6

Searching the Net for Differences of Opinion

WARREN SACK, JOHN KELLY, AND MICHAEL DALE

1 Introduction

Political theorists, at least since John Stuart Mill in his book *On Liberty* (1859), have asserted that exposure to conflicting viewpoints is beneficial for democracy. Through exposure to political viewpoints contrary to their own, citizens are said to gain political tolerance and an understanding of opposing rationales. Recent empirical work has confirmed these assertions (Fishkin 1992; Mutz 2002). However, there is no clear means by which a citizen can find opposing opinions. Factors such as the consolidation of media ownership (Bagdikian 2004), neighborhood segregation (by, for example, race, and class), lack of weak ties in personal and cross-community-oriented social networks (Putnam 2000; Granovetter 1973), proliferation of ideologically exclusive weblogs and radio and television talk shows, and recent technological developments that allow the 'filtering' of Internet-distributed news (Sunstein 2001), all make it difficult for individual citizens to find significantly different opinions. Contrary to Negroponte (1995), we posit the development of a software technology to facilitate the construction of a 'Daily Not Me', a sort of 'search engine' that, when given a topic (e.g., 'abortion'), will return a range of diverse opinions about the topic (e.g., 'pro-choice' and 'pro-life').

In this chapter, we present some preliminary results towards this long-term goal. Our work bootstraps recent, prior work in which one of the

Online Deliberation: Design, Research, and Practice.
Todd Davies and Seeta Peña Gangadharan (eds.).

co-authors used qualitative content analysis to characterize the political leanings of 120 prolific Usenet newsgroup authors (Kelly 2004). Software was developed to automatically download, from a Usenet newsgroup archive, tens of thousands of discussion threads containing over one million individual messages. Within these threads of discussion, we were able to find several thousand 'mixed exchanges' in which known discussants (i.e., two or more discussants identified by Kelly) of differing political opinion exchanged messages. We have performed an empirical analysis of the structural characteristics (e.g., size, branching factor) of the discussion threads surrounding these mixed exchanges. Our goal is to identify a set of computable, search heuristics that might be employed in a 'Daily Not Me' technology for finding opposing, political viewpoints as expressed in the archives of online discussion groups.

We understand this work to be complementary to the work of Fishkin (1991) and others who have created new environments and situations where deliberative discussion can take place. We hypothesize that in vast, online discussion spaces—like the space of Usenet newsgroups—there must exist places, or at least moments, when deliberative discussion already takes place 'in the wild'. We envision a search engine that, when given a topic, will find likely threads of discussion where opposing opinions are being or have been expressed. In this chapter, we report on our initial efforts to implement the first preprocessing step of such a search engine. We need to identify one or more quick and relatively accurate heuristics that can be used to comb through a large database of newsgroup or weblog postings to identify likely places of political exchange. The output of the mechanisms we describe here will be the input to further processing steps of the search engine that perform a detailed analysis of the contents of the messages. In short, the heuristics described here are 'triage' techniques intended to narrow down which message threads should be given more detailed analyses.

First, we give an overview of the newsgroup messages we have examined and briefly describe the results of a previous study by one of the co-authors upon which we rely for the present work (Kelly 2004). Second, we describe a set of independent variables associated with the discussion threads. Our dependent variable concerns whether or not liberals and conservatives exchanged messages in a discussion thread. We seek a heuristic in which some combination of easily measured, independent variables can be used to predict the likelihood that a thread contains an exchange of views between at least one liberal discussion participant and one conservative discussant. Third, we present such a thread categorization heuristic as a simple discriminant function. We further simplify this function by eliminating some of the independent variables that are closely correlated with others. Finally, we present our conclusions and briefly discuss future work.

2 Messages, Discussion Threads, and Newsgroups

Kelly (2004) read several thousand posts made to six Usenet newsgroups: (1) alt.fan.noam-chomsky, (2) alt.politics.bush, (3) alt.politics.democrats.d, (4) alt.politics.economics, (5) talk.abortion, and (6) talk.politics.mideast. All of these newsgroups are public, online discussions archived on thousands of newsgroup (NNTP) servers throughout the Internet. Kelly was able to identify about twenty high-frequency posters from each of the six newsgroups. Extensive study of the messages posted by these 119 frequent participants led to the articulation of a set of political categories and the identification of the political category associated with 97 of the 119 posters. The political point of view of 22 of the posters was uncharacterizable.

We will not review Kelly's results in this chapter but rather explain how we have incorporated a simplified version of some of his results into the present study. While Kelly identified twelve different political positions occupied by newsgroup participants, we have (perhaps too insensitively) coerced these twelve positions into just three categories: left, right, and unrecognized. Thus, each of the participants studied by Kelly has been labeled as unrecognizable or political left or right.

Recall that a discussion thread is constituted from an initial message, all of the replies to the initial message, all of the replies to these replies, etc. Our aim has been to study the structure of those discussion threads in which at least one known person of the left exchanged a message with at least one known person of the right. We call an exchange of messages between posters of opposing political positions a 'mixed exchange'. We are interested in threads that contain one or more mixed exchanges because they are potentially deliberative exchanges. We hope to be able to formulate heuristics to automatically detect threads that are likely to contain a mixed exchange.

Because Kelly made known to us the political position of 119 frequent Usenet newsgroup posters, it was a straightforward task to download another set of threads—from the same time period—that contain messages posted by one or more of the these known participants. We downloaded over one million messages from an NNTP server but chose to focus on a subset of about sixteen hundred (1664 out of 25,590) discussion threads in which at least two messages were posted to the thread and in which at least two-thirds (66%) of the messages posted in the threads were posted by one or more of our known participants. A total of 13,156 messages were posted to these 1664 discussion threads.

3 Variables: On the Structure of Discussion Threads

From a graph theoretic perspective, discussion threads are trees because when one hits the 'reply' button in an email program, one is replying to one and only one previous message.

We can therefore define a set of variables that characterize the size and shape of the discussion threads:

M: the number of messages posted to a thread;

L: the number of leaves in a thread tree (leaves are messages that received no replies);

P: the number of people who posted a message to a thread;

maxMp: the maximum number of messages posted by one person to a thread;

maxD: the maximum depth from the root of the tree (i.e., the initial post) to one of the leaves of the thread tree;

meanD: the mean depth from the root to the leaves of the tree;

maxB: the maximum branching factor in the tree (corresponding to the message in the thread with the greatest number of replies);

meanB: the mean branching factor in the tree;

meanMp: the mean number of messages posted by a person participating in the thread; and,

meanT: the mean amount of time (in seconds) between messages posted to the thread.

In addition we assigned a score to each thread, where a score of '1' indicates a mixed exchanged (as defined above): a person of the left replied to a message from a person of the right or vice versa. A score of '0' indicates that no such exchange happened in the thread. Scores of greater than one occurred when more than one mixed exchange occurred. We calculated a Spearman's r correlation between each of our independent variables and the score. A linear regression model works quite well for threads with 25 or fewer messages (correlation 0.72). But, the linear model does not seem to fit as well for threads of size larger than 25 messages. Examination of the correlation for this subset of threads ($25 \leq M$) shows it to be weaker (correlation = 0.48).

It is unfortunate that the correlation between M and a thread's score is weak for large discussion threads, because we need a model that will work for large threads as well as for relatively small threads. Large threads are of interest because they are more likely than small threads to contain a deliberatively elaborated point of view. While small threads containing one or more long messages might contain a detailed explanation of someone's point of view, it is only through an extended back-and-forth with an interlocutor that the strengths and weaknesses of a point of view can be unpacked and explored in detail. So, we assume that long threads are more likely to be representative of some sort of deliberative exchange than short, small threads.

Second, recall that our immediate goal is to find a set of quick and computationally inexpensive heuristics for predicting if a thread is likely to contain a mixed exchange and thus for determining if more computational resources should be devoted to analyzing the thread in detail. A linear model (like this correlation) would roughly predict that we should look at all of the large threads and none of the small ones. However, simply because they contain a large number of messages, large threads are computationally expensive to analyze in detail. If we can eliminate even some of the large threads, then we are likely to save many computational resources in the subsequent phases of analysis. Consequently, we desire a model that works for small and large threads but especially for large threads.

4 A Thread Categorization Model: Search Heuristics

To create a model that will work for small and large discussion threads, we first simplify the problem. Rather than attempting to predict the number of mixed exchanges in a thread, we will be satisfied with sorting threads into one of two categories: (1) those containing mixed exchanges; and, (2) those containing no mixed exchanges. Consequently, the problem we now face is this: Can a categorization function (i.e., a *discriminant function*) be designed such that, given a thread, when it is greater than zero it is more likely that the thread contains one or more mixed exchanges, and, when it is zero or less than zero, it is more likely that the thread does not contain a mixed exchange?

This can be formalized as follows. Associated with each thread is a vector of independent variables, as detailed in the first part of this chapter (M, L, P, maxMp, maxD, meanD, maxB, meanB, meanMp, meanT). We are exploring 1664 discussion threads, thus we can order the thread trees from 1 to 1664. For a given thread tree j, in this order, we will denote the vector of independent variables simply as v_j. Since we are examining thread trees in which participants provided a political position known to us in 66% of the messages posted, each of these threads has an associated score. However, we are restricting our interest to the distinction between those threads with scores greater than zero (score > 0) versus those threads with scores of zero (score = 0).

Using the associated vectors and scores for our 1664 threads, we estimate the following two sets of conditional probabilities:

$P(\mathbf{v}_j|\text{score=0}) = \prod_{k=1..12} p(\mathbf{v}_k|\text{score=0})$; and,
$P(\mathbf{v}_j|\text{score>0}) = \prod_{k=1..12} p(\mathbf{v}_k|\text{score>0})$.

Thanks to Bayes formula, we can convert the estimated prior probabilities (i.e., in which we know the score) into posterior probabilities (in which we want to predict the score). So, our estimated discriminant function is this:

$g(\mathbf{v}_j) = P(\text{score>0}|\mathbf{v}_j) - P(\text{score=0}|\mathbf{v}_j)$
where if $g(\mathbf{v}_j) \geq 0$ then the score is more likely to be positive;
else the score more likely to be zero.

But this estimated discriminant function cannot be applied to discussion threads outside our original set of 1664 thread trees unless the unseen discussion thread has a vector associated with it that exactly matches the vector of some tree in our original set of trees. We, rather crudely, address this problem by dividing the values for each variable into equally populated quartiles that we call small, medium, large and extra large. For instance, the quartile divisions for M, the number of messages in the thread trees are small ($M < 3$); medium ($M = 3$); large ($3 < M < 6$); and, extra large ($M \geq 6$). This allows one to see, for example, that if a thread has an extra large number of messages, then it is more likely to have a positive score (i.e., more like to contain one or more mixed exchanges) than to have a score of zero.

Given these definitions and this simplification of values into quartiles a discriminant function can be calculated and then tested against the same 1664 thread trees to get some idea of how accurate it might be. Using all variables in the function, we find that it predicts the correct category (either score = 0 or score > 0) for 1156 out of the 1664 thread trees (accuracy of 69%). But, since we are searching for threads likely to contain a mixed exchange, the power of the model is better measured according to the usual criteria of information retrieval where recall denotes the completeness of the retrieval and precision denotes the purity of the retrieval.[1] Using Kelly's

[1] 'Consider an example information request I (of a test reference collection) and its set R of relevant documents. Let |R| be the number of documents in this set. Assume that a given retrieval strategy (which is being evaluated) processes the information request I and generates a document answer set A. Let |A| be the number of documents in this set. Further, let $|R_a|$ be the number of documents in the intersection of the sets R and A. … The recall and precision measures are defined as follows. Recall is the fraction of the relevant documents (the set R) which

(2004) analysis we know that 660 of the threads had a positive score, and 1004 had a score of zero. The model miscategorized 508 threads out of which 122 of them were miscategorized as having a mixed exchange (when their actual score was zero), and 386 of them were mistakenly assigned a score of zero. Consequently, the results for this model (with all the variables) are: precision = 69% and recall = 42%.

As is always the case in information retrieval tasks, there is a tradeoff that must be made between precision and recall (Buckland and Gey 1994). In our case, a low—but nonzero—recall rate is fine because discussion threads are not a scarce commodity. For example, Google Groups (http://groups.google.com/) has hundreds of millions of newsgroup messages indexed. We would, however, like a precision score that is as high as possible so that fewer threads without mixed exchanges are given further scrutiny.

We can refine this estimated discriminant function by first simplifying it. Our initial discriminant function—which includes all the variables—is based on the assumption that each of the variables which describe the size and shape of the thread trees, is independent from all of the other variables. This is clearly not the case. For example, the mean depth of the tree (meanD) is likely to be correlated with maximum depth (maxD) and the number of messages in the tree (M); and, such is the case: r(meanD,maxD) = 0.97; r(meanD,M) = 0.93. So, a refined discriminant function need not contain all of the variables.

Our simplified, discriminant function contains three almost independent variables: maxB, meanMp, and maxMp/M (i.e., the maximum number of messages posted by one person to a thread divided by the number of messages posted to a thread). This function of three parameters, g(maxMp/M, maxB, meanMp), accurately categorizes 68% of the threads with precision of 75% and recall of 29%.

When maxMp/M is extra large, mixed exchanges are unlikely. Intuitively one can understand the logic of this: when maxMp/M is large one participant has posted many more messages than the other participants in the thread. Thus, the thread is dominated by one voice and more likely to be monological rather than dialogical in nature.

When maxB is small, the score for the thread is more likely to be zero. This too is relatively intuitive: threads containing at least one message that received a lot of replies are more likely to incorporate many engaged discussants than threads containing only messages with few replies.

---

has been retrieved; i.e., Recall = $|R_a|$ / $|R|$. Precision is the fraction of the retrieved documents (the set A) which is relevant; i.e., Precision = $|R_a|$ / $|A|$' (Baeza-Yates and Ribeiro-Neto 1999: 75).

Finally, we are interested in threads in which meanMp is relatively large as this is an indication that several people are contributing substantially to the discussion.

5 Verification of the Model

A set of one hundred discussion threads was randomly selected from the same six newsgroups. To approximate the size distribution of our original collection of 1664 threads, 25 threads with 2 messages, 25 threads with 3 messages, 25 threads with 4 or 5 messages, and 25 threads with 6 or more messages were selected. Each of the authors of this chapter independently read and tagged the threads as either containing or not containing a mixed exchange. Our purpose was to verify our discriminant function on a manually tagged corpus of discussion threads.

It is noteworthy that even the three of us did not always agree on which threads did or did not contain a mixed exchange. All three of us were in agreement only 58% of the time. To test our model we used a majority vote: if two of us agreed that a mixed exchange had taken place in the thread, then the thread was marked as having a mixed exchange. This difficulty in manual tagging indicates a much deeper problem: can even a well educated, interested, and motivated person recognize a deliberative discussion when he or she sees one? While we would like the computer to recognize such an exchange, it is not clear what criteria people use to recognize such an exchange.

For discussion threads containing six or more messages the refined discriminant function (of only three variables) performed with 71% recall and 94% precision.

6 Conclusions, Discussion, and Future Work

The approach demonstrated in this work, to attempt to automatically identify mixed, possibly deliberative, exchange in discussion threads by examining the thread trees' structures—their topologies and morphologies—might strike some as quixotic. Or, perhaps, at least as quixotic as the enterprise of Chomskyan linguistics in its attempts to tell us something about language and the human mind by closely reading syntax trees. Nevertheless, even outside of Chomskyan linguistics, there is a long history of employing structural characteristics in order to define and distinguish social (e.g., social network analysis) and cultural or literary genres or discourses (Propp 1928).

For the purposes of this project we are tactical—not committed—structuralists. Our work essentially boils down to this: if one wants to find a mixed—potentially deliberative—exchange in a large set of Usenet newsgroup threads, look for those threads in which (a) no one person dominates the discussion, (b) everyone participating in the thread has posted at least a couple of messages, and (c) there is at least one message with multiple replies. This chapter details our search for this heuristic and presents the heuristic in a more precise form, as what one in the discipline of pattern classification might call a 'discriminant function' (Duda et al. 2001).

In future work, we plan to extend these simplest of models for identifying discussion threads containing mixed (political) exchanges to include a set of linguistic and social network criteria—criteria that we have already implemented in computational form in the Conversation Map system (Sack 2001). This, we hope, will bring us closer to achieving our long-term goal to implement a search engine that, when given a topic, will find likely threads of discussion where opposing opinions have been expressed.

Acknowledgements

This work has been supported by Award #0416353 ('An Interface and Search Engine for Deliberation') from the National Science Foundation, Directorate for Computer and Information Science and Engineering; and by National Science Foundation Award #IIS-0441999. Any opinions, findings and conclusions or recommendations expressed in this chapter are those of the authors and do not necessarily reflect those of the National Science Foundation.

References

Baeza-Yates, R. and B. Ribeiro-Neto. 1999. *Modern Information Retrieval*. New York: Addison-Wesley.

Bagdikian, B. 2004. *The New Media Monopoly*. Boston: Beacon Press.

Buckland, M. K. and F. Gey. 1994. The Relationship between Recall and Precision, *Journal of the American Society for Information Science* 45(1): 12-19.

Duda, R. O., P. E. Hart, and D. G. Stork. 2001. *Pattern Classification*. New York: Wiley.

Fishkin, J. S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven: Yale University Press.

Fishkin, J. S. 1992. Beyond Teledemocracy: 'America on the Line'. *The Responsive Community* 2(3): 13-19.

Granovetter, M. 1973. The Strength of Weak Ties. *American Journal of Sociology* 78: 1360-1380.

Kelly, J. 2004. *Community and Selectivity in Unstructured, Online, Political Discourse*. Paper presented at Internet Research 5.0 (Association of Internet Researchers), Sussex, September 2004.

Mill, J. S. 1993. *On Liberty*. New York: Bantam.

Mutz, D. C. 2002. Cross-cutting Social Networks: Testing Democratic Theory in Practice. *American Political Science Review* 96(1): 111-126.

Negroponte, N. 1995. *Being Digital*. New York: Knopf.

Propp, V. 1928(1968). *Morphology of the Folktale*. Austin, TX: University of Texas Press.

Putnam, R. D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Sack, W. 2001. Conversation Map: An Interface for Very Large-Scale Conversations. *Journal of Management Information Systems* 17(3): 73-92.

Sunstein, C. R. 2001. *Republic.com*. Princeton, NJ: Princeton University Press.