86	0.62	0.61
DAPC-DE	0.60	0.69	0.58	0.84	0.58	0.78
DAPC-SE	0.62	0.70	0.60	0.83	0.61	0.62
DAPC-INT	0.54	0.88	0.52	0.91	0.57	0.85
DAPC-NO_AR	0.62	0.72	0.60	0.84	0.64	0.60
DAPC-NER	0.61	0.71	0.58	0.86	0.63	0.59
DAPC-NP	0.63	0.75	0.59	0.84	0.65	0.62

Table 7: Impact of system components on the performance

can also be noted from the results that NER contributes more to the system performance. Finally, the results support Jakob and Gurevych (2010) findings that anaphora resolution aids opinion mining systems.

5 Conclusions

In this paper, we presented an approach for subgroup detection in ideological discussions. Our system uses linguistic analysis techniques to identify the attitude the participants of online discussions carry toward each other and toward the aspects of the discussion topic. Attitude prediction as well as interaction frequency to construct an attitude vector for each participant. The attitude vectors of discussants are then clustered to form subgroups. Our experiments showed that our system outperforms text clustering and interaction graph clustering. We also studied the contribution of each component in our system to the overall performance.

References

Amjad Abu-Jbara and Dragomir Radev. 2011. Clairlib: A toolkit for natural language processing, information retrieval, and network analysis. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 121–126, Portland, Oregon, June. Association for Computational Linguistics.

Pranav Anand, Marilyn Walker, Rob Abbott, Jean E. Fox Tree, Robeson Bowmani, and Michael Minor. 2011. Cats rule and dogs drool!: Classifying stance in online debate. In *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011)*, pages 1–9, Portland, Oregon, June. Association for Computational Linguistics.

Alina Andreevskaia and Sabine Bergler. 2006. Mining wordnet for fuzzy sentiment: Sentiment tag extraction from wordnet glosses. In *EACL'06*.

Carmen Banea, Rada Mihalcea, and Janyce Wiebe. 2008. A bootstrapping method for building subjectivity lexicons for languages with scarce resources. In *LREC'08*.

Mohit Bansal, Claire Cardie, and Lillian Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework.

Razvan Bunescu and Raymond Mooney. 2005. A shortest path dependency kernel for relation extraction. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 724–731, Vancouver, British Columbia, Canada, October. Association for Computational Linguistics.

Aaron Clauset, Mark E. J. Newman, and Cristopher Moore. 2004. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111.

Sanjoy Dasgupta. 2002. Performance guarantees for hierarchical clustering. In *15th Annual Conference on Computational Learning Theory*, pages 351–363. Springer.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 363–370, Stroudsburg, PA, USA. Association for Computational Linguistics.

Claire Grover, Colin Matheson, Andrei Mikheev, and Marc Moens. 2000. Lt ttt - a flexible tokenisation tool. In *In Proceedings of Second International Conference on Language Resources and Evaluation*, pages 1147–1154.

Ahmed Hassan and Dragomir Radev. 2010. Identifying text polarity using random walks. In *ACL'10*.

Ahmed Hassan, Vahed Qazvinian, and Dragomir Radev. 2010. What's with the attitude?: identifying sentences with attitude in online discussions. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1245–1255.

Ahmed Hassan, Amjad AbuJbara, Rahul Jha, and Dragomir Radev. 2011. Identifying the semantic orientation of foreign words. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 592–597, Portland, Oregon, USA, June. Association for Computational Linguistics.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *EACL'97*, pages 174–181.

Vasileios Hatzivassiloglou and Janyce Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *COLING*, pages 299–305.

Hochbaum and Shmoys. 1985. A best possible heuristic for the k-center problem. *Mathematics of Operations Research*, 10(2):180–184.

Minqing Hu and Bing Liu. 2004a. Mining and summarizing customer reviews. In *KDD'04*, pages 168–177.

Minqing Hu and Bing Liu. 2004b. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Niklas Jakob and Iryna Gurevych. 2010. Using anaphora resolution to improve opinion target identification in movie reviews. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 263–268, Uppsala, Sweden, July. Association for Computational Linguistics.

Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. 2004. Using wordnet to measure semantic orientations of adjectives. In *National Institute for*, pages 1115–1118.

Soo-Min Kim and Eduard Hovy. 2004. Determining the sentiment of opinions. In *COLING*, pages 1367–1373.

Dan Klein and Christopher D. Manning. 2003. Accurate unlexicalized parsing. In *IN PROCEEDINGS OF THE 41ST ANNUAL MEETING OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, pages 423–430.

Nozomi Kobayashi, Kentaro Inui, and Yuji Matsumoto. 2007. Extracting aspect-evaluation and aspect-of relations in opinion mining. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL*.

Adrienne Lehrer. 1974. Semantic fields and lezical structure. North Holland, Amsterdam and New York.

Bing Liu. 2009. *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer, 1st ed. 2007. corr. 2nd printing edition, January.

Ulrike Luxburg. 2007. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, December.

J. B. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In L. M. Le Cam and J. Neyman, editors, *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. University of California Press.

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schtze. 2008. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai. 2007. Topic sentiment mixture: modeling facets and opinions in weblogs. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 171–180, New York, NY, USA. ACM.

Soo min Kim and Eduard Hovy. 2007. Crystal: Analyzing predictive opinions on the web. In *In EMNLP-CoNLL 2007*.

Tetsuya Nasukawa and Jeonghee Yi. 2003. Sentiment analysis: capturing favorability using natural language processing. In *K-CAP '03: Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *HLT-EMNLP'05*, pages 339–346.

Ellen Riloff and Janyce Wiebe. 2003. Learning extraction patterns for subjective expressions. In *EMNLP'03*, pages 105–112.

Swapna Somasundaran and Janyce Wiebe. 2009. Recognizing stances in online debates. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 226–234, Suntec, Singapore, August. Association for Computational Linguistics.

Veselin Stoyanov and Claire Cardie. 2008. Topic identification for fine-grained opinion analysis. In *In Coling*.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *ACL'05*, pages 133–140.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *In Proceedings of EMNLP*, pages 327–335.

Peter Turney and Michael Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21:315–346.

Yannick Versley, Simone Paolo Ponzetto, Massimo Poesio, Vladimir Eidelman, Alan Jern, Jason Smith, Xiaofeng Yang, and Alessandro Moschitti. 2008. Bart: A modular toolkit for coreference resolution. In *Proceedings of the ACL-08: HLT Demo Session*, pages 9–12, Columbus, Ohio, June. Association for Computational Linguistics.

Janyce Wiebe. 2000. Learning subjective adjectives from corpora. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pages 735–740.

Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan. 2005a. Opinionfinder: a system for subjectivity analysis. In *Proceedings of HLT/EMNLP on Interactive Demonstrations*, HLT-Demo '05, pages 34–35, Stroudsburg, PA, USA. Association for Computational Linguistics.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005b. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT/EMNLP'05*, Vancouver, Canada.

Ainur Yessenalina, Yisong Yue, and Claire Cardie. 2010. Multi-level structured models for document-level sentiment classification. In *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP*.

Hong Yu and Vasileios Hatzivassiloglou. 2003. Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences. In *EMNLP'03*, pages 129–136.