Unifying Local and Global Agreement and Disagreement Classification in Online Debates

Jie Yin
CSIRO ICT Centre
NSW, Australia
jie.yin@csiro.au

Nalin Narang
University of New South Wales
NSW, Australia
nalinnarang@gmail.com

Paul Thomas
CSIRO ICT Centre
ACT, Australia
paul.thomas@csiro.au

Cecile Paris
CSIRO ICT Centre
NSW, Australia
cecile.paris@csiro.au

Abstract

Online debate forums provide a powerful communication platform for individual users to share information, exchange ideas and express opinions on a variety of topics. Understanding people's opinions in such forums is an important task as its results can be used in many ways. It is, however, a challenging task because of the informal language use and the dynamic nature of online conversations. In this paper, we propose a new method for identifying participants' agreement or disagreement on an issue by exploiting information contained in each of the posts. Our proposed method first regards each post in its local context, then aggregates posts to estimate a participant's overall position. We have explored the use of sentiment, emotional and durational features to improve the accuracy of automatic agreement and disagreement classification. Our experimental results have shown that aggregating local positions over posts yields better performance than non-aggregation baselines when identifying users' global positions on an issue.

1 Introduction

With their increasing popularity, social media applications provide a powerful communication channel for individuals to share information, exchange ideas and express their opinions on a wide variety of topics. An online debate is an open forum where a participant starts a discussion by posting his opinion on a particular topic, such as regional politics, health or the military, while other participants state their support or opposition by posting their opinions. Understanding participants' opinions in online debates has become an increasingly important task as its results can be used in many ways. For example, by analysing customers' online discussions, companies can better understand customers' reviews about their products or services. For government agencies, it could help gather public opinions about policies, legislation, laws, or elections. For social science, it can assist scientists to understand a breadth of social phenomena from online observations of large numbers of individuals.

Despite the potentially wide range of applications, understanding participants' positions in online debates remains a difficult task. One reason is that online conversations are very dynamic in nature. Unlike spoken conversations (Thomas et al., 2006; Wang et al., 2011), users in online debates are not guaranteed to participate in a discussion at all times. They may enter or exit the online discussion at any point, so it is not appropriate to use models assuming continued conversation. In addition, most discussions in online debates are essentially dialogic; participants could choose to implicitly respond to a previous post, or explicitly quote some content from an earlier post and make a response. Therefore, an assumption has to be made about what a participant's post is in response to, particularly when an explicit quote is not present; in most cases, a post is assumed to be in response to the most recent post in the thread (Murakami and Raymond, 2010).

In this paper, we address the problem of detecting users' positions with respect to the main topic in online debates; we call this the *global* position of users on an issue. It is inappropriate to identify each user's global position with respect to a main topic directly, because most expressions of opinion are made not

Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis, pages 61–69,
Jeju, Republic of Korea, 12 July 2012.

for the main topic but for posts in a *local* context. This poses a difficulty in directly building a global classifier for agreement and disagreement. We illustrate this with the example below. Here, the topic of the thread is "Beijing starts gating, locking migrant villages" and the discussion is started with a seed post criticising the Chinese government[1].

> **Seed post:** I'm most sure there will be some China sympathisers here justifying these actions imposed by the Communist Chinese government. ...
>
> **Reply 1:** Not really seeing a problem there. From you article. They can come and go. People in my country pay hundreds of thousands of pounds for security like that in their gated communities..
>
> **Reply 2:** So, you are OK with living in a Police State? ...

The author of Reply 1 argues that the Chinese policy is not as presented, and is in fact defensible. This opposes the seed post, so that the author's global position for the main topic is "disagree". The opinion expressed in Reply 2, however, is not a response to the seed post: it relates to Reply 1. It indicates that the author of Reply 2 disagrees with the opinion made in Reply 1, and thus indirectly implies agreement with the seed post. From this example, we can see that it is hard to infer the global position of Reply 2's author only from the text of their post. However, we can exploit information in the local context, such as the relationship between Replies 1 and 2, to indirectly infer the author's opinion with regard to the seed post.

Motivated by this observation, we propose a three-step method for detecting participants' global agreement or disagreement positions by exploiting local information in the posts within the debate. First, we build a local classifier to determine whether a pair of posts agree with each other or not. Second, we aggregate over posts for each pair of participants in one discussion to determine whether they agree with each other. Third, we infer the global positions of participants with respect to the main topic, so that participants can be classified into two classes: agree and disagree. The advantage of our proposed method is that it builds a unified framework which enables the classification of participants' local and global positions in online debates; the aggregation of local estimates also tends to reduce error in the global classification.

In order to evaluate the performance of our method, we have conducted experiments on data sets collected from two online debate forums. We have explored the use of sentiment, emotional and durational features for automatic agreement and disagreement classification, and our feature analysis suggests that they can significantly improve the performance of baselines using only word features. Experimental results have also demonstrated that aggregating local positions over posts yields better performance for identifying users' global positions on an issue.

The rest of the paper is organised as follows. Section 2 discusses previous work on agreement and disagreement classification. Section 3 presents our proposed method for both local and global position classification, which we validate in Section 4 with experiments on two real-world data sets. Section 5 concludes the paper and discusses possible directions for future work.

2 Related Work

Previous work in automatic identification of agreement and disagreement has mainly focused on analysing conversational speech. Thomas et al. (2006) presented a method based on support vector machines to determine whether the speeches made by participants represent support or opposition to proposed legislation, using transcripts of U.S. congressional floor debates. This method showed that the classification of participants' positions can be improved by introducing the constraint that a single speaker retains the same position during one debate. Wang et al. (2011) presented a conditional random field based approach for detecting agreement/disagreement between speakers in English broadcast conversations. Galley et al. (2004) proposed the use of Bayesian networks to model pragmatic dependencies of previous agreement or disagreement on the current utterance. These differ from our work in that the speakers are assumed to

[1] Spelling of the posts is per original on the website.

be present all the time during the conversation, and therefore, user speech models can be built, and their dependencies can be explored to facilitate agreement and disagreement classification. Our aggregation technique does, however, presuppose consistency of opinions, in a similar way to Thomas et al. (2006).

There has been other related work which aims to analyse informal texts for opinion mining and (dis)agreement classification in online discussions. Agrawal et al. (2003) described an observation that reply-to activities always show disagreement with previous authors in newsgroup discussions, and presented a clustering approach to group users into two parties: support and opposition, based on reply-to graphs between users. Murakami and Raymond (2010) proposed a method for deriving simple rules to extract opinion expressions from the content of posts and then applied a similar graph clustering algorithm for partitioning participants into supporting and opposing parties. By combining both text and link information, this approach was demonstrated to outperform the method proposed by Agrawal et al. (2003). Due to the nature of clustering mechanisms, the output of these methods are two user parties, in each of which users most agree or disagree with each other. However, users' positions in the two parties do not necessarily correspond to the global position with respect to the main issue in a debate, which is our interest here. Balasubramanyan and Cohen (2011) proposed a computational method to classify sentiment polarity in blog comments and predict the polarity based on the topics discussed in a blog post. Finally, Somasundaran and Wiebe (2010) explored the utility of sentiment and arguing opinions in ideological debates and applied a support vector machine based approach for classifying stances of individual posts. In our work, we focus on classifying people's global positions on a main issue by exploiting and aggregating local positions expressed in individual posts.

3 Our Proposed Method

To infer support or opposition positions with respect to the seed post, we propose a three-step method. First, we consider each post in its local context and build a local classifier to classify each pair of posts as agreeing with each other or not. Second, we aggregate over posts for each pair of participants in one discussion to determine whether they agree with each other. Third, we infer global positions of participants with respect to the seed post based on the thread structure.

3.1 Classifying Local Positions between Posts

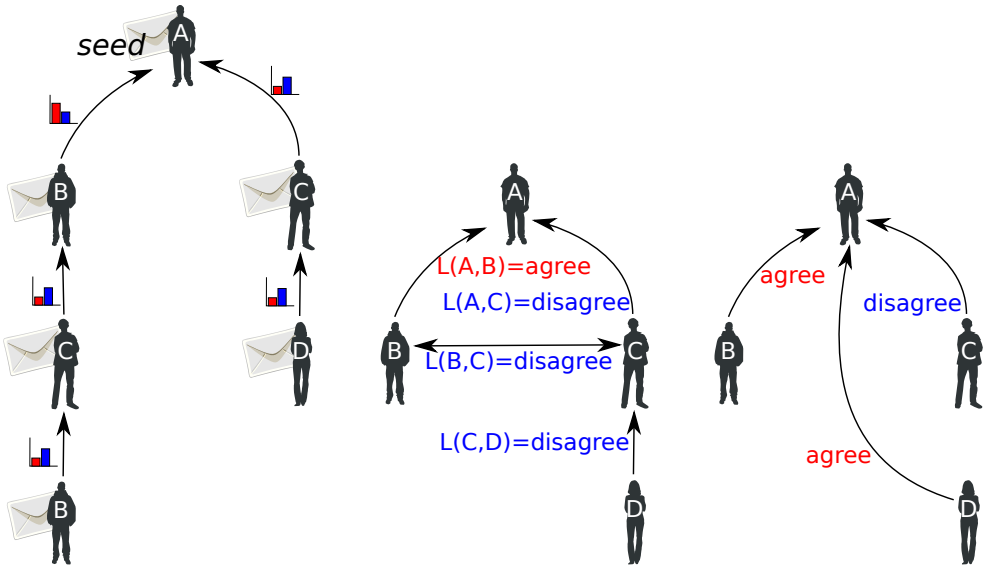
To classify local positions between posts, we need to extract the reply-to pairs of posts from the threading structure. The web forums we work with tend not to present thread structure, so we consider two types of reply-to relationships between individual posts. When a post explicitly quotes the content from an earlier post, we create an *explicit* link between the post and the quoted post. When a post does not contain a quote, we assume that it is a reply to the preceding post, and thus create an *implicit* link between the two adjacent posts. After obtaining explicit/implicit links, we build a classifier to classify each pair of posts as agreeing or disagreeing with each other.

3.1.1 Features

To build a classifier for identifying local agreement and disagreement, we explored different types of features from individual posts with the aim to understand which have predictive power for our agreement/disagreement classification task.

Words We extract unigram and bigram features to capture the lexical information from each post. Since many words are topic related and might be used by both parties in a debate, we mainly use unigrams for *adjectives*, *verbs* and *adverbs* because they have been demonstrated to possess discriminative power for sentiment classification (Benamara et al., 2007; Subrahmanian and Regorgiato, 2008). Typical examples of such unigrams include "agree", "glad", "indeed", and "wrong". In addition, we extract bigrams to capture phrases expressing arguments, for example, "don't think" and "how odd" could indicate disagreement, while "I concur" could indicate agreement.

Sentiment features In order to detect sentiment opinions, we use a sentiment lexicon referred to as SentiWordNet (Baccianella et al., 2010). This lexicon assigns a positive and negative score to a large number of words in WordNet. For example, the



(a) Estimate $P(y|x)$ for each post

(b) Aggregate these over pairs of users to get local agreement $L(m, n)$

(c) Infer the global position of each user by walking the tree

Figure 1: Local agreement/disagreement and participants' global positions. We first estimate $P(y|x_i, x_j)$, the probability of two posts x_i and x_j being in agreement or disagreement with each other, then aggregate over posts to determine $L(m, n)$, the position between two users. Finally, we infer the global position for any user by walking this graph back to the seed.

word "odd" has a positive score of 1.125, and a negative score of 1.625. To aggregate the sentiment polarity of each post, we calculate the overall positive and negative scores for all the words that can be found in SentiWordNet, and use these two sums as two features for each post.

Emotional features We observe that personal emotions could be a good indicator of agreement/disagreement expression in online debates. Therefore, we include a set of emotional features, including occurrences of emoticons, number of capital letters, number of foul words, number of exclamation marks, and number of question marks contained in a post. Intuitively, use of foul words might be linked to emotion in a visceral way, which if used, could be a sign of strong argument and disagreement. The presence of question marks could be indicative of disagreement, and the use of exclamation marks and capital letters could be an emphasis placed on opinions.

Durational features Inspired by conversation analysis (Galley et al., 2004; Wang et al., 2011), we extract durational features, such as the length of a post in words and in characters. These features are analogous to the ones used to capture the duration of a speech for conversation analysis. Intuitively, people tend to respond with a short post if they agree with a previous opinion. Otherwise, when there is a strong argument, people tend to use a longer post to state and defend their own opinions. Moreover, we also consider the time difference between adjacent posts as additional features. Presumably, when a debate is controversial, participants would be actively involved in the discussions, and the thread would unfold quickly over time. Thus, the time difference between adjacent posts would be smaller in the debate.

3.1.2 Classification Model

We use logistic regression as the basic classifier for local position classification because it has been demonstrated to provide good predictive performance across a range of text classification tasks, such as document classification and sentiment analysis (Zhang and Oles, 2001; Pan et al., 2010). In addition to the predicted class, logistic regression can also generate probabilities of class memberships,

which are quite useful in our case for aggregating local positions between participants.

Formally, logistic regression estimates the conditional probability of y given $\mathbf{x}$ in the form of

$$P_{\mathbf{w}}(y = \pm 1|\mathbf{x}) = \frac{1}{1 + e^{-y\mathbf{w}^T\mathbf{x}}}, \quad (1)$$

where $\mathbf{x}$ is the feature vector, y is the class label, and $\mathbf{w} \in R^n$ is the weight vector. Given the training data $\{\mathbf{x}_i, y_i\}_{i=1}^l$, $\mathbf{x}_i \in R^n$, $y_i \in \{1, -1\}$, we consider the following form of regularised logistic regression

$$\min_{\mathbf{w}} f(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l} \log\left(1 + e^{-y_i\mathbf{w}^T\mathbf{x}_i}\right), \quad (2)$$

which aims to minimise the regularised negative log-likelihood of the training data. Above, $\mathbf{w}^T\mathbf{w}/2$ is used as a regularisation term to achieve good generalisation abilities. Parameter $C > 0$ is a penalty factor which controls the balance of the two terms in Equation 2. The above optimisation problem can be solved using different iterative methods, such as conjugate gradient and Newton methods (Lin et al., 2008). As a result, an optimal estimate of $\mathbf{w}$ can be obtained.

Given a representation of a post $\mathbf{x}_m$, we can use Equation 1 to estimate its membership probability of belonging to each class, $P(\textit{agree}|\mathbf{x}_m)$ and $P(\textit{disagree}|\mathbf{x}_m)$, respectively.

3.2 Estimating Local Positions between Participants

After obtaining local position between posts, this step aims to aggregate over posts to determine whether each pair of participants agree with each other. The intuition is that, in one threaded discussion, most of the participants tend to retain their positions in the course of their arguments. This assumption holds for the ground-truth annotations we have obtained in our data sets. Given local predictions obtained from the previous step, we adopt the weighted voting scheme to determine the local position for each pair of participants. Specifically, given a pair of users i and j, we aggregate over all the reply-to posts between them to calculate the overall agreement score $r(i, j)$ as follows:

$$r(i,j) = \sum_{k=1}^{N(i,j)} P(\textit{agree}|\mathbf{x}_k) - \sum_{k=1}^{N(i,j)} P(\textit{disagree}|\mathbf{x}_k). \quad (3)$$

Here, $N(i, j)$ denotes the number of post exchanges between users i and j, and $r(i, j)$ indicates the degree of agreement between users i and j. Let $L(i, j)$ denote the local position between two users i and j. If $r(i, j) > 0$, we have $L(i, j) = \textit{agree}$, that is, user i agrees with user j. Otherwise, if $r(i, j) \leq 0$, we have $L(i, j) = \textit{disagree}$, that is, user i disagrees with user j.

Let us consider the example in Figure 1(a) and 1(b). There are two posts exchanged between users B and C. For each of these posts, two probabilities of class membership can be obtained:

$$P(\textit{agree}|\mathbf{x}_1) = 0.1,\ P(\textit{disagree}|\mathbf{x}_1) = 0.9,$$
$$P(\textit{agree}|\mathbf{x}_2) = 0.3,\ P(\textit{disagree}|\mathbf{x}_2) = 0.7.$$

Then we can calculate the agreement score $r(B, C)$ between users B and C by aggregating over two posts, that is, $r(B, C) = (0.1 + 0.3) - (0.9 + 0.7) = -1.2 < 0$. We can conclude that user B disagrees with user C in the threaded discussion and that $L(B, C) = \textit{disagree}$.

3.3 Identifying Participants' Global Positions

After estimating local positions between participants, we now can infer a participant's global support or opposition position with regards to the seed post. For this purpose, a thread structure must be considered. A thread begins with a seed post, which is further followed by other response posts. Of these responses, many employ a quote mechanism to explicitly state which post they reply to, whereas others are assumed to be in response to the most recent post in the thread. We construct a tree-like thread structure by examining all the posts in a thread and determining the parent of each post. Then, traversing through the thread structure from top to bottom allows us to infer the global position of each user with respect to the seed post. When there is more than one path from the seed to a user, the shortest path is used to infer the user's global position on the main issue.

We illustrate this inference process using Figure 1, an example thread with four users and six posts.

Let $L(m, n)$ denote the local position between two users m and n. In the figure, the local position between user B and user A (the author of the seed post), $L(A, B)$, is in agreement, while users B and C, A and C, as well as C and D each disagree. Walking the shortest path between D and the seed in Figure 1(a), we have $L(C, D) =$ *disagree* and $L(A, C) =$ *disagree*, so we can infer that the global position between user D and user A is in agreement. That is, user D agrees with the seed post. Had the local position between user A and user C, $L(A, C)$, been in agreement, then we would have concluded that user D disagrees with the seed post.

4 Experiments

In this section, we describe our experiments on two real-world data sets and report our experimental results for local and global (dis)agreement classification.

4.1 Data Sets

We used two data sets to evaluate our proposed method in our experiments. They were crawled from the U.S. Message Board (`www.usmessageboard.com`) and the Political Forum (`www.politicalforum.com`). The two data sets are referred to as **usmb** and **pf**, respectively, in our discussion. The detailed characteristics of the two data sets are given in Table 1.

Table 1: Characteristics of data sets

	usmb	**pf**
# of threads	88	33
# of posts	818	170
# of participants	270	103
Mean # of posts per thread	9.3	5.2
Mean # of participants per thread	3.1	3.1
Mean # of posts per participant	3.0	1.7

For the evaluation, each post was labelled with two annotations. The first was a global annotation with respect to the thread's seed post, and the other was a local annotation with respect to the immediate parent. Seed posts themselves were not annotated, nor were they classified by our algorithms.

Global annotations were made by two postgraduate students. Each was instructed to read all the posts in a thread, then label each post with *agree* if the author agreed with the seed post; *disagree* if they disagreed; or *neutral* if opinions were mixed or unclear. The annotators used training data until they reached 85% agreement, then annotated posts separately. At no time were they allowed to confer. Local annotations were reverse-engineered from these global annotations. The ratio of posts annotated as *agree* to those as *disagree* is about 2 to 1 on both datasets.

For our proposed three-stage method, local annotations were taken as input to train the classifier and then used as ground truth to evaluate the performance of local agreement/disagreement classification, while the global annotations were only used to evaluate our final accuracy of global agreement/disagreement identification. In contrast, the baseline classifiers that we compare against for global classification were directly trained and evaluated using global annotations.

4.2 Evaluation Metrics

We used two evaluation metrics to evaluate the performance of agreement/disagreement classification. The first metric is accuracy, which is computed as the percentage of correctly classified examples over all the test data:

$$\text{accuracy} = \frac{|\{\mathbf{x} : \mathbf{x} \in \mathcal{D}_{test} \bigcap h(x) = y\}|}{|\mathcal{D}_{test}|},$$

where $\mathcal{D}_{test}$ denotes the test data, y is the ground truth annotation label and $h(\mathbf{x})$ is the predicted class label.

Accuracy can be biased in situations with uneven division between classes, so we also evaluate our classifiers with the F-measure. For each class $i \in \{\text{agree}, \text{disagree}\}$, we first calculate precision $P(i)$ and recall $R(i)$, and the F-measure is computed as

$$F1(i) = \frac{2P(i)R(i)}{P(i) + R(i)}.$$

For our binary task, we report the average F-measure over both classes.

4.3 Local Agree/Disagree Classification

In our experiments, we used the implementation of L2-regularised logistic regression in Fan et al. (2008) as our local classifier. For each data set,

Table 2: Classification performance for local (dis)agreement

	usmb		**pf**	
	Accuracy	F-measure	Accuracy	F-measure
Naive Bayes, all features	0.46	0.42	0.52	0.51
SVM, all features	0.56	0.60	0.55	0.52
Logistic regression, all features	0.62	0.65	0.68	0.77

Table 3: Feature analysis for local (dis)agreement using logistic regression

	usmb		**pf**	
	Accuracy	F-measure	Accuracy	F-measure
words	0.50	0.55	0.55	0.63
words, sentiment	0.53	0.59	0.61	0.71
words, sentiment, emotional	0.54	0.51	0.55	0.65
words, sentiment, durational	0.58	0.61	0.64	0.72
words, sentiment, emotional, durational	0.62	0.65	0.68	0.77

we used 70% of posts as training and the other 30% were held out for testing. We compared regularised logistic regression against two baselines: naive Bayes and support vector machines (SVMs), which have been used for (dis)agreement classification in previous works (Thomas et al., 2006; Somasundaran and Wiebe, 2010). For SVMs, we used the toolbox LIBSVM in Chang and Lin (2011) to implement the classification and probability estimation. We tuned the parameter C in regularised logistic regression and SVM, using cross-validation on the training data, and thereafter the optimal C was used on the test data for evaluation.

Table 2 compares the local classification accuracy of the three methods on data sets **usmb** and **pf**, respectively. We can see from the table that logistic regression outperforms naive Bayes and SVM on the two evaluation metrics for local classification. Although logistic regression and SVM have been shown to yield comparable performance on some text categorisation tasks Li and Yang (2003), in our problem, regularised logistic regression was observed to outperform SVM for local (dis)agreement classification.

Experiments were also carried out to investigate how the performance of local classification would be changed by using different types of features. Table 3 shows the classification accuracy of logistic regression using different types of features on the two data sets. We can see from the table that using both words and sentiment features can improve the performance as compared to using only words features. On the **usmb** dataset, adding emotional features slightly improves the accuracy but degrades F-measure, while on the **pf** dataset, it degrades on accuracy and F-measure. In addition, durational features substantially improve the classification performance on the two metrics. Overall, the highest classification accuracy and F-measure can be achieved by using all four types of features.

4.4 Global Support/Opposition Identification

We also conducted experiments to validate the effectiveness of our proposed method for global position identification. Table 4 reports the performance of global classification using the three methods on the two data sets. Classifiers "without aggregation" were trained directly on global annotations, without considering local positions at all; those "with aggregation" were developed with our three-stage method, estimating global positions by aggregating local positions $L(m, n)$.

As before, logistic regression generally outperforms SVM or naive Bayes classifiers, although SVM does well on **usmb** when aggregation (via $L(m, n)$) is used. Although SVM scores well for

Table 4: Classification performance for global (dis)agreement

	usmb		pf	
	Accuracy	F-measure	Accuracy	F-measure
Without aggregation				
Naive Bayes, all features	0.42	0.41	0.48	0.47
SVM, all features	0.62	0.46	0.68	0.40
Logistic regression, all features	0.60	0.63	0.65	0.77
With aggregation				
Naive Bayes, all features	0.54	0.67	0.65	0.70
SVM, all features	0.64	0.77	0.48	0.60
Logistic regression, all features	0.64	0.77	0.68	0.76

classification accuracy without aggregation, it has degraded and classifies everything as the majority class in these cases. The F-measure is correspondingly poor due to a low recall. This observation is consistent with the findings reported in Agrawal et al. (2003).

In all cases — bar logistic regression on the **pf** set — aggregation of local classifications improves the performance of global classification. This is more marked in the **usmb** data set, which has slightly more exchanges between each pair of users (mean 1.33 per pair per topic, vs. 1.19 for the **pf** data set) and therefore more potential for aggregation. We believe that this improvement is because local classification is sometimes error prone, especially when opinions are not expressed clearly in individual posts. If so, and assuming that users tend to retain their stances within a debate, aggregation can "wash out" local classification errors.

5 Conclusion and Future Work

In this paper, we have proposed a new method for identifying participants' agreement or disagreement on an issue by exploiting local information contained in individual posts. Our proposed method builds a unified framework which enables the classification of participants' local and global positions in online debates. To evaluate the performance of our proposed method, we conducted experiments on two real-world data sets collected from two online debate forums. Our experiments have shown that regularised logistic regression is useful for this type of task; it has a built-in automatic feature selection by assigning a coefficient to each specific feature, and directly estimates probabilities of class memberships, which is quite useful for aggregating local positions between users. Our feature analysis has suggested that using sentiment, emotional and durational features can significantly improve the performance over only using word features. Experimental results have also shown that, for identifying users' global positions on an issue, aggregating local positions over posts results in better performance than no-aggregation baselines and that more benefit seems to accrue as users exchange more posts.

We consider extending this work along several directions. First, we would like to examine what other factors would have predictive power in online debates and thus could be utilised to improve the performance of agreement/disagreement classification. Second, we have so far focused on classifying users' positions into two categories: agree and disagree. However, there do exist a portion of posts falling into the neutral category; that means posts/users do not express any position towards an issue. We will explore how to extend our computational framework to classify the neutral class. Finally, in online debates, it is not uncommon to have off-topic or topic-drift posts, especially for long threaded discussions. Off-topic posts are the ones totally irrelevant to the main issue being discussed, and topic-drift posts usually exist when the topic of a debate has shifted over time. Taking these posts into consideration would increase the difficulty of automatic agreement and disagreement classification, and therefore it is another important issue we plan to investigate.

References

Rakesh Agrawal, Sridhar Rajagopalan, Ramakrishnan Srikant, and Yirong Xu. 2003. Mining newsgroups using networks arising from social bahavior. In *Proceedings of the 12th International World Wide Web Conference*, pages 529–535, Budapest, Hungary, May.

Stefano Baccianella, Andrea Esuli, , and Fabrizio Sebastiani. 2010. SENTIWORDNET 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on Internatinal Language Resources and Evaluation*, pages 2200–2204, Valletta, Malta, May.

Ramnath Balasubramanyan and William W. Cohen. 2011. What pushes their buttons? Predicting comment polarity from the content of political blog posts. In *Proceedings of the ACL Workshop on Language in Social Media*, pages 12–19, Porland, Oregon, USA, June.

Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgiato, and V. S. Subrahmanian. 2007. Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Boulder, CO, USA, March.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(27):1–27.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Michel Galley, Kathleen McKeown, Julia Hirschberg, and Elizabeth Shriberg. 2004. Identifying agreement and disagreement in conversational speech: Use of Bayesian networks to model pragmatic dependencies. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, pages 669–676, Barcelona, Spain, July.

Fan Li and Yiming Yang. 2003. A loss function analysis for classification methods in text categorisation. In *Proceedings of the 20th International Conference on Machine Learning*, pages 472–479, Washington, DC, USA, July.

Chih-Jen Lin, Ruby C. Weng, and S. Sathiya Keerthi. 2008. Trust region Newton method for large-scale logistic regression. *Journal of Machine Learning Research*, 9:627–650.

Akiko Murakami and Rudy Raymond. 2010. Support or oppose? Classifying positions in online debates from reply activities and opinion expressions. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 869–875, Beijing, China, August.

Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *Proceedings of the 19th International World Wide Web Conference*, pages 751–760, Raleigh, NC, USA, April.

Swapna Somasundaran and Janyce Wiebe. 2010. Recognizing stances in ideological on-line debates. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, Los Angeles, CA, USA, June.

V. S. Subrahmanian and Diego Regorgiato. 2008. AVA: Adjective-verb-adverb combinations for sentiment analysis. *Intelligent Systems*, 23(4):43–50.

Matt Thomas, Bo Pang, and Lillian Lee. 2006. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 327–335, Sydney, Australia, July.

Wen Wang, Sibel Yaman, Kristin Precoda, Colleen Richey, and Geoffrey Raymond. 2011. Detection of agreement and disagreement in broadcast conversations. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 374–378, Porland, Oregon, USA, June.

Tong Zhang and Frank J. Oles. 2001. Text categorisation based on regularised linear classification methods. *Information Retrieval*, 4(1):5–31.