
Topicality, Time, and Sentiment in Online News Comments

Nicholas Diakopoulos

School of Communication and Information
Rutgers University
diakop@rutgers.edu

Mor Naaman

School of Communication and Information
Rutgers University
mor@rutgers.edu

Abstract

In this paper we examine the relationships between news comment topicality, temporality, sentiment, and quality in a dataset of 54,540 news comments. Initial observations indicate that comment sentiments, both positive and negative, can be useful indicators of discourse quality, and that aggregate temporal patterns in positive sentiment exist on comment threads.

Keywords

Commenting systems, online journalism, news discourse

ACM Classification Keywords

H.5.3 Group and Organization Interfaces:
Asynchronous interaction

Introduction

Recent work in the study of online commenting systems has shown that the quality of online commentary around news stories can affect the efficacy of journalistic information gathering efforts [3]. With this overarching motivation in mind, we undertook a quantitative study of 54,540 news comments collected from the website of the Californian newspaper Sacramento Bee, SacBee.com. Our study attempts to better understand features or factors of the discourse on the site, and could inform the moderation process for such an online newspaper. We primarily focus on features related to topicality, frequency, and other

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

temporal patterns, combined with sentiment analysis, as these were suggested as promising features by previous work [1-3, 5].

In particular, Burke and Kraut [1], show that politeness in online forums is often linked to the topic that those forums address. In prior work we present additional evidence for this effect based on interviews with moderators and editors at a newspaper [3]. Current work provides more detail in the domain of news commenting to assess how news topics affect the quality and tone of online discourse. For instance, what are the relationships between topicality and quality or sentiment? Sood and Churchill present several ideas with respect to using sentiment analysis to better inform moderation systems [5]. They raise questions for future work about modeling and mapping sentiment trajectories through comment threads. In this paper we begin to address such questions by asking what temporal features could potentially inform moderation, whether relating to frequency of commenting, the tempo of conversation, or the evolution of comment threads.

Data

We obtained a dataset of 54,540 comments made on the SacBee.com site spanning the month of August of 2009. The comments were extracted directly from the moderation system, and included fields indicating whether the comment had been deleted via moderation or hidden from the site because the user had been blocked by a moderator. We also recorded the article title and link that each comment referred to as well as the timestamp for the comment and the username of the person who wrote the comment. We enriched the comment dataset in two ways: (1) by collecting topic

categorizations for all articles in the dataset, and (2) by automatically tagging each comment with sentiment ratings based on the use of positive and negative words.

Since we were unable to collect topic categorization information for articles from the SacBee system we re-created this information, albeit with some loss of precision, using Amazon Mechanical Turk to label the topic of each article based on its title. Unfortunately, the original article text was not available due to the lack of access to the SacBee archive. The article title alone was thus available for our coders when assigning categories. We used the IPTC Media Topic News Codes¹ as a basis for topic labeling. This taxonomy consists of 17 top-level categories for news topics including, for example, Education, Human Interest, Labor, and Politics. The full list of topics is shown in Table 1 and descriptions for each topic can be found online. Each of the 3,755 articles was shown to two Mechanical Turk coders who were asked to mark all relevant topics for the article based on the article title. An article could therefore be marked as relating to several topics. Some articles had titles for which it was impossible to tell the topic (e.g. "Letters to the Editor") thus we included a category "Not Sure" for taggers, which was used for 6.0% of the articles. In order to maintain the quality of ratings we restricted our task to users of the service in the USA who had a greater than 95% task acceptance rate. This meant that the taggers could be expected to be fluent in English and be conscientious workers. We did an informal check of the collected tags and a majority seemed germane given the article title.

1. *Arts, Culture, and Entertainment*
2. *Conflicts, War, and Peace*
3. *Crime, Law, and Justice*
4. *Disaster and Accident*
5. *Economy, Business, and Finance*
6. *Education*
7. *Environment*
8. *Health*
9. *Human Interest*
10. *Labour*
11. *Lifestyle and Leisure*
12. *Politics*
13. *Religion and Belief*
14. *Science and Technology*
15. *Society*
16. *Sport*
17. *Weather*
18. *Not Sure*

Table 1. IPTC Media Topic News Codes

¹ <http://www.iptc.org/cms/site/index.html?channel=CH0103>

We also enriched our comment dataset with sentiment information by assigned a positivity and a negativity score to each comment ranging from zero to one. The algorithm we implemented was based on a simple classifier using a lexicon of words marked as positive or negative, including their negations [4]. For instance, the positivity score was computed by summing up the number of positive words in a comment and dividing by the total number of words in the comment to normalize the score to [0,1].

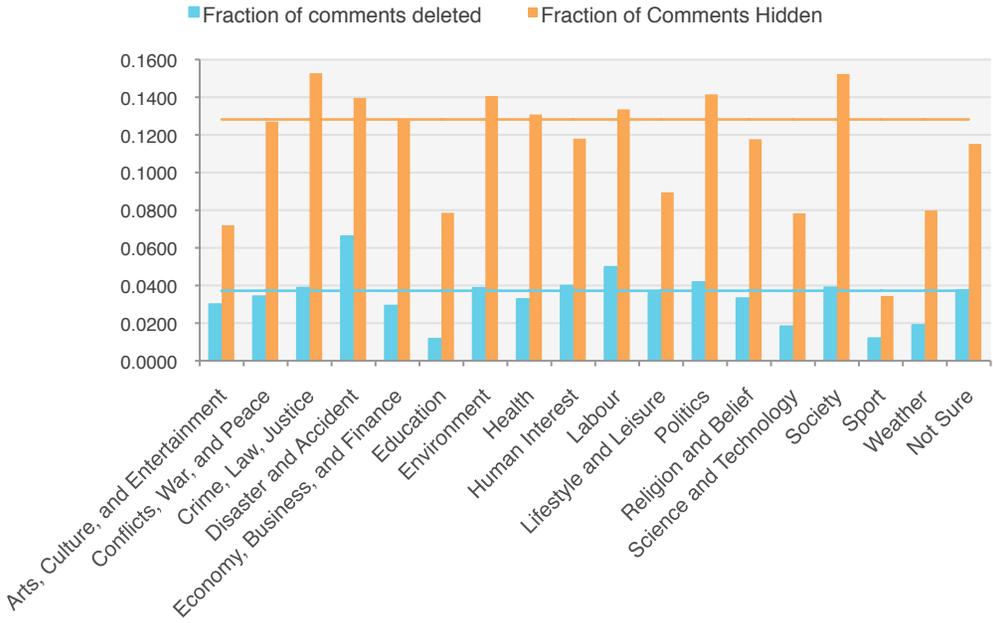


Figure 1. Fraction of comments deleted and hidden for comments on articles of each news topic. Horizontal lines show averages.

System Characterization

Here we describe the commenting system on SacBee.com as it relates to the analysis in the next section. Comments on SacBee.com are listed below each story in blocks of ten. Users can click through numbered pages for each subsequent block of ten. Comments are sorted in reverse chronological order by default (i.e. more recent comments first). In order to leave a comment a user must register a unique screen name with an email address. Comments are not threaded thus we are limited in our ability to understand how people are interacting with each other. The ability to comment is never switched off on articles.

Like many online communities SacBee.com explicitly states its commenting policy. The policy includes warnings against posting profanities, hate speech, and personal attacks, as well as off-topic, repeat, or spam comments. If a comment is deemed to be abusive by the moderators (professionals in the newsroom) then it is removed from the site and only visible within the moderation system marked as *deleted*. Users who repeatedly abuse the system can be blocked by moderators. Blocked users' comments remain visible to themselves but are *hidden* from other users (colloquially referred to as a "bozo filter"). In our analysis we use deleted and hidden comments as proxies for quality as they indicate the degree to which discourse does not coincide with the commenting policy of the site.

Analysis

We first describe the dataset and sentiment analysis, then move on to topicality, and finally temporal analysis.

Description

Of the 54,540 comments collected in our dataset, 2,124 (3.9%) were marked as *deleted*. An additional 7,225 (13.3%) comments were marked as *hidden*. Perhaps unsurprisingly the deleted comments in our dataset were found to be more negative ($M=0.041$ $SD=0.051$) than non-deleted comments ($M=0.033$ $SD=0.042$). The difference is statistically significant ($t(54,538)=8.45$, $p < .0001$). Moreover, deleted comments had a higher variance in negativity than non-deleted comments indicating that deleted comments were more varied in their range of negativity. There was no such difference found between deleted and non-deleted comments on their positivity scores. Overall comments were found to be more positive (0.046) than negative (0.033). In other words, across all 54,540 comments, people were more apt to use the positively associated words in our lexicon than the negatively associated words.

Topicality, Quality, and Sentiment

Figure 1 shows the fraction of comments deleted and hidden across each of the news topic categories. We can see that categories such as "Disaster and Accident"

and "Labour" have higher than average proportions of deleted comments. Similarly we can also see that categories such as "Crime, Law, Justice", "Disaster and Accident", "Environment", "Politics", and "Society" have a higher proportion of hidden comments indicating that people who were already blocked by moderators were attracted to those topics. If we look at topicality in relationship with sentiment we find that topics that are more positive tend to have less comments per article ($r = -0.548$, $p = .019$) and that topics that are more negative tend to have a higher fraction of deleted comments ($r = 0.698$, $p < .01$).

Temporal Analysis

FREQUENCY AND SENTIMENT

In prior work [3] we suggest that frequency of commenting is a useful measure of the quality of a user's comments. In particular we found that users who commented less than 10 times in a month were blocked less often than people who commented more than 10 times. Here we add sentiment information to the analysis of commenting frequency. We found a significant positive correlation between negativity and frequency of commenting ($r = .027$, $p = .014$). Even if we hold out the more negatively shaded deleted comments this correlation remains ($r = .025$, $p = .024$) indicating that even for only messages that passed through moderation, people who comment more are still more negative in their comments.

CONVERSATIONAL TEMPO

We calculated the tempo of a comment by measuring the amount of time (in minutes) between the current comment and the last comment on any given thread. Figure 2 shows the distribution of the fraction of deleted comments across comment tempos ranging

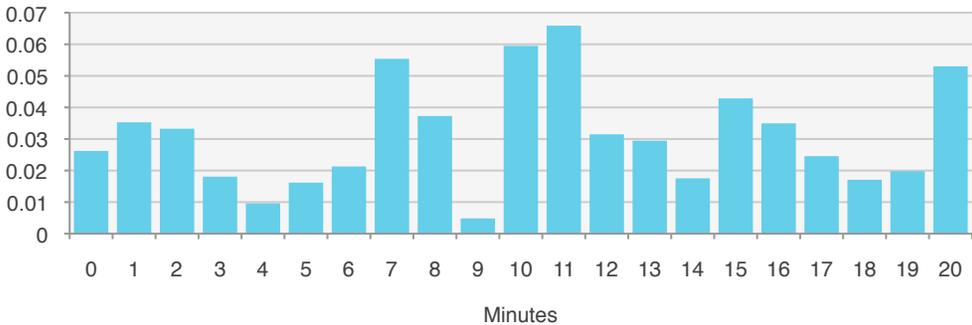


Figure 2. The fraction of comments deleted across minutes since the last comment.

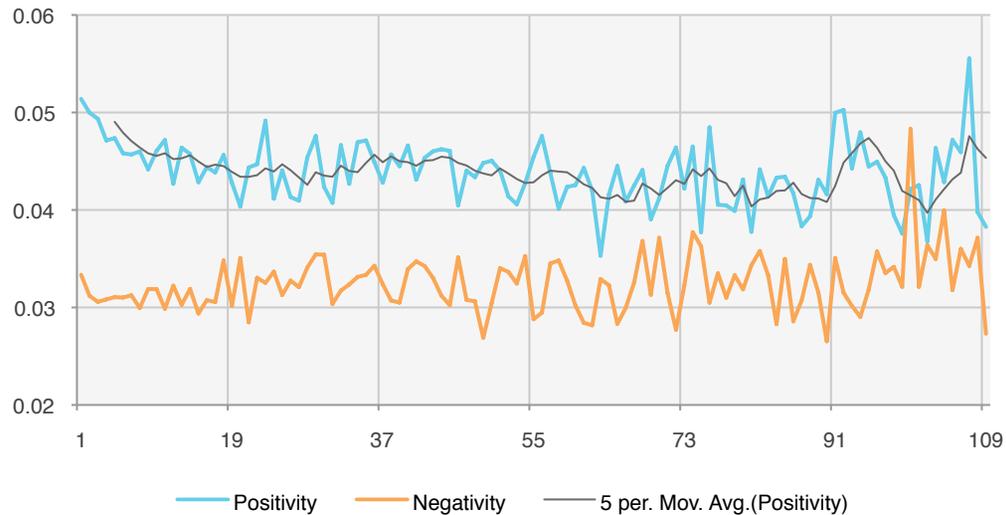


Figure 3. Aggregated sentiment per comment rank. A smoothed trend line is shown in gray for the positive signal.

from zero minutes (the two comments were made practically contemporaneously) up to twenty minutes. By inspecting Figure 2 we can see that the fraction of deleted comments for fast-paced comments (less than six minutes) is below average, whereas for slower tempo comments the fraction of deleted comments is above average. We compared the number of deleted comments for fast tempo (≤ 6 minutes) and slower tempo (> 6 minutes) comments using a chi-square test and found a significant difference (chi-square = 47.02, dof = 1, $p = 0$) from the expected distribution. This would seem to indicate that during the “frenzy” of commenting on articles, when there is a fast tempo, comments are less likely to get deleted than if the tempo is (even a bit) slower. The exact reason for this difference is unclear; it could be due to a paucity of moderation attention (from users flagging or the

newsroom moderating) during the high-volume commenting frenzy (which accounts for 32.7% of comments), or it could be due to characteristics of low-quality commenters taking longer to craft their troll responses.

SENTIMENT OVER TIME

In order to better understand how comment sentiment changes over time, we aggregated sentiment by comment rank, such that, for instance, the positivity of each of the 10th comments on a story were averaged together. This data (which excludes comments that were hidden or deleted) is plotted in Figure 3. There are a few observations we can make from Figure 3. For one, we can again see that across (almost) all ranks comments are more positive than they are negative. Secondly, we can see a somewhat steep decline in positivity in the first 10-20 comment ranks, whereas the negativity remains more or less constant across ranks (with some variance). If we correlate the aggregate positivity to rank up to rank 10 there is a strong negative correlation ($r = -.903$, $p < .001$). There is no such correlation for negativity. This result indicates a propensity for comments early on in a thread to be more positive than those that occur later. While future work remains to be done to assess why this is, one possibility is that early on in the comments people are responding directly to the story, whereas later on they are interacting more with each other.

In higher ranks, the positivity signal appears to enter a periodic state (shown via the moving average in gray), which is most prominent between ranks 91 to 109. The number of stories with 109 comments or less was 64. We ran an autocorrelation to assess the degree of periodicity and found the largest autocorrelation

corresponds to a period of rank 18. The Box-Ljung statistic indicates the autocorrelation at this period was significant ($p < .017$). There was no significant autocorrelation for the negativity scores. Future work remains to be done to better understand the nature of this periodicity. One interpretation is that because of the reverse chronological ordering of the comments and their chunking into groups of 10, the sentiment of writers is in response to the most recent 10 comments on the page. There may be a natural tendency to make your comment more or less positive based on your most recent exposure to other comments (a form of anchoring), as well as the overall degree of positivity (causing the signal to bounce).

Discussion and Conclusions

In this paper we have begun shedding light on the relationships between news comment topicality, temporality, sentiment, and quality. There is much future work to do in this area, but we believe our initial results include several interesting observations. While there were no substantial surprises in terms of the topics that aroused more deleted comments, the correlation between negativity and the fraction of deleted comments across topics suggests that measures of comment sentiment such as negativity could inform moderators about topics (or perhaps even threads) that need additional moderation attention. Beyond the coarse analysis we did here, in future work we would like to use the full text of articles to build topic models that could then be related to comment attributes such as sentiment or deleted status.

From our temporal analysis we observed that frequency of commenting, while connected to the propensity for an individual to be blocked on the site, is also

correlated with the negativity of a user's comments. This is another signal that negativity, even using our relatively simplistic measure, is a useful indicator of quality. Our future work also aims to assess the veracity of our current observations about conversational tempo (e.g. that slow conversation is more likely to have deleted comments) and about comment sentiment over time (e.g. an initial steep decline in positivity followed by a periodic swing in positivity). Ideally we would like to repeat these analyses with another dataset to see if our observations are reliable, as well as to develop theoretical explanations of our findings.

Acknowledgements

We wish to thank the Sacramento Bee for their cooperation with our study. The first author would also like to acknowledge the AAAS as a member of the Mass Media Fellowship program, and the CRA as part of a Computing Innovation Fellowship (CIF-197).

References

1. Burke, M. and Kraut, R., Mind Your Ps and Qs: The Impact of Politeness and Rudeness in Online Communities. *Computer Supported Cooperative Work (CSCW)*, (2008).
2. Chmiel, A., Sienkiewicz, J., Paltoglou, G., Buckley, K., Thelwall, M. and Holyst, J. Negative Emotions Boost Users Activity at BBC Forum. 2010. <http://arxiv.org/abs/1011.5459>
3. Diakopoulos, N. and Naaman, M., Towards Quality Discourse in Online News Comments. *Proc. CSCW*, (2011).
4. Riloff, E. and Wiebe, J., Learning Extraction Patterns for Subjective Expressions. *Empirical Methods in Natural Language Processing (EMNLP)*, (2003).
5. Sood, S.O. and Churchill, E.F., Anger Management: Using Sentiment Analysis to Manage Online Communities. *Grace Hopper Celebration*, (2010).