

# Classification of Controversial News Issues Based on Disputant Relation: A Survey

Priya P.Raut

PG Student, Department of Computer Engineering,  
SES's R. C. Patel Institute of  
Technology, Shirpur, North Maharashtra University, MS, India  
priyaraute52@gmail.com

Nitin N.Patil<sup>2</sup>

Head & Associate Professor, Department of Computer  
Engineering, SES's R. C. Patel Institute of  
Technology, Shirpur, North Maharashtra University, MS, India  
er\_nitinpatil@rediffmail.com

**Abstract**— The controversial news issues draw much interest from the public. But it is not simple for an ordinary user to search and contrast the opposing arguments and have complete understanding of issues. Disputant relation based method classifies the opposing views of the news or issues which can help readers to easily understand the issue. For classifying news articles on contentious issues disputant relation-based method is used. It is known that the disputants of a contention are an important feature for understanding the conversation so, the disputant relation based method performs unsupervised classification on news articles based on disputant relations, and helps readers naturally view the articles through the opponent- based frame and attain balanced understanding, free from a specific biased viewpoint.

This method is performed in three stages: disputant mining, disputant separation and article classification. Also a modified version of HITS algorithm is used in disputant partitioning and an SVM classifier is used for article analysis. In this paper we review the said classification based on disputant relation.

**Keywords:** Classification, document analysis, text mining.

\*\*\*\*\*

## I. INTRODUCTION

In today's world, news issues become an important part of everyone's life. It is essential to know the information of the surrounding as to remain updated. This covering of news issues work is done by journalism. The covering of controversial issues of the surrounding is an important function of journalism as controversial issues grow in various areas such as business, politics, sports, etc and these issues include various participants and their various views. It is noticed that sometimes news articles get biased and unable to present conflicting views of the issue. That's why it is very difficult for ordinary readers to analyze the conflicting views and understand the controversy. Readers generally make their views through single articles advanced news presentation models are needed to increase the awareness on differing views. In this paper, we survey on a disputant relation based method for classifying news articles on controversial issues. It is observed that the disputants i.e. the participants who take his appropriate position and participate in the controversy such as businessmen, politician, sportsmen, experts, news writer, and so on are an important aspects for understanding the conversation. News producers primarily shape articles on a contention by selecting and covering specific disputants [2]. Readers also naturally understand the controversy by identifying who the opposing Disputants are. The method helps readers naturally view the news articles through the 'opponent-based frame' [1]. It performs classification in an unsupervised mode: it dynamically identifies the opposing disputant groups and classifies the articles according to their positions. As such, it successfully helps readers compare articles of a controversy and achieve balanced understanding, free from a specific biased viewpoint. The surveyed method differs from those used in related tasks as it aims to achieve classification under the opponent based frame. Most research on sentiment classification and debate position appreciation

takes a topic-oriented view, and attempts to perform classification under the 'positive versus negative' frame for the given topic, for example, positive versus negative about television. However, news articles of a controversy are hardly classified under such frames.

### A. Argument Frame Comparison

Launching a suitable argument frame is important. It provides a framework that allows readers to naturally understand the controversy. It also determines how classification methods should classify articles of the issue.

## II. RELATED WORK IN DOCUMENT CLASSIFICATION

Turney et al., and Pang et al., have been made researched on sentiment classification in document-level. It aims to automatically recognize and classify the sentiment of documents into positive or negative. Opinion summarization aims a similar goal, to identify special opinions on a topic and generate summaries of them [2], [3].

Paul et al., developed an unsupervised method for generating summaries of contrastive opinions on a common topic. Many of these works make a number of assumptions that are difficult to apply to the conversation of controversial news issues. They usually apply a single static classification frame, 'positive versus negative,' to the topic [4].

Somasundaran et al., had proposed that a number of works deal with debate attitude recognition, which is a closely related task. They try to identify a position of a dispute, such as ideological. This debate frame is often not appropriate for controversial issues for similar reasons as the positive/negative frame. In contrast, surveyed method does not assume a fixed debate frame, and rather develops one based on the opponents of the controversy at hand [5].

Thomas et al., and Agrawal et al., had proposed that the several works have used the relation between speakers or authors for classifying their debate stance. However, these works also assume the same debate frame and use the debate mass, e.g., floor debates in the House of Representatives, online dispute forums. Their approaches are also supervised, and require training data for relation analysis, e.g., voting records of congress people [6], [7].

Schon et al., had proposed the conversation of controversial issues in news articles shows different characteristics from that studied in the sentiment classification tasks. First, the opponents of a controversial issue often discuss different topics. Second, the frame of dispute is not fixed as, positive vs. negative [9].

### III. DISPUTANT RELATION-BASED CLASSIFICATION METHOD

The disputant relation-based method implements the opponent based frame for classification. It attempts to recognize the two opposing groups of the matter at hand, and determines whether an article replicates the position of a definite side more. The method is based on the observation that there are usually two opposing groups of disputants, and the groups compete for news coverage. They strive to influence readers' understanding, estimate of the issue, and grow support from them [2]. In this challenging process, news articles may give more chance of speaking to a detailed side, explain or detailed them, or supply facts helpful of that side. The surveyed method is performed in three stages:

The first step, disputant extraction, mines the disputants appearing in an article set.

The second step, disputant partition, divisions the mined disputants into two opposing groups.

Lastly the news classification step classifies the articles into three categories, i.e., two for the articles influenced to each group, and one for the others.

This method assumes polarization for controversial issues. This statement 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. For example, in the second issue Entrance of retailers to supermarket business the government was in the middle between the big retail companies and the small store owners. The government commented that this is a difficult problem to solve and did not support a specific side. Based on this observation, the method is designed to identify opposing two groups of disputants and recognize articles biased to a specific side.

#### A. Disputant Extraction

In this step, the participants who participate in the controversy have to be extracted or mined. It is expanded that many disputants appear as the subject of quotes in the news article set. The articles actively quote or cover their action to deliver the controversy actively. It uses straight-forward methods for extraction of the subjects. The methods were efficient in practice as quotes of articles often had a regular pattern. The subjects of direct and indirect quotes are mined. The sentences including a statement inside double quotes are considered as direct quotes. The sentences that express a statement without double quotes, and those describing the action of a disputant

are considered as indirect quotes (see the example 1 below). The indirect quotes are identified based on the morphology of the ending word. The ending word of the indirect quotes often has a verb as its root or includes a verbalization suffix. Other sentences, typically, those describing the reporter's explanation or comments are not considered as quotes (see example sentence 2).

1. The government **clarified** that there would not be any talks unless Pakistan apologizes for the attack.
2. The government's **belief** is that a demanding response is the only solution for the current emergency.

#### B. Disputant Partitioning

"Key opponent-based partitioning" method is developed for disputant partitioning step. The method initially identifies two key opponents, each representing one side, and uses them as a pivot for partitioning other disputants [1]. The other disputants are divided according to their relation with the key opponents, i.e., which key opponent they stand for or against.

Disputant partitioning it is a meaningful task to explore for more optimized methods. The discussed method is our first round solution to the task, based on the observation of the criticizing structure that is frequent in news article sets of controversial issues, i.e., existence of the key opponents who actively criticize and are criticized by others, and existence of other minor disputants who commonly criticize the key opponents but are not criticized often. The discussed key opponent-based partitioning is an initial algorithm that operates this criticizing structure. The perception behind the method is that there frequently exists key opponents who represent the controversy, and many participants disagree about the key opponents, whereas they rarely identify and talk about minor disputants.

Selecting key opponents: To identify the key opponents of the issue, it is to be searched 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 clear their argument, and also have more chance to face counter-arguments of other disputants. This is done in two steps. First, for each disputant, it is to be analyzing whom he or she criticizes and by whom he or she is criticized. The method goes through each sentence of the article set and investigates for both disputant's criticisms and the criticisms about the disputant. Based on the criticisms, it analyzes the associations among disputants. On the other hand, if the disputant is not the subject but demonstrates in the quote, the sentence is considered to convey a criticism about the disputant from another disputant. Second, a modified version of the HITS graph algorithm to uncover 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 connection is made from a criticizing disputant to a Criticized disputant.

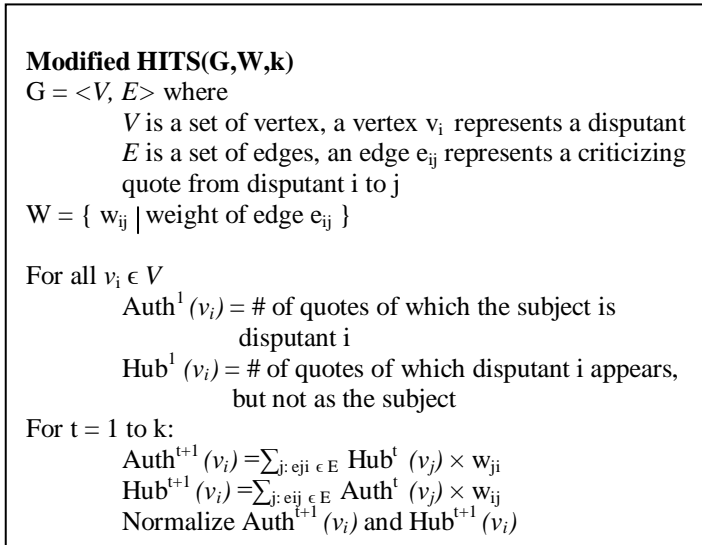


Figure 3.2: Modified HITS algorithm.

Originally, the HITS algorithm is designed to rate WebPages 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 webpage, 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. The HITS algorithm is adopted 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). The aim is to find the nodes that 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 adapted HITS algorithm is shown in Fig. 3.2. Some adaptation is to make the algorithm reproduce the disputants' uniqueness. The initial hub score of a node is set to the number of quotes in which the matching 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 convey such criticism.

Partitioning minor disputants: Given the two key opponents, we have to partition the rest of disputants based on their relations with the key opponents. For this, it is to be identifying whether each disputant has a positive or negative relation with the key opponents. The disputant is classified to the side of the key opponent with whom the disputant shows a more positive relation. If the disputant shows a negative relation, the disputant is classified to the opposite side. Here are the four features to capture the positive and negative relationships between the disputants:

- 1) Positive Quote Rate (PQRab): Given two disputants (a key opponent a, and a minor disputant b), the

feature measures the ratio of positive quotes between them.

- 2) Negative Quote Rate (NQRab): This feature is an opposite version of PQR. It measures the ratio of negative quotes between the two disputants.
- 3) Frequency of Standing Together (FSTab): This feature attempts to capture whether the two disputants share a position.
- 4) Frequency of Division (FDab): This feature is an opposite version of the FST. It counts how many times they are not collocated in the sentences.

### C. Article Classification

Each news article of the set is classified by evaluating which side is prominently enclosed. The method classifies the articles into three categories, either to one of the two sides or the category "other".

It is observed that the major components that shape an article on a controversy are quotes from disputants and journalists' commentary. Thus, this method believes two points for classification: First, from which side the article's quotes came; second, for the rest of the article's text, the correspondence of the text to the arguments of each side.

As for the quotes of an article, the method computes the amount of the quotes from each side based on the disputant partitioning step's result. As for the rest of the sentences, a similarity analysis is conducted with an SVM classifier [8]. The SVM classifier receives a sentence as input, establishes its class to one of the three types, i.e., one of the two categories, or other. It is qualified with the quotes from each side. The related number of quotes from each side is used for training. It is automatically obtained based on the partitioning result of the earlier stage.

According to survey, an article is classified to a precise 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,

Classify a to b if,

$$\frac{Q_b}{S} > \frac{Q_c}{S}$$

Classify a to c if,

Classify a to other, otherwise, where

Su: Number of all sentences of the article

Qi: Number of quotes from the side i.

Qij: Number of quotes from either side i or j.

Si: Number of sentences classified to i by classifier.

Sij: Number of sentences classified to either i or j.

## IV. CONCLUSIONS AND FUTURE WORK

In this paper the problem of classifying news articles on controversial issues is studied. It involves new challenges as the conversation of controversial issues is complex, and news articles show different characteristics from commonly studied amount, such as product reviews. The conversation involves many topics and the arguments often do not fit the 'positive versus negative' frame.

It suggests that opponent-based frame is a clear and effective frame for understanding controversial issues. The frame does not require articles to cover a common topic nor the arguments to explicitly express positive or negative sentiments. In this study, it can be find that the participants easily identified the

opposing disputants, and classified articles written for or against the disputants. Instead of taking a topic-oriented view, the disputant relation-based method focuses on the disputants of the controversy.

For better performance of Disputant relation based method, 'Naive Bayes algorithm' can be used in article classification step as; SVM classifier cannot classify the issue that involves three or more disputants.

#### REFERENCES

- [1] Souneil Park, Jungil Kim, Kyung Soon Lee, and Junehwa Song, Disputant Relation Based Classification for Contrasting Opposing Views of Contentious News Issues, *IEEE Transactions On Knowledge And Data Engineering*, Vol. 25, No. 12, December 2013.
- [2] P. Turney, Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews, *Proc. 40th Ann. Meeting Assoc. Computational Linguistics (ACL '02)*, pp. 417-424, 2002.
- [3] B. Pang, L. Lee, and S. Vaithyanathan, Thumbs up? Sentiment Classification Using Machine Learning Techniques", *Proc. Conf. Empirical Methods Natural Language Processing (EMNLP '02)*, pp. 79-86, 2002.
- [4] M.J. Paul, C. Zhai, and R. Girju, Summarizing Contrastive Viewpoints in Opinionated Text, *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 66-76, 2010.
- [5] S. Somasundaran and J. Wiebe, Recognizing Stances in Ideological Online Debates, *Proc. NAACL HLT Workshop Computational Approaches Analysis and Generation Emotion in Text (CAAGET '10)*, pp. 116-124, 2010.
- [6] M. Thomas, B. Pang, and L. Lee, Get Out the Vote: Determining Support or Op- position from Congressional Floor-Debate Transcripts, *Proc. Conf. Empirical Methods Natural Language Processing (EMNLP '06)*, pp. 327-335, 2006.
- [7] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu, Mining Newsgroups Using Networks Arising from Social Behavior, *Proc. 12th Int'l Conf. World Wide Web (WWW '03)*, pp. 529-535, 2003.
- [8] T. Joachims, Making Large-Scale SVM Learning Practical, *Advances in Kernel Methods-Support Vector Learning*, B. Scholkopf, C. Burges, and A. Smola, eds., MIT Press, 1999.

#### Books:

- [9] D.A. Schon and M. Rien, Frame Reaction: Toward the Resolution of Intractable Policy Controversies, (IBasicBooks, 1994).