

Contrasting Opposing Views of News Articles on Contentious Issues

Souneil Park¹, KyungSoon Lee², Junehwa Song¹

¹Korea Advanced Institute of
Science and Technology
291 Daehak-ro, Yuseong-gu,
Daejeon, Republic of Korea

{spark, junesong}@nclab.kaist.ac.kr

²Chonbuk National
University
664-14 1ga Deokjin-dong Jeonju,
Jeonbuk, Republic of Korea
selfsolee@chonbuk.ac.kr

Abstract

We present disputant relation-based method for classifying news articles on contentious issues. We observe that the disputants of a contention are an important feature for understanding the discourse. It performs unsupervised classification on news articles based on disputant relations, and helps readers intuitively view the articles through the opponent-based frame. The readers can attain balanced understanding on the contention, free from a specific biased view. We applied a modified version of HITS algorithm and an SVM classifier trained with pseudo-relevant data for article analysis.

1 Introduction

The coverage of contentious issues of a community is an essential function of journalism. Contentious issues continuously arise in various domains, such as politics, economy, environment; each issue involves diverse participants and their different complex arguments. However, news articles are frequently biased and fail to fairly deliver conflicting arguments of the issue. It is difficult for ordinary readers to analyze the conflicting arguments and understand the contention; they mostly perceive the issue passively, often through a single article. Advanced news delivery models are required to increase awareness on conflicting views.

In this paper, we present *disputant relation-based method* for classifying news articles on con-

tentious issues. We observe that the disputants of a contention, i.e., people who take a position and participate in the contention such as politicians, companies, stakeholders, civic groups, experts, commentators, etc., are an important feature for understanding the discourse. News producers primarily shape an article on a contention by selecting and covering specific disputants (Baker. 1994). Readers also intuitively understand the contention by identifying who the opposing disputants are.

The method helps readers intuitively view the news articles through the *opponent-based frame*. It performs classification in an unsupervised manner: it dynamically identifies opposing disputant groups and classifies the articles according to their positions. As such, it effectively helps readers contrast articles of a contention and attain balanced understanding, free from specific biased viewpoints.

The proposed method differs from those used in related tasks as it aims to perform classification under the opponent-based frame. Research on sentiment classification and debate stance recognition takes a topic-oriented view, and attempts to perform classification under the ‘positive vs. negative’ or ‘for vs. against’ frame for the given topic, e.g., positive vs. negative about iPhone.

However, such frames are often not appropriate for classifying news articles of a contention. The coverage of a contention often spans over different topics (Miller. 2001). For the contention on the health care bill, an article may discuss the enlarged coverage whereas another may discuss the increase of insurance premiums. In addition, we observe that opposing arguments of a contention are often complex to classify under these frames. For exam-

ple, in a political contention on holding a referendum on the Sejong project¹, the opposition parties strongly opposed and criticized the president office. Meanwhile, the president office argued that they were not considering holding the referendum and the contention arose from a misunderstanding. In such a case, it is difficult to classify any argument to the “positive” category of the frame.

We demonstrate that the opponent-based frame is clear and effective for contrasting opposing views of contentious issues. For the contention on the referendum, ‘president office vs. opposition parties’ provides an intuitive frame to understand the contention. The frame does not require the documents to discuss common topics nor the opposing arguments to be positive vs. negative.

Under the proposed frame, it becomes important to analyze which side is more centrally covered in an article. Unlike debate posts or product reviews news articles, in general, do not take a position explicitly (except a few types such as editorials). They instead quote a specific side, elaborate them, and provide supportive facts. On the other hand, the opposing disputants compete for news coverage to influence more readers and gain support (Miller et al. 2001). Thus, the method focuses on identifying the disputants of each side and classifying the articles based on the side it covers.

We applied a modified version of HITS algorithm to identify the key opponents of an issue, and used disputant extraction techniques combined with an SVM classifier for article analysis. We observe that the method achieves acceptable performance for practical use with basic language resources and tools, i.e., Named Entity Recognizer (Lee et al. 2006), POS tagger (Shim et al. 2002), and a translated positive/negative lexicon. As we deal with non-English (Korean) news articles, it is difficult to obtain rich resources and tools, e.g., WordNet, dependency parser, annotated corpus such as MPQA. When applied to English, we believe the method could be further improved by adopting them.

2 Background and Related Work

Research has been made on sentiment classification in document-level (Turney et al., 2002, Pang et al., 2002, Seki et al. 2008, Ounis et al. 2006). It aims to automatically identify and classify the sen-

timent of documents into positive or negative. Opinion summarization aims a similar goal, to identify different opinions on a topic and generate summaries of them. Paul et al. (2010) developed an unsupervised method for generating summaries of contrastive opinions on a common topic. These works make a number of assumptions that are difficult to apply to the discourse of contentious news issues. They assume that the input documents have a common opinion target, e.g., a movie. Many of them primarily deal with documents which explicitly reveal opinions on the selected target, e.g., movie reviews. They usually apply one static classification frame, positive vs. negative, to the topic.

The discourse of contentious issues in news articles show different characteristics from that studied in the sentiment classification tasks. First, the opponents of a contentious issue often discuss different topics, as discussed in the example above. Research in mass communication has showed that opposing disputants talk across each other, not by dialogue, i.e., they martial different facts and interpretations rather than to give different answers to the same topics (Schon et al., 1994).

Second, the frame of argument is not fixed as ‘positive vs. negative’. We frequently observed both sides of a contention articulating negative arguments attacking each other. The forms of arguments are also complex and diverse to classify them as positive or negative; for example, an argument may just neglect the opponent’s argument without positive or negative expressions, or emphasize a different discussion point.

In addition, a position of a contention can be communicated without explicit expression of opinion or sentiment. It is often conveyed through objective sentences that include carefully selected facts. For example, a news article can cast a negative light on a government program simply by covering the increase of deficit caused by it.

A number of works deal with debate stance recognition, which is a closely related task. They attempt to identify a position of a debate, such as ideological (Somasundaran et al., 2010, Lin et al., 2006) or product comparison debate (Somasundaran et al., 2009). They assume a debate frame, which is similar to the frame of the sentiment classification task, i.e., for vs. against the debate topic. All articles of a debate in their corpus cover a coherent debate topic, e.g., iPhone vs. Blackberry, and explicitly express opinions for or

¹ http://www.koreatimes.co.kr/www/news/nation/2010/07/116_61649.html

against to the topic, e.g., for or against iPhone or Blackberry. The proposed methods assume that the debate frame is known apriori. This debate frame is often not appropriate for contentious issues for similar reasons as the positive/negative frame. In contrast, our method does not assume a fixed debate frame, and rather develops one based on the opponents of the contention at hand.

The news corpus is also different from the debate corpus. News articles of a contentious issue are more diverse than debate articles conveying explicit argument of a specific side. There are news articles which cover both sides, facts without explicit opinions, and different topics unrelated to the arguments of either side.

Several works have used the relation between speakers or authors for classifying their debate stance (Thomas et al., 2006, Agrawal et al., 2003). However, these works also assume the same debate frame and use the debate corpus, e.g., floor debates in the House of Representatives, online debate forums. Their approaches are also supervised, and require training data for relation analysis, e.g., voting records of congresspeople.

3 Argument Frame Comparison

Establishing an appropriate argument frame is important. It provides a framework which enable readers to intuitively understand the contention. It also determines how classification methods should classify articles of the issue.

We conducted a user study to compare the opponent-based frame and the positive (for) vs. negative (against) frame. In the experiment, multiple human annotators classified the same set of news articles under each of the two frames. We compared which frame is clearer for the classification, and more effective for exposing opposing views.

We selected 14 contentious issues from Naver News (a popular news portal in Korea) issue archive. We randomly sampled about 20 articles per each issue, for a total of 250 articles. The selected issues range over diverse domains such as politics, local, diplomacy, economy; to name a few for example, the contention on the 4 river project, of which the key opponents are the government vs. catholic church; the entrance of big retailers to the supermarket business, of which the key opponents are the small store owners vs. big retail companies; the refusal to approve an integrated civil servants'

union, of which the key opponents are government vs. Korean government employees' union.

We use an internationally known contention, i.e., the dispute about the Cheonan sinking incident, as an example to give more details on the disputants. Our data set includes 25 articles that were published after the South Korea's announcement of their investigation result. Many disputants appear in the articles, e.g., South Korean Government, South Korea defense secretary, North Korean Government, United States officials, Chinese experts, political parties of South Korea, etc.

Three annotators performed the classification. All of them were students. For impartiality, two of them were recruited from outside the team, who were not aware of this research.

The annotators performed two subtasks for classification. As for the positive vs. negative frame, first, we asked them to designate the main topic of the contention. Second, they classified the articles which mainly deliver arguments for the topic to the "positive" category and those delivering arguments against the topic to the "negative" category. The articles are classified to the "Other" category if they do not deal with the main topic nor cover positive or negative arguments.

As for the opponent-based frame, first, we asked them to designate the competing opponents. Second, we asked to classify articles to a specific side if the articles cover only the positions, arguments, or information supportive of that side or if they cover information detrimental or criticism to its opposite side. Other articles were classified to the "Other" category. Examples of this category include articles covering both sides fairly, describing general background or implications of the issue.

Issue #	Free-marginal kappa		Issue #	Free-marginal kappa	
	Pos.-Neg.	Opponent		Pos.-Neg.	Opponent
1	0.83	0.67	8	0.26	0.58
2	0.57	0.48	9	0.07	1.00
3	0.44	0.95	10	0.48	0.84
4	0.75	0.87	11	0.71	0.86
5	0.36	0.64	12	0.71	0.71
6	0.30	0.70	13	0.63	0.79
7	0.18	0.96	14	0.48	0.87
			Avg.	0.50	0.78

Table 1. Inter-rater agreement result.

The agreement in classification was higher for the opponent-based frame in most issues. This indicates that the annotators could apply the frame more clearly, resulting in smaller difference between them. The kappa measure was 0.78 on aver-

age. The kappa measure near 0.8 indicates a substantial level of agreement, and the value can be achieved, for example, when 8 or 9 out of 10 items are annotated equally (Table 1).

In addition, fewer articles were classified to the “Other” category under the opponent-based frame. The annotators classified about half of the articles to this category under the positive vs. negative frame whereas they classified about 35% to the category under the opponent-based frame. This is because the frame is more flexible to classify diverse articles of an issue, such as those covering arguments on different points, and those covering detrimental facts to a specific side without explicit positive or negative arguments.

The kappa measure was less than 0.5 for near half of the issues under the positive-negative frame. The agreement was low especially when the main topic of the contention was interpreted differently among the annotators; the main topic was interpreted differently for issue 3, 7, 8, and 9. Even when the topic was interpreted identically, the annotators were confused in judging complex arguments either as positive or negative. One annotator commented that “it was confusing as the arguments were not clearly for or against the topic often. Even when a disputant was assumed to have a positive attitude towards the topic, the disputant’s main argument was not about the topic but about attacking the opponent” The annotators all agreed that the opponent-based frame is more effective to understand the contention.

4 Disputant relation-based method

Disputant relation-based method adopts the opponent-based frame for classification. It attempts to identify the two opposing groups of the issue at hand, and analyzes whether an article more reflects the position of a specific side. The method is based on the observation that there exists two opposing groups of disputants, and the groups compete for news coverage. They strive to influence readers’ interpretation, evaluation of the issue and gain support from them (Miller et al. 2001). In this competing process, news articles may give more chance of speaking to a specific side, explain or elaborate them, or provide supportive facts of that side (Baker 1994).

The proposed method is performed in three stages: the first stage, disputant extraction, extracts

the disputants appearing in an article set; the second stage, disputant partition, partitions the extracted disputants into two opposing groups; lastly, the news classification stage classifies the articles into three categories, i.e., two for the articles biased to each group, and one for the others.

4.1 Disputant Extraction

In this stage, the disputants who participate in the contention have to be extracted. We utilize that many disputants appear as the subject of quotes in the news article set. The articles actively quote or cover their action in order to deliver the contention lively. We used straight forward methods for extraction of subjects. The methods were effective in practice as quotes of articles frequently had a regular pattern.

The subjects of direct and indirect quotes are extracted. The sentences including an utterance inside double quotes are considered as direct quotes.

The sentences which convey an utterance without double quotes, and those describing the action of a disputant are considered as indirect quotes (See the translated example 1 below). The indirect quotes are identified based on the morphology of the ending word. The ending word of the indirect quotes frequently has a verb as its root or includes a verbalization suffix. Other sentences, typically, those describing the reporter’s interpretation or comments are not considered as quotes. (See example sentence 2. The ending word of the original sentence is written in boldface).

- (1) The *government* **clarified** that there won’t be any talks unless *North Korea* apologizes for the attack.
- (2) The *government’s* **belief** is that a stern response is the only solution for the current crisis

A named entity combined with a topic particle or a subject particle is identified as the subject of these quotes. We detect the name of an organization, person, or country using the Korean Named Entity Recognizer (Lee et al. 2006). A simple anaphora resolution is conducted to identify subjects also from abbreviated references or pronouns in subsequent quotes.

4.2 Disputant Partitioning

We develop key opponent-based partitioning method for disputant partitioning. The method first identifies two key opponents, each representing

one side, and uses them as a pivot for partitioning other disputants. The other disputants are divided according to their relation with the key opponents, i.e., which key opponent they stand for or against.

The intuition behind the method is that there usually exists key opponents who represent the contention, and many participants argue about the key opponents whereas they seldom recognize and talk about minor disputants. For instance, in the contention on “investigation result of the Cheonan sinking incident”, the government of North Korea and that of South Korea are the key opponents; other disputants, such as politicians, experts, civic group of South Korea, the government of U.S., and that of China, mostly speak about the key opponents. Thus, it is effective to analyze where the disputants stand regarding their attitude toward the key opponents.

Selecting key opponents: In order to identify the key opponents of the issue, we search for the disputants who frequently criticize, and are also criticized by other disputants. As the key opponents get more news coverage, they have more chance to articulate their argument, and also have more chance to face counter-arguments by other disputants.

This is done in two steps. First, for each disputant, we analyze whom he or she criticizes and by whom he or she is criticized. The method goes through each sentence of the article set and searches for both disputant’s criticisms and the criticisms about the disputant. Based on the criticisms, it analyzes relationships among disputants.

A sentence is considered to express the disputant’s criticism to another disputant if the following holds: 1) the sentence is a quote, 2) the disputant is the subject of the quote, 3) another disputant appears in the quote, and 4) a negative lexicon appears in the sentence.

On the other hand, if the disputant is not the subject but appears in the quote, the sentence is considered to express a criticism about the disputant made by another disputant (See example 3. The disputants are written in italic, and negative words are in boldface.).

(3) the *government* defined that “the **attack** of *North Korea* is an act of **invasion** and also a **violation** of North-South Basic Agreement”

The negative lexicon we use is carefully built from the Wilson lexicon (Wilson et al. 2005). We translated all the terms in it using the Google trans-

lation, and manually inspected the translated result to filter out inappropriate translations and the terms that are not negative in the Korean context.

Second, we apply an adapted version of HITS graph algorithm to find major disputants. For this, the criticizing relationships obtained in the first step are represented in a graph. Each disputant is modeled as a node, and a link is made from a criticizing disputant to a criticized disputant.

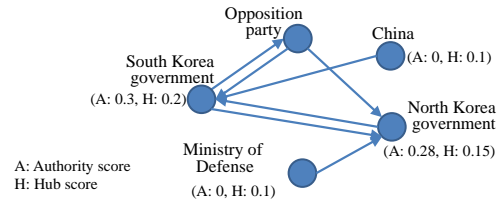


Figure 1. Example HITS graph illustration

Originally, the HITS algorithm (Kleinberg, 1999) is designed to rate Web pages regarding the link structure. The feature of the algorithm is that it separately models the value of outlinks and inlinks. Each node, i.e., a web page, has two scores: the authority score, which reflects the value of inlinks toward itself, and the hub score, which reflects the value of its outlinks to others. The hub score of a node increases if it links to nodes with high authority score, and the authority score increases if it is pointed by many nodes with high hub score.

We adopt the HITS algorithm due to above feature. It enables us to separately measure the significance of a disputant’s criticism (using the hub score) and the criticism about the disputant (using the authority score). We aim to find the nodes which have both high hub score and high authority score; the key opponents will have many links to others and also be pointed by many nodes.

The modified HITS algorithm is shown in Figure 2. We make some adaptation to make the algorithm reflect the disputants’ characteristics. The initial hub score of a node is set to the number of quotes in which the corresponding disputant is the subject. The initial authority score is set to the number of quotes in which the disputant appears but not as the subject. In addition, the weight of each link (from a criticizing disputant to a criticized disputant) is set to the number of sentences that express such criticism.

We select the nodes which show relatively high hub score and high authority score compared to other nodes. We rank the nodes according to the sum of hub and authority scores, and select from

the top ranking node. The node is not selected if its hub or authority score is zero. The selection is finished if more than two nodes are selected and the sum of hub and authority scores is less than half of the sum of the previously selected node.

Modified HITS(G, W, k)
$G = \langle V, E \rangle$ where V is a set of vertex, a vertex v_i represents a disputant E is a set of edges, an edge e_{ij} represents a criticizing quote from disputant i to j $W = \{w_{ij}\}$ weight of edge e_{ij}
For all $v_i \in V$ $Auth^1(v_i) = \#$ of quotes of which the subject is disputant i $Hub^1(v_i) = \#$ of quotes of which disputant i appears, but not as the subject
For $t = 1$ to k : $Auth^{t+1}(v_i) = \sum_{j: e_{ji} \in E} Hub^t(v_j) \times w_{ji}$ $Hub^{t+1}(v_i) = \sum_{j: e_{ij} \in E} Auth^t(v_j) \times w_{ij}$ Normalize $Auth^{t+1}(v_i)$ and $Hub^{t+1}(v_i)$

Figure 2. Algorithm of the Modified HITS

More than two disputants can be selected if more than one disputant is active from a specific side. In such cases, we choose the two disputants whose criticizing relationship is the strongest among the selected ones, i.e., the two who show the highest ratio of criticism between them.

Partitioning minor disputants: Given the two key opponents, we partition the rest of disputants based on their relations with the key opponents. For this, we identify whether each disputant has positive or negative relations with the key opponents. The disputant is classified to the side of the key opponent who shows more positive relations. If the disputant shows more negative relations, the disputant is classified to the opposite side.

We analyze the relationship not only from the article set but also from the web news search results. The minor disputants may not be covered importantly in the article set; hence, it can be difficult to obtain sufficient data for analysis. The web news search results provide supplementary data for the analysis of relationships.

We develop four features to capture the positive and negative relationships between the disputants.

1) Positive Quote Rate (PQR_{ab}): Given two disputants (a key opponent a , and a minor disputant b), the feature measures the ratio of positive quotes between them. A sentence is considered as a positive quote if the following conditions hold: the sentence is a direct or indirect quote, the two disputants appear in the sentence, one is the subject of the quote, and a positive lexicon appears in the

sentence. The number of such sentences is divided by the number of all quotes in which the two disputants appear and one appears as the subject.

2) Negative Quote Rate (NQR_{ab}): This feature is an opposite version of PQR. It measures the ratio of negative quotes between the two disputants. The same conditions are considered to detect negative quotes except that negative lexicon is used instead of positive lexicon.

3) Frequency of Standing Together (FST_{ab}): This feature attempts to capture whether the two disputants share a position, e.g., “*South Korea* and *U.S.* both criticized North Korea for...” It counts how many times they are co-located or connected with the conjunction “and” in the sentences.

4) Frequency of Division (FD_{ab}): This feature is an opposite version of the FST. It counts how many times they are not co-located in the sentences.

The same features are also calculated from the web news search results; we collect news articles of which the title includes the two disputants, i.e., a key opponent a and a minor disputant b .

The calculation method of PQR and NQR is slightly adapted since the titles are mostly not complete sentences. For PQR (NQR), it counts the titles which the two disputants appear with a positive (negative) lexicon. The counted number is divided by the number of total search results. The calculation method of FST and FD is the same except that they are calculated from the titles.

We combine the features obtained from web news search with the corresponding ones obtained from the article set by calculating a weighted sum. We currently give equal weights.

The disputants are partitioned by the following rule: given a minor disputant a , and the two key opponents b and c ,

classify a to b 's side if,

$$(PQR_{ab} - NQR_{ab}) > (PQR_{ac} - NQR_{ac}) \text{ or } ((FST_{ab} > FD_{ab}) \text{ and } (FST_{ac} = 0));$$

classify a to c 's side if,

$$(PQR_{ac} - NQR_{ac}) > (PQR_{ab} - NQR_{ab}) \text{ or } ((FST_{ac} > FD_{ac}) \text{ and } (FST_{ab} = 0));$$

classify a to *other*, otherwise.

4.3 Article Classification

Each news article of the set is classified by analyzing which side is importantly covered. The method classifies the articles into three categories, either to one of the two sides or the category “other”.

We observed that the major components which shape an article on a contention are quotes from disputants and journalists’ commentary. Thus, our method considers two points for classification: first, from which side the article’s quotes came; second, for the rest of the article’s text, the similarity of the text to the arguments of each side.

As for the quotes of an article, the method calculates the proportion of the quotes from each side based on the disputant partitioning result. As for the rest of the sentences, a similarity analysis is conducted with an SVM classifier. The classifier takes a sentence as input, determines its class to one of the three categories, i.e., one of the two sides, or other. It is trained with the quotes from each side (tf.idf of unigram and bigram is used as features). The same number of quotes from each side is used for training. The training data is pseudo-relevant: it is automatically obtained based on the partitioning result of the previous stage.

An article is classified to a specific side if more of its quotes are from that side and more sentences are similar to that side: given an article a , and the two sides b and c ,

$$\text{classify } a \text{ to } b \text{ if } \frac{Q_b + S_b}{S_U} \geq \frac{\alpha \cdot Q_{bc} + \beta \cdot S_{bc}}{S_U}$$

$$\text{classify } a \text{ to } c \text{ if } \frac{Q_c + S_c}{S_U} \geq \frac{\alpha \cdot Q_{bc} + \beta \cdot S_{bc}}{S_U}$$

classify a to *other*, otherwise.

where S_U : number of all sentences of the article

Q_i : number of quotes from the side i .

Q_{ij} : number of quotes from either side i or j .

S_i : number of sentences classified to i by SVM.

S_{ij} : number of sentences classified to either i or j .

We currently set the parameters heuristically. We set 0.7 and 0.6 for the two parameters α and β respectively. Thus, for an article written purely with quotes, the article is classified to a specific side if more than 70% of the quotes are from that side. On the other hand, for an article which does not include quotes from any side, more than 60% of the sentences have to be determined similar to a specific side’s quotes. We set a lower value for β to classify articles with less number of biased sentences (Articles often include non-quote sentences unrelated to any side to give basic information).

5 Evaluation and Discussion

Our evaluation of the method is twofold: first, we evaluate the disputant partitioning results, second, the accuracy of classification. The method was

evaluated using the same data set used for the classification frame comparison experiment.

A gold result was created through the three human annotators. To evaluate the disputant partitioning results, we had the annotators to extract the disputants of each issue, divide them into opposing two groups. We then created a gold partitioning result, by taking a union of the three annotators’ results. A gold classification is also created from the classification of the annotators. We resolved the disagreements between the annotators’ results by following the decision of the majority.

5.1 Evaluation of Disputant Partitioning

We evaluated the partitioning result of the two opposing groups, denoted as G1 and G2. The performance is measured using precision and recall. Table 2 presents the results. The precision of the partitioning was about 70% on average. The false positives were mostly the disputants who appear only a few times both in the article set and the news search results. As they appeared rarely, there was not enough data to infer their position. The effect of these false positives in article classification was limited.

The recall was slightly lower than precision. This was mainly because some disputants were omitted in the disputant extraction stage. The NER we used occasionally missed the names of unpopular organizations, e.g., civic groups, and the extraction rule failed to capture the subject in some complex sentences. However, most disputants who frequently appear in the article set were extracted and partitioned appropriately.

Issue #	G1		G2		Issue #	G1		G2	
	Pr.	Re.	Pr.	Re.		Pr.	Re.	Pr.	Re.
1	0.45	0.63	0.63	0.47	8	0.80	1.00	1.00	0.40
2	0.75	0.36	0.82	0.70	9	1.00	1.00	1.00	0.67
3	0.80	0.47	0.50	0.50	10	0.67	0.63	0.50	0.75
4	0.75	0.53	0.80	0.60	11	0.67	0.33	0.63	0.36
5	0.75	0.55	0.44	0.67	12	0.83	0.38	0.63	0.63
6	0.50	0.33	0.33	0.50	13	0.71	0.56	0.50	0.33
7	0.67	0.80	1.00	0.57	14	0.55	0.75	0.56	0.45
					Avg.	0.71	0.59	0.67	0.54

Table 2. Disputant Partitioning Result

5.2 Evaluation of Article Classification

We evaluate our method and compare it with two unsupervised methods below.

Similarity-based clustering (Sim.): The method implements a typical method. It clusters articles of an issue into three groups based on text similarity

Issue #	Total	G1	G2	Other	Issue #	Method	wF	Group 1			Group 2			Other			Issue #	Method	wF	Group 1			Group 2			Other		
								F	P	R	F	P	R	F	P	R				F	P	R	F	P	R	F	P	R
1	24	9	9	6	1	DrC	0.47	0.64	0.47	1.00	0.62	1.00	0.44	N/A	0.00	0.00	8	DrC	0.90	0.86	0.75	1.00	1.00	1.00	0.86	1.00	0.75	
2	23	11	4	8		QbC	0.50	0.62	0.47	0.89	0.71	1.00	0.55	N/A	0.00	0.00		QbC	0.48	0.57	0.50	0.67	0.57	0.50	0.67	0.33	0.50	0.25
3	18	2	10	6		Sim.	0.27	0.20	1.00	0.11	0.20	1.00	0.11	0.47	0.30	1.00		Sim.	0.56	0.67	0.67	0.67	0.50	0.40	0.67	0.50	1.00	0.33
4	25	9	12	4	2	DrC	0.65	0.67	0.62	0.73	0.86	1.00	0.75	0.53	0.57	0.50	9	DrC	0.77	N/A	0.00	N/A	0.57	0.50	0.67	0.82	1.00	0.70
5	18	5	9	4		QbC	0.65	0.76	0.80	0.73	0.60	0.50	0.75	0.53	0.57	0.50		QbC	0.79	N/A	0.00	N/A	0.67	0.67	0.67	0.82	1.00	0.70
6	10	0	5	5		Sim.	0.37	0.63	0.48	0.91	N/A	0.00	0.00	0.22	1.00	0.13		Sim.	0.49	N/A	0.00	N/A	0.00	0.00	0.00	0.63	0.67	0.60
7	22	4	11	7	3	DrC	0.72	0.57	0.40	1.00	0.67	1.00	0.50	0.86	0.75	1.00	10	DrC	0.66	0.71	0.56	1.00	0.73	1.00	0.57	0.40	0.50	0.33
8	10	3	3	4		QbC	0.74	0.57	0.40	1.00	0.75	1.00	0.60	0.77	0.71	0.83		QbC	0.72	0.77	0.63	1.00	0.77	0.83	0.71	0.50	1.00	0.33
9	13	0	3	10		Sim.	0.59	N/A	0.00	0.00	0.70	0.62	0.80	0.60	0.75	0.50		Sim.	0.40	0.33	1.00	0.20	0.44	1.00	0.29	0.40	0.25	1.00
10	15	5	7	3	4	DrC	0.80	0.82	0.69	1.00	0.86	1.00	0.75	0.57	0.67	0.50	11	DrC	0.61	0.73	0.80	0.67	0.50	0.43	0.60	0.57	0.67	0.50
11	15	6	5	4		QbC	0.81	0.90	0.82	1.00	0.86	1.00	0.75	0.44	0.40	0.50		QbC	0.39	0.62	0.57	0.67	0.20	0.20	0.20	0.29	0.33	0.25
12	13	2	3	8		Sim.	0.67	0.80	1.00	0.67	0.80	0.67	1.00	N/A	0.00	0.00		Sim.	0.47	0.63	0.46	1.00	0.33	1.00	0.20	0.40	1.00	0.25
13	19	12	5	2	5	DrC	0.60	0.63	0.50	0.83	0.71	0.83	0.63	0.33	0.50	0.25	12	DrC	0.67	0.29	0.20	0.50	0.67	0.67	0.67	0.77	1.00	0.63
14	25	13	10	2		QbC	0.55	0.40	0.50	0.33	0.71	0.67	0.75	0.44	0.40	0.50		QbC	0.38	0.33	0.25	0.50	0.44	0.33	0.67	0.36	0.47	0.25
						Sim.	0.51	0.63	0.46	1.00	0.67	1.00	0.50	N/A	0.00	0.00		Sim.	0.43	N/A	0.00	0.00	0.55	0.38	1.00	0.50	0.75	0.38
					6	DrC	0.89	N/A	0.00	N/A	0.89	1.00	0.80	0.89	1.00	0.80	13	DrC	0.65	0.79	0.69	0.92	0.33	1.00	0.20	0.67	1.00	0.50
						QbC	0.50	N/A	0.00	N/A	0.50	0.67	0.40	0.50	0.67	0.40		QbC	0.59	0.75	0.75	0.75	0.33	1.00	0.20	0.29	0.20	0.50
						Sim.	0.55	N/A	0.00	N/A	0.77	0.63	1.00	0.33	1.00	0.20		Sim.	0.54	0.71	0.63	0.83	0.33	1.00	0.20	N/A	0.00	0.00
					7	DrC	0.48	0.67	1.00	0.50	0.71	0.55	1.00	N/A	N/A	0.00	14	DrC	0.61	0.77	0.77	0.77	0.50	0.57	0.44	0.25	0.20	0.33
						QbC	0.48	0.67	1.00	0.50	0.62	0.53	0.73	0.17	0.20	0.14		QbC	0.66	0.83	0.75	0.92	0.53	0.67	0.44	0.33	0.33	0.33
						Sim.	0.44	0.40	0.27	0.75	0.57	0.60	0.55	0.25	1.00	0.14		Sim.	0.37	0.29	1.00	0.17	0.60	0.43	1.00	N/A	0.00	0.00

*N/A: The metric could not be calculated in some cases. This happened when no articles were classified to a category.

Table 3. Number of articles of each issue and group (left), and classification performance (right)

ty. It uses tf.idf of unigram and bigram as features, and cosine similarity as the similarity measure. We used the K-means clustering algorithm.

Quote-based classification (QbC): The method is a partial implementation of our method. The disputant extraction and disputant partitioning is performed identically; however, it classifies news articles merely based on quotes. An article is classified to one of the two opposing sides if more than 70% of the quotes are from that side, or to the “other” category otherwise.

Results: We evaluated the classification result of the three categories, the two groups G1 and G2, and the category Other. The performance is measured using precision, recall, and f-measure. We additionally used the weighted f-measure (wF) to aggregate the f-measure of the three categories. It is the weighted average of the three f-measures. The weight is proportional to the number of articles in each category of the gold result.

The disputant relation-based method (DrC) performed better than the two comparison methods. The overall average of the weighted f-measure among issues was 0.68, 0.59, and 0.48 for the DrC, QbC, and Sim. method, respectively (See Table 3). The performance of the similarity-based clustering was lower than that of the other two in most issues.

A number of works have reported that text similarity is reliable in stance classification in political domains. These experiments were conducted

in political debate corpus (Lin et al. 2006). However, news article set includes a number of articles covering different topics irrelevant to the arguments of the disputants. For example, there can be an article describing general background of the contention. Similarity-based clustering approach reacted sensitively to such articles and failed to capture the difference of the covered side.

Quote-based classification performs better than similarity-based approach as it classifies articles primarily based on the quoted disputants. The performance is comparable to DrC in many issues. The method performs similarly to DrC if most articles of an issue include many quotes. DrC performs better for other issues which include a number of articles with only a few quotes.

Error analysis: As for our method, we observed three main reasons of misclassification.

1) Articles with few quotes: Although the proposed method better classifies such articles than the quote-based classification, there were some misclassifications. There are sentences that are not directly related to the argument of any side, e.g., plain description of an event, summarizing the development of the issue, etc. The method made errors while trying to decide to which side these sentences are close to. Detecting such sentences and avoiding decisions for them would be one way of improvement. Research on classification

of subjective and objective sentences would be helpful (Wiebe et al. 99).

2) Article criticizing the quoted disputants: There were some articles criticizing the quoted disputants. For example, an article quoted the president frequently but occasionally criticized him between the quotes. The method misclassified such articles as it interpreted that the article is mainly delivering the president's argument.

3) Errors in disputant partitioning: Some misclassifications were made due to the errors in the disputant partitioning stage, specifically, those who were classified to a wrong side. Articles which refer to such disputants many times were misclassified.

6 Conclusion

We study the problem of classifying news articles on contentious issues. It involves new challenges as the discourse of contentious issues is complex, and news articles show different characteristics from commonly studied corpus, such as product reviews. We propose opponent-based frame, and demonstrate that it is a clear and effective classification frame to contrast arguments of contentious issues. We develop disputant relation-based classification and show that the method outperforms a text similarity-based approach.

Our method assumes polarization for contentious issues. This assumption was valid for most of the tested issues. For a few issues, there were some participants who do not belong to either side; however, they usually did not take a particular position nor make strong arguments. Thus, the effect on classification performance was limited. Discovering and developing methods for issues which involve more than two disputants groups is a future work.

References

- Rakesh Agrawal, Sridhar Rajagopalan, Ramakrishnan Srikant, and Yirong Xu. 2003. Mining newsgroups using networks arising from social behavior. In *Proceedings of WWW*.
- Baker, B. 1994. How to Identify, Expose and Correct Liberal Media Bias. Media Research Center.
- Mohit Bansal, Claire Cardie, and Lillian Lee. 2008. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-2008)*.
- Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. In *Journal of ACM*, 46(5): 604-632.
- Landis JR, Koch G. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.
- Changki Lee, Yi-Gyu Hwang, Hyo-Jung Oh, Soojong Lim, Jeong Heo, Chung-Hee Lee, Hyeon-Jin Kim, Ji-Hyun Wang, Myung-Gil Jang. 2006. Fine-Grained Named Entity Recognition using Conditional Random Fields for Question Answering. In *Proceedings of Human & Cognitive Language Technology (HCLT)*, pp. 268~272. (in Korean)
- Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, and Alexander Hauptmann. 2006. Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL-2006)*, pages 109–116, New York.
- Mark M. Miller and Bonnie P. Riechert. 2001. Spiral Opportunity and Frame Resonance: Mapping Issue Cycle in News and Public Discourse. In *Framing Public Life: Perspectives on Media and our Understanding of the Social World*, NJ: Lawrence Erlbaum Associates.
- I Ounis, M de Rijke, C Macdonald, G Mishne, and I Soboroff. 2006. Overview of the TREC-2006 Blog Track. In *Proceedings of TREC*.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Paul, M. J., Zhai, C., Girju, R. 2010. Summarizing Contrastive Viewpoints in Opinionated Text. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

- Schon, D.A., and Rien, M. 1994. Frame reflection: Toward the resolution of intractable policy controversies. New York: Basic Books.
- Y. Seki, D. Evans, L. Ku, L. Sun, H. Chen, and N. Kando. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In *Proceedings of 7th NTCIR Evaluation Workshop*, pages 185-203
- Kwangseob Shim and Jaehyung Yang. 2002. MACH : A Supersonic Korean Morphological Analyzer, *Proceedings of the 19th International Conference on Computational Linguistics (COLING-2002)*, pp.939-945.
- 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.
- Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological online debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text (CAAGET '10)*.
- Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get outthe vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 327-335, Sydney, Australia, July. Association for Computational Linguistics.
- Turney, Peter D. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, *Proceedings of ACL-02*, Philadelphia, Pennsylvania, 417-424
- Wiebe, Janyce M., Bruce, Rebecca F., & O'Hara, Thomas P. 1999. Development and use of a gold standard data set for subjectivity classifications. In *Proc. 37th Annual Meeting of the Assoc. for Computational Linguistics (ACL-99)*. June, pp. 246-253.
- T. Wilson, J. Wiebe, and P. Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.