

# Support or Oppose? Classifying Positions in Online Debates from Reply Activities and Opinion Expressions

Akiko Murakami<sup>1,2</sup> Rudy Raymond<sup>1</sup>

<sup>1</sup> IBM Research - Tokyo

<sup>2</sup> Graduate School of Interdisciplinary Information Studies, The University of Tokyo

{akikom, raymond}@jp.ibm.com

## Abstract

We propose a method for the task of identifying the *general* positions of users in online debates, i.e., support or oppose the main topic of an online debate, by exploiting *local* information in their remarks within the debate. An online debate is a forum where each user post an opinion on a particular topic while other users state their positions by posting their remarks within the debate. The supporting or opposing remarks are made by directly replying to the opinion, or indirectly to other remarks (to express local agreement or disagreement), which makes the task of identifying users' general positions difficult. A prior study has shown that a link-based method, which completely ignores the content of the remarks, can achieve higher accuracy for the identification task than methods based solely on the contents of the remarks. In this paper, we show that utilizing the textual content of the remarks into the link-based method can yield higher accuracy in the identification task.

## 1 Introduction

Social computing tools, such as a SNS (Social Network Service) or an online discussion board have become very powerful communication tools for discussing topics with people around the world. Many companies use these kinds of social computing tools to understand their customers' requirements and their marketing activities. Social

computing tools are very useful not only for aggregating customers' opinions outside the companies, but also for aggregating their employees' ideas. For example, IBM has held Jam<sup>1</sup> sessions, which are short-term online discussions to aggregate ideas from employees and customers. The results of Jam sessions help management decisions, for instance the technology areas to invest.

Not just enterprises, but some nations are trying to aggregate their citizens' ideas in the Internet and provide systems for discussions at the people-to-people levels as part of the movement for open government. The United States government has the *Idea Factory*<sup>2</sup> website for collecting ideas to enhance activities of Department of Homeland Security (DHS) and the *Open For Questions*<sup>3</sup> to collect requests for the US government.

The motivation for creating these kinds of online discussions is not limited to collecting ideas but also to help understand the trends of opinions about the ideas or topics. This means that getting a quick overview of opinions about ideas is a key point for the success of online discussions.

In this paper we propose a method to show quick overview of participants' positions, "Support" or "Oppose" for the main idea or topic in online debates. It is difficult to identify each person's position for a topic directly, since most of opinionative expressions are made not for main topic but for adjacent remarks. This causes a difficulty in building answer sets for classifier. The following example shows opinion expressions for a main topic focused on an adjacent remark in a

<sup>1</sup><https://www.collaborationjam.com>

<sup>2</sup><http://www.whitehouse.gov/open/innovations/IdeaFactory>

<sup>3</sup><http://www.whitehouse.gov/OpenForQuestions/>

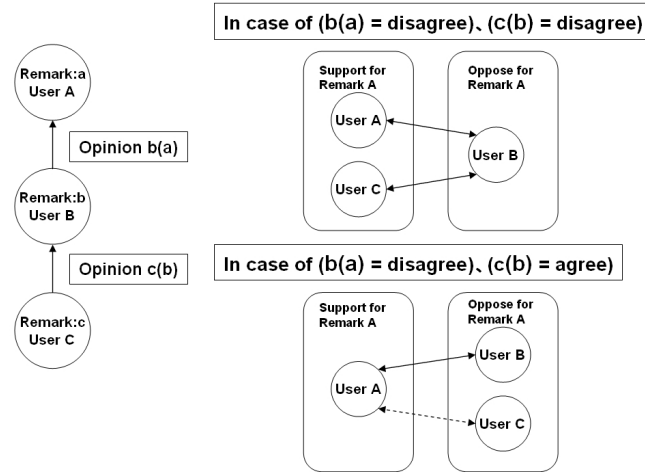


Figure 1: Identifying users' positions from their opinions about previous remarks

debate. In this example, The main topic is “Travel and F2F (face-to-face) meeting is fundamental to business”.

**Remark A** Travel isn't necessary because besides the high cost of travels around the world, today we have a lot of communication tools, for instance web conference, video chat that can easily contribute to join leaders around the world in a cheaper way.

**Remark B** I disagree. Without travel and F2F meetings global integration does not work as well or as quickly. It doesn't mean that everybody has to travel all the time, but at least some meetings are key to success.

The author of Remark A mentions that travel is not necessary to business. This opinion opposed to the main topic, so that the position for the main topic is “Oppose”. In contrast, the opinion expression in Remark B is not an opinion about the main topic, but relates to the previous opinion in Remark A. This opinion expression indicates that the author of Remark B disagrees with the opinion of Remark A, and indirectly implies agreement with the main topic. Thus, although it is hard to infer the global position of Remark B from only the sur-

face expressions, it is straightforward to infer that an opinion in Remark B about Remark A is negative (i.e., Remark B expresses disagreement with Remark A).

In this paper, positions with regards to the main topic (global positions) are classified into two classes: support and oppose, while opinions about the previous remarks (local positions) are classified into three: *agree*, *disagree*, and *neutral*. For example, let us consider the case in Fig. 1, where Remark “a” is the main topic, and Remark “b” is the reply to Remark “a” and Remark “c” is the reply to Remark “b”. Here, let  $b(a)$  be the local position, that is, opinion (agree/disagree/neutral) in Remark “b” on the topic in Remark “a”. For example, if  $b(a)$  and  $c(b)$  are *disagree*, one can determine that the authors of the corresponding remarks are in the opposition. That is, the author of Remark “c” agrees with Author A (the author of Remark “a”), that is, the main topic, while Author B is against the others. On the contrary, if  $b(a)$  is *disagree* and  $c(b)$  is *agree*, then Author C agrees with Author B and therefore it implies that Author C is against Author A. In this case, only Author A supports the main topic while Author B and Author C oppose to the main topic.

To infer supporting or opposing positions with regards to the main topic, two steps are used. First, the degree of disagreement between any two users is computed from the link structure and the text of

each pair of their adjacent replies. This is used as the link weight between nodes (which correspond to users in a debate) in the network. Second, the bipartition of the users in the weighted network is computed by finding a bipartition that induces the *maximum cut* of the network, a partition of nodes into two disjoint sets that maximizes the sum of the weights of the links connecting nodes in different sets. Since the weight of the links is higher (more positive) when the degree of disagreement is higher, the bipartition is expected to express two groups of opposing positions.

In order to evaluate the performance of our method, we conducted some experiments to identify the supporting and opposing positions of participants in online debates. The experimental results indicate that our method leads to higher precision than the baseline method, which is described in (Agrawal et al., 2003).

The rest of this paper is organized as follows. First we describe related work in Section 2, and in Section 3 we propose our method for identifying participants' positions from their reply activities and text contents. In Section 4 we explain the data sets used for the evaluations and show the experimental results of an opinion classifier for adjacent remarks and a support/oppose classifier for the participants in online debates. We conclude this paper and describe future work in Section 5.

## 2 Related Work

There are some research papers published on analysis of online discussions. Some researches reported on how to analyze and navigate IBM Jam sessions. Millen et al. pointed out the importance of supporting the participants in discussions and demonstrated the effectiveness of their methods in one of these jams (Millen and Fontaine, 2003). Dave et al. described ways for jam participants to navigate using visualization techniques (Dave et al., 2004). One of the authors previously also proposed a method to mine discussion records using XML annotations (Murakami et al., 2001) and a method to find important remarks in a discussion thread based on the reply-to structure and participants' opinions (Murakami et al., 2007).

Classifying agree/disagree opinions in conversational debates using Bayesian networks was

presented in (Galley et al., 2004). Agrawal et al. described an observation that reply activities show disagreement with previous authors, and showed a method to classify the supporting/opposing position of users based on this observation in (Agrawal et al., 2003). Thomas et al. (Thomas et al., 2006) introduced some constraints that a single speaker retains the same position for the classification of participants' positions from floor-debate transcripts.

## 3 Proposed Method

### 3.1 Calculating the Reaction Coefficient between participants

We call the degree of divergence in the opinions between participants a *reaction coefficient*. This reaction coefficient is defined as a function of the participants  $i, j$ , represented as  $r(i, j)$ . To calculate reaction coefficients, we extracted pairs of a remark and its reply remark, and assigned "local position flags" to the pairs. There are three local position flags, "agree", "disagree", and "neutral". The reaction coefficient  $r(i, j)$  between participants  $i$  and  $j$  is defined as:

$$r(i, j) = \alpha N_{\text{disagree}}(i, j) + \beta N_{\text{neutral}}(i, j) + \gamma N_{\text{agree}}(i, j), \quad (1)$$

where  $N_{\text{opinion}}(i, j)$  is the number of remark pairs with opinion as the corresponding local position flag between participants  $i$  and  $j$ .

Typically we assign a positive value to  $\alpha$ , a slightly positive value to  $\beta$ , and a negative value to  $\gamma$ . This means that  $r(i, j)$  is positive when there are only neutral remarks between user  $i$  and  $j$ . This is based on the hypothesis in (Agrawal et al., 2003) that replies usually indicate disagreement with previous remarks. There is no directionality in reaction coefficients so that  $r(i, j) = r(j, i)$ .

### 3.2 Classification of Participants' Positions based on the Max Cut Problem

Let the graph corresponding to the activity network of the participants in an online debate be  $G(V, E)$ , where  $V$  is the set of nodes that corresponds to participants and  $E$  is the set of edges each of which links participants that exchanged remarks. For any  $i, j \in V$ , let  $r(i, j)$  be the weight of the link between  $i$  and  $j$ . A partition of the

Table 1: Ideas and Number of Comments and Participants for the Ideas

Idea ID	Title	# of Comments	# of Participant	# of Remarks per Participant
1	Making “IT” Education as a Compulsory Subject in Schools	75	45	1.7
2	Making Personal-Computer Makers to Supplying Service Parts	130	21	6.2
3	Adoption of “Basic Income”	118	57	2.1
4	Votes in elections using Closed Networks	108	40	2.7
5	Computerized Books in Libraries	50	12	4.2

participants into supporting and opposing parties,  $S_{\text{sup}}$  and  $S_{\text{opp}}$  respectively, is computed by solving the *max cut* problem on  $G(V, E)$  defined as follows.

**[Max cut problem]** Given  $G(V, E)$  as above, find a bipartition of  $V$  into  $S_{\text{sup}}$  and  $S_{\text{opp}} = V \setminus S_{\text{sup}}$  so that  $\sum_{i \in S_{\text{sup}}, j \in S_{\text{opp}}} r(i, j)$  is maximized.

The max cut problem is known to be NP-hard, and thus in general is difficult to solve. However, good approximation algorithms based on Linear Programming and Semidefinite Programming have been developed recently, and combined with branch-and-bound techniques a good exact max-cut solver called *BiqMac* exists (Rendl et al., 2010). We used *BiqMac* for solving the max cut problem *exactly* on the activity network. Although a faster approximate max cut solver is used in (Agrawal et al., 2003), it is based on the limiting assumption that the size of  $S_{\text{opp}}$  is approximately the same as  $S_{\text{sup}}$ . This cannot be assumed for the networks in this paper.

## 4 Experiments

### 4.1 Corpus

The Ministry of Economy, Trade and Industry in Japan (METI) was accepting public opinions on e-government programs via the “e-METI Idea Box<sup>4</sup>” from February 23 to March 15 2010. Participants could show their positions for the ideas since the site accepted comments on the main idea and other comments, so this discussion can be regarded as a kind of debate. We used this data

<sup>4</sup>[http://www.meti.go.jp/policy/it\\_policy/open-meti/](http://www.meti.go.jp/policy/it_policy/open-meti/)

to evaluate our proposed method. The ideas and comments were written in Japanese and the data is available at the METI website.

For the 936 ideas that were posted to the Idea Box, we examined 17 ideas with more than 40 comments. Finally we selected five ideas for the evaluation. The numbers of remarks (a main idea and comments), participants, and remarks per participants are shown in Table 1.

We extracted the reply-to structure information in textual contents. The Idea Box system had a capability to adding a comment to a main topic or the other comment, and the system inserted an identifier in comment’s text. Each identifier started with “#” and the IDs of the previous comments followed the identifier, such as “#003” (with #001 referring to the main topic in the thread). An idea or comments may have several comments as replies, so this reply-to structure in a debate is a tree structure whose root node is the main topic. A typical reply-to tree structure is shown in Fig 2.

### 4.2 Agree/Disagree Classification

To calculate the reaction coefficients, we need to extract the reply-to pairs and classify these pairs into the agree/disagree/neutral classes. To classify these remark pairs we use opinionative and sentiment expressions. If a reply remark contains an expression of “I agree with you” then it should be classified into the agree class. Another example of expressions of the agree class would be “That’s a good idea!”.

To extract expressions of opinion, we created a simple pattern dictionaries that contains

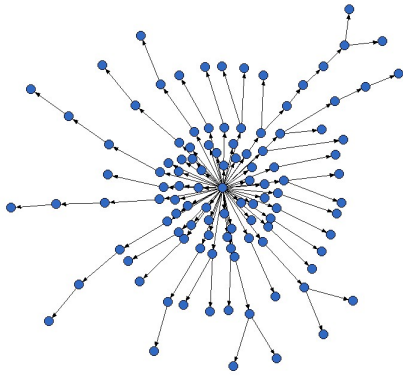


Figure 2: Reply-to Structure of a Debate

agree/disagree expressions. For instance, “I disagree with your idea” and “I don’t agree with you” are in the disagree pattern dictionary. At the same time we use a sentiment analysis tool to extract sentiment expressions. The tool we used for sentiment expression extractions is the same as described in (Kanayama et al., 2004), which use machine translation techniques to identify sentiment expressions in text. The tool returns sentiment expressions with a sentiment label, favorable or unfavorable.

After identifying opinionative and sentiment expressions in the remarks, scores for the opinion classification are calculated. The score of each reply-to pair is the number of agreeing and favorable expressions minus the number of disagreeing and unfavorable expressions in the reply remark. When the score is positive, the opinion of the pair is identified as agree, and if the score is negative then the opinion of the pair identified as disagree. If the score equals zero, then the opinion is identified as neutral.

To evaluate this opinion classifier, we did an experiment with the METI data, which was manually assigned agree/disagree/neutral flags. The answers for these evaluation were created by us for three of the idea threads (Idea IDs #1,#2 and #3). Since most remarks do not have agree or disagree expressions, most reply-to pairs are classified into the neutral class. This means that calculating precision and recall for the neutral class are not important. For the evaluation of the clas-

Table 2: Accuracy of opinion classification for reply-to pairs

Idea ID	Precision	Recall
1	0.63	0.25
2	0.62	0.14
3	0.44	0.38
Ave.	0.56	0.26

sifier we calculated precisions and recalls only for agree and disagree classes. The results are shown in Table 2.

### 4.3 Support/Oppose Classification

Using the numbers of agree/disagree/neutral reply-to pairs, we can calculate the reaction coefficients for each pair of participants. After calculating the reaction coefficients for all of the participants’ pairs, we can classify each participant into support or oppose sets using the max cut technique. In this subsection, we explain how to evaluate our proposed method and show experimental results.

#### 4.3.1 Answer Sets for Global Position Classification

To evaluate our method we created answer sets for a global position classifier, consisting of participant sets with the position labels *Support* or *Oppose*. We identified the positions of the participants’ remarks with contexts, but we assigned the “Unclear” label for some participants since their remarks did not contain enough information to classify their global positions. For showing the validity of the answer sets, two annotators annotated three ideas and calculated a  $\kappa$  value. The  $\kappa$  value is 0.69 so that this answer set is appropriate as an evaluation set. The use of the answer set annotated by a single annotator for the evaluation of support/oppose classification is justified since the agreement rate (the  $\kappa$  value) is enough for the evaluation.

#### 4.3.2 Evaluation Index for Position Classification

For evaluation we defined the estimation index *accuracy* since the number of participants in the Support position is not always the same as the

number of participants in the Oppose position. If the answers are grossly one-sided, the general accuracy does not work well, since the system can lead to a high score when it classifies all of the participants into the larger side. To minimize this potential bias, we defined an estimation index *accuracy* using the average of the accuracies for the Support/Oppose sets. The estimation index *accuracy* is defined as:

$$\text{accuracy} = \frac{1}{2} \left( \frac{|A_{\text{sup}} \cap S_{\text{sup}}|}{|A_{\text{sup}}|} + \frac{|A_{\text{opp}} \cap S_{\text{opp}}|}{|A_{\text{opp}}|} \right), \quad (2)$$

where  $A_{\text{sup}}$  and  $A_{\text{opp}}$  are the Support and Oppose participant sets in the answer set and  $S_{\text{sup}}$  and  $S_{\text{opp}}$  are the Support and Oppose participant sets generated by the system, respectively. For *accuracy*, we ignore “Unclear” users since the system is a two-class classifier.

### 4.3.3 Experimental Results

In the experiments we use the reaction coefficients  $r(i, j)$  calculated based on the results of the agree/disagree/neutral Classifier, and classify participants into Support/Oppose position sets using BiqMac. Since we assumed that the main topic of the debate is the first remark of the debate thread, we assume that the set which includes the author of the first remark as the “Support” set and the other set as the “Oppose” set<sup>5</sup>.

We conducted experiments for  $(\alpha, \beta, \gamma) = (1, 0, 0), (1, 0.5, 0), (1, 0.5, -1)$  in Eq. (1) to examine the dependency of the accuracy on the coefficients  $r(i, j)$ . We also conducted an experiment for  $(1, 1, 1)$ , which is regarded as a baseline method described in (Agrawal et al., 2003), since all of the reply actions represent “disagree” opinions for the previous remarks with these parameter. The experimental results are shown in Table 3.

The ideas other than ID 1 show better accuracy than the baseline and their accuracies tend to increase in the order of  $(1, 0, 0), (1, 0.5, 0), (1, 0.5, -1)$ . This result shows that the effectiveness of distinguishing between “disagree” and “agree” replies. This distinction makes it possible to introduce the constraint in which the user pairs

<sup>5</sup>For this reason, the values of the accuracies can be lower than 0.5.

Table 3: Accuracy of Support/Oppose position classification

ID	Baseline	(1,0,0)	(1,0.5,0)	(1,0.5,-1)
1	47.86	<b>66.67</b>	54.52	54.05
2	66.43	76.43	76.43	<b>89.29</b>
3	46.47	48.88	42.63	<b>55.45</b>
4	53.19	51.52	55.36	<b>77.60</b>
5	66.67	58.33	66.67	<b>75.00</b>

with “disagree” and “neutral” should be classified into opposing positions and user pairs with “agree” should be classified into same position in the Support/Oppose user sets.

At the same time, ID 1 shows lower accuracy for  $(1, 0.5, 0), (1, 0.5, -1)$  even though the accuracy of agree/disagree classifier is good. In idea ID 1, the number of remarks per participant is the lowest in data sets, so the errors of the Agreement/Disagreement classifier strongly affect the results of the Support/Oppose classifier.

## 5 Conclusion and Future Work

We have shown how to classify users in an online debate based on their general positions with regards to the main topic by the textual contents of their remarks and the link structure of their replies. The previous work used the assumption that the replies are usually disagreements and based on this assumption used a link-based method to classify the participants. However, in an online debate the replies are also used for clarifying previous remarks and quite often for supporting the previous ones. Our proposed method uses not only the link structure of the replies, but also the textual contents of the local agreement/disagreement positions between the remarks to boost the accuracy of the task of classifying users into the supporting and opposing parties.

The proposed method is based on the observation that it is easier to use the textual contents for classifying the local positions of a user’s replies with regards to the previous remarks, than to use them (e.g., by aggregating them) for classifying his/her global position with regards to the main topic of the debate. In our experiments, we used a rule-based classifier to classify the replies into

agree, disagree, and neutral (with regards to the previous replies) and used these classifier's result to determine the weight of the corresponding links in the link structure of the reply network. The max cut algorithm is then applied to the network, which results in a classification of the users into supporting or opposing parties (with regards to the main topic of the debate). The experiments indicate that the accuracies of the link-based method of (Agrawal et al., 2003) can be significantly increased by considering the textual contents of the replies.

There are several directions to extend our method. When an expression of opinion appears in a reply, we have to locate the target of the opinion. In the current method the target is determined by the ID of the remark pointed by the reply. When the ID is not available, we assume that the reply is with regards to the main topic. However, we also observed that even though a reply was directed to a particular remark, it often also contained opinions about the main topic. Identifying such replies can be used to yield higher accuracy in the classification task.

Much work remains for ultimate understanding of the participants' opinions in debate corpus. Understanding the reasons for the position for the main topic is one of the ways to understand their opinions and it may help to decide the next steps for companies or governments which held the debate sessions. An integrated system that includes a discussion system and an analysis system showing the ratio of positions and the reasons would support such purposes.

## Acknowledgments

The authors would like to acknowledge Kenji Hiramoto and Manabu Morita, who are responsible for the IdeaBox, for their helpful comments and conversation.

## References

Agrawal, Rakesh, Sridhar Rajagopalan, Ramakrishnan Srikant, and Yirong Xu. 2003. Mining news-groups using networks arising from social behavior. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 529–535, New York, NY, USA. ACM.

Dave, Kushal, Martin Wattenberg, and Michael Muller. 2004. Flash forums and forumreader: navigating a new kind of large-scale online discussion. In *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 232–241, New York, NY, USA. ACM.

Galley, Michel, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: use of bayesian networks to model pragmatic dependencies. In *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, pages 669–676, Morristown, NJ, USA. Association for Computational Linguistics.

Kanayama, Hiroshi, Tetsuya Nasukawa, and Hideo Watanabe. 2004. Deeper sentiment analysis using machine translation technology. In *COLING '04: Proceedings of the 20th international conference on Computational Linguistics*, page 494, Morristown, NJ, USA. Association for Computational Linguistics.

Millen, David R. and Michael A. Fontaine. 2003. Multi-team facilitation of very large-scale distributed meetings. In *ECSCW'03: Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work*, pages 259–275, Norwell, MA, USA. Kluwer Academic Publishers.

Murakami, Akiko, Katashi Nagao, and Koichi Takeda. 2001. Discussion Mining: Knowledge discovery from online discussion records. In *NLPRS Workshop XML and NLP, 2001*.

Murakami, Akiko, Tetsuya Nasukawa, Fusashi Nakamura, Hironori Takeuchi, Risa Nishiyama, Pnina Veisberg, and Hideo Watanabe. 2007. Innovation-Jam: Analysis of online discussion records using text mining technology. In *International Workshop on Intercultural Collaboration 2007 (IWIC2007)*.

Rendl, Franz, Giovanni Rinaldi, and Angelika Wiegele. 2010. Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations. *Math. Programming*, 121(2):307.

Thomas, Matt, Bo Pang, and Lillian Lee. 2006. Get out the 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.