DETERMINING LEADERSHIP IN CONTENTIOUS DISCUSSIONS

Siddharth Jain  
USC Information Science Institute  
University of Southern California  
4676 Admiralty Way  
Marina Del Rey, CA 90292  
sjain@isi.edu

Eduard Hovy  
Language Technology Institute  
Carnegie Mellon University  
500 Forbes Avenue  
Pittsburgh, PA 15213  
hovy@cmu.edu

ABSTRACT

Participants in online decision making environments assume different roles. Especially in contentious discussions, the outcome often depends critically on the discussion leader(s). Recent work on automated leadership analysis has focused on collaborations where all the participants have the same goal. In this paper we focus on contentious discussions, in which the participants have different goals based on their opinion, which makes the notion of leader very different. We analyze discussions on the Wikipedia Articles for Deletion (AfD) forum. We define two complementary models, Content Leader and SilentOut Leader. The models quantify the basic leadership qualities of participants and assign leadership points to them. We correlate the leaders’ ranks produced by the two models using the Spearman Coefficient. We also propose a method to verify the quality of the leaders identified by each model.

Index Terms: discussion leader discovery, discussion participant role, contentious discussion, Wikipedia, social multimedia, natural language processing

1. INTRODUCTION

What is a leader? As social beings, humans are adept at recognizing and responding to leadership in group settings. Yet leadership is surprisingly hard to define or recognize computationally.

In this paper we describe methods to identify leaders in contentious (not cooperative) online discussions. The explosive growth of social multimedia, whether targeted or untargeted (for example, personal email and tweets respectively), has made available a wealth of material that can be used for study. Using the Wikipedia Articles for Deletion (AfD) corpus, we in this paper:

- define two complementary models of leadership in online discussions,
- develop methods to automatically identify and quantify various characteristics of leaders,
- merge these results to identify discussion leaders,
- measure the correlation between leaders identified by the models,
- propose a method to verify the quality of users identified by the models.

2. RELATED WORK

Identifying leaders in different online social environments has attracted much research in recent times. The types of Opinion Leaders that have been explored the most can be categorized as: leaders in social networks and leaders in online discussion groups. Lu et al. [1] propose the LeaderRank algorithm to identify leaders in social network by quantifying the user influence over network. Clemson and Evans [2] study a network version of minority games to identify the followers in the network and in turn identify the leaders by determining the users who follows the smallest number of users. Fazeen et al. [3] present context dependent and independent models to identify leaders, lurkers, spammers and group associates in social networks. Tsai et al. [4] use probabilistic time based graph propagation model to build action specific influence chains to determine leaders in social communities. Song et al. [5] propose a data structure called the Longest Sequence Phrase tree to measure similarities between comments made by different users in order to identify positive opinion leader group in discussion forums. Another method to identify groups and leaders in discussions is by Arvapalli and Ravi using the K-means clustering algorithm [6], which uses users’ provided weights for support/ opposition to different opinions to formulate faction groups and then identify leaders in each group as the ones having the most support for their group’s opinion. Catizone et al. [7] use social network analysis methods to infer social and instrumental roles and relationships in online discussions.

The work that comes closest to ours is by Bracewell et al. [8], who define psychologically motivated acts to determine social goals of individuals based on discourses’ intentional structure. Apart from network analysis methods, all the methods discussed here use annotators to characterize user behavior in the social network or discussion forums and use that to train learning systems. Ours is the first model to quantify leadership automatically and identify the leaders in an unsupervised way.
3. CORPUS

3.1. Articles for Deletion

In our analysis we focus on the discussions on the Wikipedia forum called “Wikipedia: Articles for Deletion (AfD)” [9]. Wikipedia, being a very large peer production system, has its own decentralized governance system to maintain the quality of articles created by the users. Any Wikipedia user can nominate any article on Wikipedia for consideration for deletion. Articles that are nominated for deletion are typically discussed for a minimum of 7 days if they do not qualify for the speedy deletion or speedy keep criteria. During this period, feedback from the community is sought to reach a consensus. Each user can participate in the discussion by declaring his stance towards the discussion and then adding comments.

A user participating in the discussion can take one of the following stances:

- **Keep**: suggesting that article should be kept.
- **Delete**: suggesting that article should be deleted.
- **Merge**: suggesting that article should be merged with another existing article.
- **Redirect**: suggesting that article should be redirected to another existing article.
- **Transwiki**: suggesting that article should be moved to another Wikipedia project.

In the end, a Wikipedia user with administrative authority reviews the discussion and declares the decision that reflects the consensus of the users participating in the discussion. The result of the discussions can be one of **Keep, Delete, Merge, Redirect, Transwiki, Withdraw** (suggesting that nominator withdrew his nomination), or **No Consensus** (suggesting that no consensus can be reached as per the discussion).

3.2. Dataset

AfD discussions are public, written with good language, well-structured, and freely available. We collected all AfD discussions that took place in the period of January 1, 2009 to June 30, 2012. The data contains 92066 discussions with 781909 distinct comments and 47066 distinct users. Each discussion includes the title of the article in question, the nominator of the discussion, all the comments in the discussion, and the final outcome alone with the admin who imposed it. Each comment in the discussion includes the user who posted it, the stance of the user if specified, the time of the comment, and the level of the comment. The level refers to information about whether the comment is a new thread or a reply to some previous comment/reply. The data extracted was verified by cross checking 25 random discussions with the original discussions on the website.

The corpus consists of a total of 457668 distinct instances of stance. Since merge, redirect, and transwiki combine to a total of only 7.16% and as they rarely occur together in the same discussion, we simplified the data by combining them to the single stance compromise. Similarly, the outcome of the discussion was also converted to compromise if appropriate. Also, as an outcome, withdraw implicitly means keeping the article. Therefore, it was converted to keep. We create a timeline for each discussion based on the chronological order of each comment. Whenever users state their stance for the first time in the discussion, it is propagated to their subsequent comments unless they explicitly state a change in stance.

3.3. Contentious Discussion

The outcomes of discussions on AfD are not decided based on the number of votes for each stance as stated in the guidelines for the deletion process, but depend on the nature of the discussion. Hence the content, flow, and structure of the discussion become very significant. We define a contentious discussion as any discussion containing at least one user each for at least two stances. But not all contentious discussions are suited for leadership analysis. If a stance shows an overwhelming majority, then there is no incentive for a user to take up the leadership role, as he would be sure that his goal would be achieved.

| Table 1. Stance majority as factor for outcome of discussion

<table>
<thead>
<tr>
<th>Majority</th>
<th># of discussions</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>30–40%</td>
<td>463</td>
<td>38.23</td>
</tr>
<tr>
<td>40–50%</td>
<td>996</td>
<td>29.21</td>
</tr>
<tr>
<td>50–60%</td>
<td>6833</td>
<td>48.48</td>
</tr>
<tr>
<td>60–70%</td>
<td>11487</td>
<td>70.00</td>
</tr>
<tr>
<td>70–80%</td>
<td>9557</td>
<td>86.73</td>
</tr>
<tr>
<td>80–90%</td>
<td>10467</td>
<td>93.77</td>
</tr>
<tr>
<td>90–100%</td>
<td>43646</td>
<td>97.69</td>
</tr>
</tbody>
</table>

Table 1 shows the relationship between the degree of majority for any stance in a discussion and the outcome being the same as the majority for all the discussions where some consensus was concluded by the admin for the discussion (i.e., neglecting the discussions where the outcome was no consensus). The majority cannot be less than 1/3 as there are only 3 possible stances. Also, to resolve ties, we select the majority stance as the one with the larger length of content in support of it. Evident from the table, we can see that when the majority is over 60%, the outcome of the discussion is very much in favor of the majority. This reduces the incentive for leadership behavior of any participant in the discussion.

To fortify our claim, we analyze the average contribution by users in discussions. Table 2 shows the average number of comments by a user in discussions versus the majority of any stance in it. It shows that when the discussion is very contentious, users contribute more in order to sway the outcome in their favor.

As described above, a participant can reply to a comment/reply from another participant. We analyze the
degree to which participants get direct replies from other participants having a different stance in the discussion. This shows the amount of direct opposition to participants’ arguments, which signifies the level of contention in the discussion. For discussions having majority more than 60%, 88.65% of comments didn’t get any direct opposing reply. However, for discussions whose majority comments number less than or equal to 60%, only 60.46% of the comments didn’t get any direct opposing reply. This shows that in contentious discussions participants are not only focused on proving their arguments, but also are equally focused on disproving others’ arguments, which indicates potential leadership.

Table 2. Majority vs Average user comments

<table>
<thead>
<tr>
<th>Majority</th>
<th>Average user comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>30–40%</td>
<td>2.23</td>
</tr>
<tr>
<td>40–50%</td>
<td>1.95</td>
</tr>
<tr>
<td>50–60%</td>
<td>2.11</td>
</tr>
<tr>
<td>60–70%</td>
<td>1.65</td>
</tr>
<tr>
<td>70–80%</td>
<td>1.52</td>
</tr>
<tr>
<td>80–90%</td>
<td>1.41</td>
</tr>
<tr>
<td>90–100%</td>
<td>1.32</td>
</tr>
</tbody>
</table>

The analysis shown here suggests that leadership behavior is highly influenced by the degree of contention in the discussion. Therefore, for all the experiments described below we only consider discussions in which the majority stance forms less than or equal to 60% of the total.

4. MODELS

We define two principal leadership models of participants based on the basic qualities that reflect the leadership behavior of a person in contentious discussions. The idea is to quantify the degree of leadership for each participant by assigning leadership points for such behavior instances.

4.1. Content Leader

Language can be a great indicator to recognize leadership qualities of a person. As suggested by the name of the model, the Content Leader model quantifies the language use of a participant to assign leadership points. This model is built upon two basic characteristics of a leader: 1) Encourage others to follow his arguments (Attract Followers) and 2) Counter the arguments from the opposing groups (Counterattack).

4.1.1. Attract Followers (AF)

One of the prime qualities of a leader is to attract followers by convincing them with his arguments. This can be very useful in building majority in contentious discussions and thus swaying the outcome of the discussion in his/her favor. We model this attribute by matching the n-grams (word sequences) of a user against the n-grams used by another user having the same stance. The user who encourages other users to support his argument by making them reuse the same phrases acquires leadership points.

4.1.2 Counterattack (CA)

This attribute of the model quantifies the quality to stand up against opponents and try to nullify the arguments presented in oppose to the user’s stance. We model this attribute by matching the n-grams of a user against the n-grams of users holding a different stance. The user who counterattacks acquires leadership points. In addition, the user who is getting counterattacked also gets some leadership points. This is because of his/her ability to attract the attention of opposing groups, which implies a significant contribution in regards to the discussion.

Both these attributes implicitly quantify another ability of a leader, which is having command over the course of the discussion. The fact that other users having the same or different stances are using the same argumentation words implies that the leader is guiding the discussion.

We define the equation for leadership points for Content Leader model for any user in any discussion by:

\[ \alpha \times AF + \beta \times CA_{oppose} + \gamma \times CA_{opposed} \]

where

\[ AF = n\text{-gram weight for n-grams used by other users having same stance, found in word n-grams produced by the user} \]

\[ CA_{oppose} = n\text{-gram weight for n-grams used by user found also in n-grams of users with different stance} \]

\[ CA_{opposed} = n\text{-gram weight for n-grams used by other users having different stance, found in n-grams of user} \]

\[ \alpha, \beta, \gamma = \text{weights of corresponding attributes (see below)} \]

4.2. SilentOut Leader

The SilentOut Leader model quantifies the ability of a leader to silent opposing users out with his arguments. This is modeled by two attributes: 1) Giving arguments that cannot be countered (Factual Arguments) and 2) Winning the small battles in the discussion (Small Wins).

4.2.1. Factual Arguments (FA)

Based on the analysis presented earlier, we learn that users counter the arguments given by opposing stance users more often in highly contentious discussions. So, if a user presents an argument that none of the users from the opposing stances attack, it shows the quality of the argument which silences out the opposing users. This contributes to the leadership qualities of the user in the sense that he can give factual arguments that relate to the discussion and can have an impact on the outcome of the discussion. We model this attribute by giving a constant amount of leadership points to users for each comment that elicits no reply from
any opposing stance user. Fortunately, the effects of spammers in discussions on this attribute of the model are negligible since Wikipedia admins identify such spammers and strike their comment out. Using this delete mark of admins, we identify such comments and ignore them.

4.2.2. Small Wins (SW)

This attribute of the model refers to the ability of a leader to silence out other users by countering their individual arguments and thus winning a small battle over other users. This counterargument not only nullifies the original argument from the opposing user but also strengthens the arguments for leader’s own stance. To model this attribute, we divide the discussion into small argument sections. Each argument section contains an original argument and at least one reply from a user from opposing stance. For each such argument section, the user who has the last say (i.e., whose reply gets no counterargument) acquires a constant amount of leadership points.

The equation for the leadership points for SilentOut Leader model for any user in any discussion is given by

\[ \alpha \times FA + \beta \times SW \]

where
- \( FA = \# \) of comments from user which didn’t have any reply from any user from opposing stance
- \( SW = \# \) of small battles user won
- \( \alpha, \beta = \) weights of corresponding attributes (see below)

5. EXPERIMENT

5.1. Setup

We create a timeline for each discussion based on the chronological order of comments. We process each discussion comment by comment, so at any step, we have the list of all the users who have participated in the discussion so far, their current stance, the n-grams that they have used so far, and the nesting level for each comment (which defines argument sections).

To determine the weight of any n-gram, we use the Inverse Document Frequency (idf) of that n-gram across all the discussions in the whole corpus. The idf value of an n-gram denoted by \( t \) is given by

\[ idf(t) = \log \frac{|N|}{|\{n \in N : t \in n\}|} \]

where
- \( |N| = \# \) of discussions in the corpus
- \( |\{n \in N : t \in n\}| = \# \) of discussions where the n-gram \( t \) appears

We use unigrams, bigrams, and trigrams as part of our n-grams and calculate the weights across the whole corpus, not only the contentious discussions. To deal with closed class words that are not important in calculating the leadership points, we ignore all n-grams which appear in at least 25% of the discussions. The motivation behind using idf values is to assign more importance to topic words that may be more relevant to the discussion.

We also maintain a window to determine a list of active participants for Content Leader analysis. We don’t want any participants who haven’t contributed in a long span in the discussion to get leadership points for new comments. Thus, if a participant hasn’t contributed in the window of consideration, he/she won’t get any leadership points for any of the new comments. As the average length of the discussions in consideration is 13.25, we keep the window size to be 10.

5.2. Content Leader calculation

We process discussions comment by comment, extracting n-grams from the text and storing them in the bag of n-grams for the user. We also keep track of the stance of each user as the discussion progresses. Given a new comment by user B, we extract the n-grams in this comment. For each user \( A \), who has already participated in the discussion, we match the n-grams used by user B to the n-grams in the bag of user \( A \).

If user \( B \) has the same stance as user \( A \), then for each n-gram of user \( B \) that matches some n-gram in the bag of user \( A \), user \( A \) gets \( \alpha \) times the idf weight of the n-gram amount of leadership points, per the AF formula. If user \( B \) has a stance different from user \( A \), then for each n-gram of user \( B \) that matches some n-gram in the bag of user \( A \), user \( A \) gets \( \gamma \) times the idf weight of the n-gram and user \( B \) gets \( \beta \) times the idf weight of the n-gram as their respective leadership points, per the CA formula.

Here, the absolute values of \( \alpha, \beta, \) and \( \gamma \) are not significant. But their relative values are important because that reflects their relative importance to the corresponding attributes of the model. Because of the lack of gold standards for leaders, we set the values for coefficients manually. We present in Table 5 the analysis of different coefficients values along with the results.

5.3. SilentOut Leader calculation

To calculate the SilentOut Leader in a discussion, we divide the whole discussion into argument sections based on the nesting level of each comment in the discussion. Thus, each comment and the replies to that comment form one argument section. Now for each argument section, we identify the user who initiated the argument section, say user \( A \).

If an argument section contains no replies from any user with a stance different from user \( A \) (i.e., the argument section shows no contention), then user \( A \) acquires \( \alpha \) leadership points, per the FA formula. And if an argument section does have replies from any user with a stance different from user \( A \) (i.e., the argument section shows contention), then the user who replied last in the
chronological order gets β leadership points, per the SW formula.

Similar to the Content Leader model, the absolute values of α and β are not significant. But their relative values are important because they reflect the relative importance of the corresponding attributes of the model. Because of the lack of gold standards for leaders, we set the values for coefficients manually. We present in Table 6 the analysis of different coefficients values along with the results.

6. RESULTS

6.1. Correlation between Content Leaders and SilentOut Leaders

Using the two leadership models, we calculate leadership points for each user in each of the 8292 contentious Wikipedia AfD discussions. We then aggregate the leadership points for user across all the discussions and calculate average leadership points per discussion. Using the average score, we create a ranked list of leaders for both the models. We compare the ranked list using the Spearman’s rank correlation coefficient. For any two ranked lists, the coefficient is calculated as

\[
\rho = 1 - 6 \sum \frac{d^2}{N(N^2 - 1)}
\]

where

\(d\) = difference in statistical rank of corresponding user

\(N\) = total number of users in consideration

Table 3. Spearman correlation between models

<table>
<thead>
<tr>
<th>Min participation</th>
<th># of users</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9489</td>
<td>0.23</td>
</tr>
<tr>
<td>5</td>
<td>2166</td>
<td>0.30</td>
</tr>
<tr>
<td>10</td>
<td>1144</td>
<td>0.46</td>
</tr>
<tr>
<td>20</td>
<td>533</td>
<td>0.48</td>
</tr>
<tr>
<td>50</td>
<td>183</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Table 3 shows the Spearman’s rank correlation between the Content Leader and SilentOut Leader models. The first column limits the users by minimum number of different discussions they must have participated in to be considered for correlation calculation. The second column shows the number of users satisfying the criteria for minimum number of discussions and the last column shows the correlation coefficient for those users. Positive correlation coefficient implies that the models complement each other for identifying same users as leaders.

6.2. Predicting outcome of the discussion

Our work presents a possible method to test the quality of leaders identified by each model. The criterion to qualify as a quality leader is to be able to turn the outcome of the discussion into one’s favor. We compare our models with some other naïve models:

- **Majority Stance**: The outcome of the discussion is predicted as the stance with the majority of votes. To resolve ties, we select the majority stance as the one which has larger length of content in support of it. If a tie cannot be resolved, we predict the stance with equal probability.
- **Majority Content Stance**: The outcome of the discussion is predicted as the stance which has largest number of words in support. To resolve ties, we select the stance with the majority of votes. If a tie cannot be resolved, we predict the stance with equal probability.
- **Talkative Leader**: The user who has contributed the largest number of words in the discussion is chosen as the leader of the discussion. The outcome of the discussion is predicted as the stance of that leader.
- **Content Leader**: The user with the highest leadership points per the Content Leader model is chosen as the leader of the discussion. The outcome of the discussion is predicted as the stance of that leader.
- **SilentOut Leader**: The user with the highest leadership points per the SilentOut Leader model is chosen as the leader of the discussion. To resolve ties for leadership points, we choose as leader the user who has contributed a larger length of content in the discussion. The outcome of the discussion is predicted as the stance of that leader.
- **Content and SilentOut Leader**: We calculate the leadership points for each user in the discussion using both the models separately and then combine the leadership points of the model using the following equation

\[
\alpha \ast CL + \beta \ast SL
\]

where

- CL = Leadership points from Content Leader model
- SL = Leadership points from SilentOut Leader model

We set the values of α and β manually because of the lack of gold standard data. After experimenting with various values for α and β, they are set to 1 and 2 respectively. The user with the highest combined leadership points is chosen as the leader of the discussion. To resolve ties, the user having greater SilentOut leadership points is chosen to be the leader because of the higher individual accuracy of the model. The outcome of the discussion is predicted as the stance of their leader.

Table 4 shows the comparison of accuracies for predicting the outcome of Wikipedia AfD contentious discussions based on the different models described above. Table 5 and 6 show an analysis of relative values of model coefficients on the outcome prediction accuracy for Content Leader and SilentOut Leader models respectively. We choose the coefficient values related to the best accuracy for each model to create the leadership rankings, which are used to calculate Spearman correlation between the two models.
Table 4. Comparison of outcome prediction accuracy

<table>
<thead>
<tr>
<th>Model</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority Stance</td>
<td>45.59</td>
</tr>
<tr>
<td>Majority Content Stance</td>
<td>42.56</td>
</tr>
<tr>
<td>Talkative Leader</td>
<td>40.28</td>
</tr>
<tr>
<td>Content Leader</td>
<td>60.90</td>
</tr>
<tr>
<td>SilentOut Leader</td>
<td>65.01</td>
</tr>
<tr>
<td>Content and SilentOut Leader</td>
<td>68.34</td>
</tr>
</tbody>
</table>

Table 5. Comparison of accuracy for different coefficient values for Content Leader model

<table>
<thead>
<tr>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>53.59</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>52.88</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>2</td>
<td>59.60</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>55.53</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>54.68</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>55.48</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>60.90</td>
</tr>
</tbody>
</table>

Table 6. Comparison of accuracy for different coefficient values for SilentOut Leader model

<table>
<thead>
<tr>
<th>α</th>
<th>β</th>
<th>% Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>64.05</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>65.01</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>62.25</td>
</tr>
</tbody>
</table>

7. DISCUSSION

The models presented in this paper show a simple method to quantify leader qualities. The initial results show promise in terms of identifying individuals who affect the flow and the outcome of the discussions taking place on the Wikipedia AfD forum. The Spearman correlation between the leaders identified by the models show that the models complement each other. The increase in correlation for users who participate in more discussions implies that the models can identify users who have better leadership qualities in long span on participation.

An analysis of coefficient values for the Content Leader model indicates that users who tend to counter arguments from alternative stance users prove more effective in influencing the outcome of the discussion than users who just present arguments and rely on others to believe them without much further effort. The same analysis for the SilentOut Leader model similarly implies that users who win small battles by either countering others’ arguments or defending their own argument prove more effective in influencing the outcome of the discussion. Overall, the preponderance of β over α and γ indicates that the strongest leaders are those who actively address opposing points of view. This conclusion is stable regardless of the specific values of α, β and γ, which are varied from 1 to 4.

One limitation of this study is that we do not have any ground truth data for users exhibiting leadership behavior in such discussions. Thus, we must consider annotating discussions for leaders in them and then measure accuracies of models to identify the same leaders. Also, we would like to generalize the models to determine leadership behaviors in other genres of discussions like Q/A systems and discussions for collaborative tasks.

8. CONCLUSION

In this study, we construct two models to find leaders in contentious discussions. The models quantify the basic qualities of the leader behavior and assign leadership points to users participating in the discussions for each instance of such behavior. The initial results show promise for identifying users who succeed in influencing the outcome of the discussions. Also, the similarity measure between the models implies that the models complement each other to identify leaders based on different leadership qualities.

9. REFERENCES