Utilizing the Quoting System of Online Web Forums to Estimate User Agreement

Manos Karvounis
Thodoris Georgiou
Yannis Ioannidis

Department of Informatics and Telecommunications, University of Athens
manosk@di.uoa.gr, t.georgiou@di.uoa.gr, yannis@di.uoa.gr

Abstract
Online web forums have become one of the most popular places for opinion exchange and thus, complex agreement-disagreement relationships are formed between their users. In this paper, a novel method is proposed that creates a graph representation of a discussion by using the quoting mechanism offered by almost all web forums. By applying a variant of coloring on the previous graph, we are able to separate the participating users into groups, depending on their opinion homogeneity. The proposed method needs no training and can be readily applied to any discussion thread, yielding results with competitive accuracy and high recall. We have also developed an algorithm for extracting specific agreements-disagreements between pairs of users.

Keywords: agreement discovery, web forums, graph coloring

1. Introduction
Online web forums are websites that are specifically created to accommodate dynamic discussions between an unrestricted number of participants. These discussions are usually separated into discussion threads (autonomous conversations with a certain topic). Internet users can participate in a thread by writing a message and publishing it by making a post. They can also respond to a previous post of another user by quoting the target post and replying to it. This well defined and discussion-oriented structure of online web forums encourages lengthy debates, in which a rich network of relationships between the participating users is formed. The most prominent of these relationships are connections of agreement and disagreement.

Extracting these relationships using natural language processing techniques can be a very difficult task in the general case. If a method only utilizes words that directly express agreement-disagreement, then many possibilities are left out, especially cases where someone responds with an extensive argument and expresses a complex viewpoint. In an experiment conducted in [Jeong (2001)], it was found that the participants of an online discussion were more likely to respond with an argument
instead of a simple agreement-disagreement phrase, even when responding to such a phrase. Nevertheless, even if a method uses more features to identify the orientation of such complex arguments, it unavoidably becomes more complex and needs extensive training data.

In this paper, instead of using natural language processing methods, we propose viewing a discussion thread as a social network. This way, we are able to create a graph representation of a thread, which encodes information about the exchange of posts between the participating users. To create this graph, we are especially interested in replies (posts that quote other posts), since these messages show direct exchange of opinions between two users. By applying a variant of coloring on the resulting graph, we are able to perform an initial grouping of users in two sets. The semantics of this procedure is that users inside the same set express more homogenous opinions.

The main advantage of the proposed method is that it can be used in any discussion thread immediately, without the need for training, and still yield accurate results. This is possible because the coloring algorithm that is applied on the discussion graph is based on observations that are true for the majority of the discussion threads containing debates between users. Chapter 3 focuses on explaining the proposed method and establishing the validity of these observations.

We have also developed, in a concurrent project ([Georgiou et. Al. (2010)]), a method that is able to identify subtle topic changes in a thread discussion, so users that exchange opinions on precisely the same topic are grouped together. These results, combined with the results of the previously discussed algorithm on opinion homogeneity clustering, are used to extract more specific agreements and disagreements between pairs of users. This method is briefly described in chapter 4.

2. Previous Work

Research on agreement-disagreement discovery in multiparty conversations has been performed in the works of [Hillard et. Al. (2003)], [Galley et. Al. (2004)], [Hahn et. Al. (2006)] and [Germesin et. Al. (2009)]. All these works use testing data from live meetings and apply their methods on the recordings of the conversations. In [Hillard et. Al (2003)] mainly word-based and prosodic (e.g. conversation pauses) features are used in an unsupervised machine learning approach. In [Galley et. Al. (2004)] adjacency pair information is extracted and used to construct a conditional markov model. In [Hahn et. Al. (2006)] the skewness of the available test data is addressed by using semi-supervised contrast classifiers. Finally, in [Germesin et. Al. (2009)] the problem of identifying the target of the expressed opinions is also examined.

Research on opinion mining in web forums, with the aim of extracting information on the forum itself and its participants and not reviews and customer feedback, is as of
yet fairly limited. Nevertheless, a very comprehensive work on feature selection for the purpose of sentiment analysis in web forums can be found in [Abbashi et. Al. (2008)]. In [Shi et. Al. (2009)] a method is proposed that can identify the sentiment of posts by looking for words with known sentiment near nouns that relate to the thread’s topic. For the purpose of identifying disagreements in discussions, the work of Spertus in [Spertus (1997)] can provide interesting insight into the ways people use to express tense disagreement and hostility when posting messages.

3. Opinion Homogeneity Clustering

3.1 Representation of Forum Discussions as Graphs

The core idea of our approach is to shift the focus from what is being written in a discussion to how the participants interact with each other. This is typically the focus of tasks that utilize social networking techniques ([Wasserman et. Al. (1994)]).

Social networks use a graph to represent the relationships between the participating parts (actors). In our approach, we create a graph for each given discussion thread, where the nodes are the users that take part in the discussion and the edges indicate that the connected users have posted replies to each other’s posts. The edges are weighted, with the weight being the number of such replies. For example, in Figure 1 below, users A and B have replied to each other’s various posts 35 times.

![Figure 1](image1.png)

*Figure 1. A graph representation of a discussion. For clearness reasons, edges with weights less than 4 are not shown. Usernames were replaced by letters for privacy reasons.*
This representation can be used to extract potential agreements-disagreements by applying a variant of graph coloring. We support this claim with the observations that are presented in the following section.

3.2 Observations and Preconditions

Preliminary examinations of various discussion threads have indicated that the resulting graphs tend to be bipartite, with the occasional exception of some edges, usually with small weights. The two sets of nodes of the bipartite graph seemed to represent the two “opposing factions” of users, especially when the discussion was highly polarized. For example, Figure 1 shows the graph representation of a thread on a religious debate, more specifically on whether the Christian belief is compatible with the notion of punishment in hell. Users A, D, I expressed the viewpoint that it is compatible and had a heated debate with users C, B, G, H, F, E who opposed this viewpoint.

By looking closely to this and other similar cases, we came to the conclusion that a significant number of replies between two users is an indication that they disagree. To support this, we manually evaluated a sample set of 813 replies (posts that quoted another post). It was found that 74% were expressing disagreement, 4% agreement and 22% were either neutral or couldn’t be evaluated even by a human reader. Furthermore, it was observed that the more the replies between two users the more likely they are to disagree. Agreements were mostly implicit, stemming from common disagreement with other users rather than direct responses.

Another important observation is that, in most cases, a small amount of users dominate the discussion inside a thread. These power users post most of the messages and frequently exchange messages between them, forming the backbone of the viewpoints that are expressed. Typically, the rest of the users have more homogenous opinions with some of the power users than others, even if they do not totally agree with any of them. To support this observation we processed 3110 posts in the context of various threads and found that, on average, 25% of the users were responsible for 65% of the posts within a thread.

These observations seem to be valid regardless of topic, domain or participants, provided that the discussion has the nature of a debate and no objectively correct answers exist. Also, the agreement-disagreement relationships between the users must not radically change as the discussion develops, since this will confuse the results of the proposed method.

3.3 Proposed Method

The proposed algorithm receives a discussion thread as input and creates the equivalent graph representation, as described in the previous section. After applying
the algorithm described in Table 1, the output is two sets of users which is interpreted as follows:

- The users inside each set are generally assumed to have homogenous opinions regarding the discussion topic. This means that two users inside the same set do not disagree (they either agree or are neutral).
- Two users that belong to different sets are assumed to have heterogeneous opinions. This means that they do not agree (they either disagree or are neutral).

Table 1. The algorithm applied on the graph representation of a discussion thread to produce the two user sets

<table>
<thead>
<tr>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>find the edge with the greatest weight</td>
</tr>
<tr>
<td>insert the connected nodes to different sets</td>
</tr>
<tr>
<td>mark this edge as visited</td>
</tr>
<tr>
<td>while there are nodes that have no set</td>
</tr>
<tr>
<td>for every edge not yet visited in descending order of weight</td>
</tr>
<tr>
<td>if exactly one of the connected nodes has a set</td>
</tr>
<tr>
<td>insert the node that has no set to the other set</td>
</tr>
<tr>
<td>mark this edge as visited</td>
</tr>
<tr>
<td>break this loop and restart</td>
</tr>
<tr>
<td>else if both nodes have a set</td>
</tr>
<tr>
<td>mark this edge as visited</td>
</tr>
<tr>
<td>continue the loop</td>
</tr>
<tr>
<td>if no edges were visited in the previous loop</td>
</tr>
<tr>
<td>take the edge with the greatest weight not yet visited</td>
</tr>
<tr>
<td>insert the connected users to different sets</td>
</tr>
<tr>
<td>mark this edge as visited</td>
</tr>
</tbody>
</table>

In the rest of this chapter we will focus on explaining the details of this algorithm and looking into how it brings into effect the observations presented in the previous section. More specifically, the rationale behind the following choices will be examined:

- Why the edges are processed in descending order of weight
- Why only edges that connect a user that has a set with a user that doesn’t have a set are examined in each iteration
- Why the choice to insert a user in a set is never changed

We will also use a real-life example to better illustrate the way the algorithm works. The graph representation of this example is shown in Figure 2. Usernames have been replaced by random character strings for privacy reasons.

The choice to process the edges in descending order of weight is based on the observation that many replies between two users is an indication of disagreement. As
a result, we try to examine these more certain disagreements first, before moving to edges with smaller weights. Also, edges with significant weights connect power users, who, through their interaction, form the backbone of the discussion. By processing the edges that connect power users as soon as possible, we form this “viewpoint backbone” early on, so the rest of the users will be inserted in sets depending on their interaction with these power users.

**Figure 2.** A graph representation of a discussion about gun control in the United States. Bolder lines indicate stronger edges.

In Figure 2, we can see that certain edges dominate the graph’s relations, namely \([\{\text{NA, Wi}\}, \{\text{NA, Ti}\}, \{\text{Do, si}\}, \{\text{NA, Br}\}, \{\text{zs, Wi}\}, \{\text{Ol, si}\}, \{\text{Ol, Wi}\}, \{\text{Et, Wi}\}]\). These edges connect power users and will be processed as soon as possible.

More specifically, in the first step the edge with the greatest weight is considered, in our case \([\text{NA, Wi}]\). This edge corresponds to the most prevailing conversation inside the discussion thread, and it would be very unusual for an agreement to be able to sustain such a lengthy exchange of opinions. As a result, this edge is considered to be representing the most certain disagreement available. So, NA is inserted in set 1 and Wi in set 2.

In the following steps, the rest of the edges are examined in descending order of weight. Nevertheless, we are not only interested in the weight of an edge, but also in
its distance from the disagreements we have already found. This is an important choice, as it leads to the propagation of disagreements we have already found, rather than considering every edge as a disagreement a-priori. So, in our example, the next edges processed would be \{NA, Ti\} and then \{NA, Br\}, despite the fact that \{Do, si\} is stronger. This way the disagreement between NA and Wi is propagated to the users that have a more direct interaction with them, since the further away in the graph from NA and Wi, the more likely it is for the users to be talking about something slightly different. By propagating the disagreements we have found to the local discussion first, we are more certain that users talk about the same topic and so, we are able to make more straightforward decisions about their agreement-disagreement relationships.

Another important characteristic of the proposed algorithm is that it follows the greedy paradigm. As a result, when a choice is made about the set of a user, this choice is never changed. This of course helps in speeding up the algorithm significantly, but also in preventing over-fitting. To better illustrate this, consider users Ol and zs who, as the algorithm progresses, will end up having edges with significant weights to users of their own set. It could be argued that these users are outliers and should be considered for insertion in a separate third and fourth set respectively. Nevertheless, careful examination of the discussion thread by a human reader determined that users Ol and zs belonged in set 1, where users had a more pro-gun viewpoint. The edges connecting them to users of their own set were either expressing agreement or minor disagreement.

Observations like the previous one were the most important factor that led to the choice of using only two sets and not changing the set choices once they were made. More sets would create an over-fitting effect that is not desirable. Instead, we are mostly interested in the general picture, an understanding of the homogeneity of the major viewpoints expressed in a discussion. Under these requirements, minor or partial disagreements must be tolerated, in order to emphasize the more significant ones.

The final output of the algorithm is the following sets:

- NA, zs, Et, Ol, Do, No, J, T, Ch
- Wi, Ti, Br, Pe, si, Ga, P, B, Su, Se

Indeed, after examination of the actual discussion by a human reader, it was determined that the algorithm’s results were mostly accurate. More specifically, most of the users in set 1 had a clear pro-gun viewpoint, whereas users in set 2 had a more anti-gun stance.

4. Agreement and Disagreement Discovery

In [Georgiou et. Al. (2010)] we have developed a method that can identify sub-discussions, denoted by slight changes of topic inside a larger discussion thread. The
output is groups of users who participate in the same part of the conversation and thus, exchange opinions on precisely the same topic, directly or indirectly. In Figure 3 a sample output of this algorithm is shown.

![Figure 3. Users in the same group take part in the same discussion inside a larger thread. Users are represented by letters.](image)

In Figure 4 the output of the opinion homogeneity clustering algorithm is shown, applied on the same thread as Figure 3.

![Figure 4. The output of the opinion homogeneity clustering algorithm when applied on the same thread as that of Figure 3.](image)

The results of these two methods can be combined in order to extract agreements and disagreements between pairs of users. More specifically

- Two users in the same group in Figure 3 exchange opinions on the same topic.
- Two users in the same set in Figure 4 have homogenous opinions, whereas two users on different sets have heterogeneous opinions.
- As a result, two users in the same group in Figure 3 and in the same set in Figure 4 are assumed to agree, whereas if they are in different sets they are assumed to disagree. This is a result derived from logical necessity, since homogenous opinions expressed on the same topic is the definition of agreement, and heterogeneous ones of disagreement.

The previous algorithm, if applied on the results shown in Figures 3 and 4, would produce the following output:

- A disagrees with B and C. B and C agree with each other.
- E disagrees with F.
- G disagrees with I and J. H disagrees with I and J. G and H agree with each other. I and J agree with each other.

5. Evaluation of the Proposed Method

Preliminary evaluation of the opinion homogeneity clustering algorithm presented in chapter 3, has shown a high degree of accuracy in recognizing potential
disagreements. With the help of an independent tester, we have manually annotated approximately 3000 posts, in the context of 30 discussion threads. A summary of the results is shown in Table 2 below. Please note that the evaluation process is still work in progress.

The manually annotated test data consisted of pairs of users that, according to the testers’ opinions, seemed to directly or indirectly disagree. Both accuracy and recall are measured inside each evaluated discussion thread separately. If two users that are marked as disagreeing in the test data are:

- In different sets in the algorithm’s results, this is considered as correct
- In the same set in the algorithm’s results, this is considered as incorrect
- Not present in the algorithms results, this is considered as unresolved

Accuracy is measured as the amount of correct results divided by the amount of correct plus incorrect results. Recall is defined as the amount of correct plus incorrect results divided by the total amount of pairs present in the test data.

Table 2. Evaluation of the opinion homogeneity clustering algorithm across different discussions and domains

<table>
<thead>
<tr>
<th>Domain</th>
<th>Accuracy</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Philosophical Debates</td>
<td>84%</td>
<td>96%</td>
</tr>
<tr>
<td>Health Debates</td>
<td>92%</td>
<td>90%</td>
</tr>
<tr>
<td>Scientific Debates</td>
<td>87%</td>
<td>95%</td>
</tr>
<tr>
<td>Abortion Debates</td>
<td>90%</td>
<td>94%</td>
</tr>
<tr>
<td>Crime Debates</td>
<td>84%</td>
<td>92%</td>
</tr>
<tr>
<td>Religion Debates</td>
<td>80%</td>
<td>93%</td>
</tr>
<tr>
<td>Society Debates</td>
<td>86%</td>
<td>75%</td>
</tr>
<tr>
<td>Total</td>
<td>86%</td>
<td>91%</td>
</tr>
</tbody>
</table>

Of particular importance is the algorithm’s ability to retain a stable and high degree of accuracy across different domains. This is something to be expected, since no topic information or natural language processing is used.

6. Future Work

Our first priority is to further evaluate the methods we propose, in order to gain a better idea of their results and accuracy. In the near future, we are also aiming at examining the graph representations of thread discussions more closely, since interesting patterns of the way forum users interact with each other seem to emerge and useful information can be extracted from them.
In respect to the homogeneity clustering algorithm presented in chapter 3, we would like to further enhance it in order to treat strongly not bipartite graphs more robustly. We are already looking into ways to recognize typical discussion patterns, which can then be turned into certain graph formations. Success in this effort can enable us to develop highly accurate agreement-disagreement algorithms that can identify even subtle differentiations of opinion, without the need for natural language processing.

7. References

Abbashi, A., Chen, H., Salem, A. (2008), Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums, ACM Transactions on Information Systems (TOIS)


