Sentiments in Wikipedia Articles for Deletion Discussions

Lu Xiao(✉) and Niraj Sitaula(✉)

Syracuse University, Syracuse, NY 13244, USA
{lxiao04,nsitaula}@syr.edu

Abstract. Wikipedia provides a discussion forum, namely, Article for Deletion forum, for people to deliberate about whether or not an article should be deleted from the site. In this paper, we present interesting correlation between outcomes of the discussion and number of sentiments in the comments with different intensity. We performed sentiment analysis on 37,761 AfD discussions with 156,415 top-level comments and explored relationship between outcomes of the discussion and sentiments in the comments. Our preliminary work suggests: discussion that have keep or other outcomes have more than expected positive sentiment, whereas discussions that have delete outcomes have more than expected negative and neutral sentiment. This result shows that there tends to be positive sentiment in the comment when Wikipedia users suggest not to delete the article. This observation of differences in sentiments also encourages to further study influence of sentiments in decision making or outcome of the discussions. Our future analysis will include threaded comments, and examine the relationship between a discussion's sentiment and its other properties such as topic of the article and the characteristics of the participating users.

Keywords: Wikipedia · Sentiment analysis · Online discussion

1 Introduction

As a decentralized peer production system, Wikipedia uses various strategies to monitor and control the quality of the articles. One of which is its mechanisms for deleting articles from Wikipedia. According to the Deletion Discussion mechanism, an article proposed to be deleted may undergo the community discussion before a decision is made (e.g., to delete the article or to keep it). The article tagged for Deletion Discussion is called "Article for Deletion" (hereafter: AfD). The user who nominates an article for Deletion Discussion needs to provide a rationale to justify his nomination (i.e., the statement in the figure "Appears to not meet WP:GNG for reliable secondary sources to confirm notability"). In an AfD discussion, participants offer their opinions on what to do with this article as top-level comments. They may also respond to someone else' opinions by embedding their comments below the corresponding top-level comments. At the end of the discussion (e.g., after a week or two), a user who did not participate in this AfD discussion will review the discussion content and make the final decision. Wikipedia policy requires that the AfD decision should be based on the rationales provided by the participants.

G. Chowdhury et al. (Eds.): iConference 2018, LNCS 10766, pp. 81–86, 2018.
https://doi.org/10.1007/978-3-319-78105-1_10

While researchers have explored the sentiment aspect of Wikipedia talk pages' discussions (e.g., [2, 6, 10]), the sentiments in Wikipedia AfD discussions have not been looked at, not to say the role of sentiments in the decision-making process about the article.

Our work makes initial step to fill this gap and understand the sentiment aspect of AfD discussions along with its role in the discussion process and outcomes. We report here our current status of the research activities. Specifically, we examined the sentiment of the top-level comments in about 37,761 AfD discussions and their percentages in the AfD discussions of different outcomes.

2 Related Work

"On Wikipedia, notability is a test used by editors to decide whether a given topic warrants its own article" (https://en.wikipedia.org/wiki/Wikipedia:Notability). Lam and Riedl [4] found that the most common reason for deleting an article is its lack of notability. A later content analysis study (N = 229 AfD discussions) also found that notability is the most commonly used rationale for keeping or deleting an article [11]. These studies suggest that the discussions are logical and follow Wikipedia's policies in general.

On the other hand, it has been shown that besides notability of the article an AfD decision can be affected by various factors. For example, opinions offered at the early stage of an AfD discussion influence the later opinions [9]. Groups formed naturally and groups with a moderate diversity of newcomer and expert participants make better decision [5]. Additionally, Xiao and Askin [11] found an AfD discussion that has votes other than keep or delete (e.g., merge) is more likely to be suggested for actions other than delete. The authors also found that certain categories of an article correlates with the likelihood of the article to be deleted. For example, articles about people, for-profit organizations, or definitions are slightly more likely to be deleted than expected, while articles about locations or events are more likely to be kept than expected, and articles about nonprofit organizations and media are more likely to be suggested for other options (e.g., merge, redirect, etc.) than expected.

However, the sentiment in AfD discussions, and whether and how it affects the final decision has not been explored. Nonetheless, while all the AfD discussion data are available online, these earlier studies only had small samples, which made it challenging to generalize their findings to the AfD discussion study. Hence, our work explored larger number of AfD discussion and presents results from analyzing sentiments in the comments in the discussions.

3 Our AfD Corpus for Sentiment Analysis

Wikipedia manages the AfDs based on the date they are proposed for Deletion Discussion. The content of each proposed date is publicly accessible through URLs. An AfD discussion consists of several parts: article title, nomination reason, participants' votes, and outcome. A nomination reason explains why the article is proposed for

deletion. A typical participant's vote includes the participant's opinion on the outcome of the article such as keep or delete, and the rationale to justify his/her vote. An outcome includes the final decision regarding the article and the rationale of this decision. Each of these three parts ends with a user signature. A user's signature consists of a username, and a timestamp (date and time).

Our PHP script visited the web sites through the URLs of the proposed dates from May 15, 2013, to May 15, 2015 (i.e., 720 dates) and stored the HTML content. These HTML pages become our raw data. We then applied regular expressions and filtered the noise from the data, e.g., the missing HTML tags, the mis-formatted user signature, etc. There are three possible views of an AfD vote or its outcome according to our database design: keep, delete, and other (Table 1).

Table 1. The mapping between a comment's view in AfD to the database

Participant's vote in following choices	Vote stored a
Strong delete, speedy delete, delete, weak delete	Delete
Weak keep, keep	Keep
Note, comment, question, speedy close, closing administrator, relisted, text, reviews, speedy decline, withdraw, userfy, move and dab, move, oppose, merge, redirect, redirect and merge	Other

Of the 183,007 top-level comments in the corpus, we observed that there comments which were not associated with any particular vote. In other words, they were not the justifications of participants' opinions. Examples of such comments are "Please add new comments below this notice. Thanks", "Please add new comments below this notice. Thanks, Mz7 (talk) 19:27, 15 May 2014 (UTC)", "Note: This debate has been included in the list of Football-related deletion discussions. Ascii002Talk Contribs GuestBook 12:24, 23 August 2014 (UTC)", and "Note: This debate has been included in the list of……", in comments text. There were also comments with only text "–" followed by username, for e.g. "Delete–Rpclod (talk) 02:42, 16 July 2014 (UTC)".

After removing these comments, we had 156,415 comments and we noticed that the median of the comments was found to be 2 and average number of sentences in each comment was found to be close to 2.62. In addition, distribution of outcomes in 37,761 AfD discussions consisted 21,589 as Delete, 8,196 as Other, and 7,976 as Keep.

4 Sentiment Analysis of Our AfD Corpus

Given the large number of sentences in this corpus, it is difficult and time consuming to manually annotate all the sentences for their sentiments. Therefore, we used available classifier to classify the polarity and intensity of a comment's sentiment.

4.1 Sentiment Classification

We preprocessed the data using nltk [1] as follows:

- Unwanted hyperlinks and html tags embedded in the comments were cleaned.
- Unwanted characters were also removed for e.g. \\r\\n, -, *comma (,), 's*, and others.
- Capitalization and punctuations were retained as it is.

To classify the sentiment of a sentence, we used VADER which stands for Valence Aware Dictionary for sentiment Reasoning [3]. Built with corpus like movie reviews, technical product reviews, opinion news article, VADER has been shown to perform better than many other methods in predicting sentiments in Social Media Text and accounts for different intensity of a sentiment [3].

4.2 Sentiment Analysis Results

We used the VADER library from nltk [1] library for the sentiment classifications. First we used VADER to find the sentiment of a comment, and then based on the sentiments of all the top-level comments of a discussion, we calculate the sentiment of the discussion by majority of the sentiments present in the discussion. When two or more sentiments with majority were found in discussion, to break the ties 100 random selection among the sentiments were taken and the one that occurred maximum number of times was picked as the sentiment of the discussion.

The result of sentiment labeling is shown in Table 2 along with the result from the classification. As shown in the table, majority of the discussions were neutral with respect to the sentiment of their content. This result is consistent with prior studies that AfD discussions are in general rational [11].

Table 2. Sentiment label distribution

Label	Comments count	% of comments	Discussion count	% of discussion
Positive	29886	19.11%	5697	15.09%
Neutral	111754	71.45%	30041	79.55%
Negative	14775	9.45%	2023	5.36%

For each discussion labeled with majority of sentiment, we observe the count of different votes i.e. Delete, Keep, and Other. This data is summarized in Table 3. Note that neutral means the emotional tone is neutral, not that the opinion on whether not to delete the article is neutral. We conducted a chi-square test based on this result and obtained a p-value < 0.001.

Table 3. Sentiment labels on discussion outcomes

Label	Delete discussion count	Other discussion count	Keep discussion count
Negative	1488	309	226
Neutral	17641	6389	6011
Positive	2460	1498	1739

The test shows the following results. Discussions that had delete outcome are more than expected to have negative and neutral sentiment and less than expected to have positive sentiment. Discussions that had keep or other outcome are more than expected to have positive sentiment and less than expected to have negative and neutral sentiment.

5 Discussion and Conclusion

When the inclusion of an article in Wikipedia article is questioned, Wikipedia may start an open discussion that encourages Wikipedia users to offer their opinions along with their reasoning. Such a discussion is called an Article for Deletion (AfD) discussion. While researchers have explored various aspects of AfD discussions (e.g., [5, 11]), the sentiments in these discussions have not been explored.

Addressing this gap, we seek to understand the sentiment aspect of AfD discussions and its role in the discussion process and outcome. Our analysis of 156,415 top-level comments in 37,761 discussions suggest that there is a correlation between an AfD discussion's sentiment and its outcome. The results showed that discussions with outcomes of keep or other have more than expected positive sentiment, whereas more than expected negative and neutral sentiment were found in discussions with delete as outcomes. While the correlation of positive and negative sentiment with the keep and delete discussion outcome is expected, we find it particularly interesting the correlation of neutral sentiment with the discussion outcome. The fact that delete discussions had more than expected neutral discussions indicates that even when participants argue for deleting the articles their comments were not that emotional hence more than expected neutral discussions were observed. The fact that keep or other discussions that had less than expected neutral discussions indicates that when users suggest not to delete the article there tends to be a positive sentiment than just a neutral statement. Prior studies have shown that in AfD discussions Wikipedia users tend to offer advice on how to improve the articles and be inclusive on controversial articles, as opposed to just arguing to delete or keep the article [11]. Our analysis result is consistent with this previous finding, as it indicates a constructive and inclusive discussion context.

In addition, our current analysis has only considered top-level comments. We next will include the threaded comments in our analysis, and examine the relationship between a discussion's sentiment and its other properties such as the topic of the article and the characteristics of the participating users. Another limitation of our study is the lack of performance measure for VADER for AfD discussion data. According to [3], VADER has F-1 measure of .63 for classifying the sentiment of an Amazon product review (performance of a human annotator is .85). While AfD comments are somewhat comparable with a product review, it would have been better if the performance of VADER was evaluated with a small sample of AfD discussion data before applying it to all the discussion comments. On the other hand, VADER is reported to be among the best sentiment prediction tool available for classifying social media texts and online review comments [7].

Earlier work in policy making has shown that the community sentiments have great influence in decision making processes of law makers [8]. It has also been shown that

sentiment as a feature in a machine learning model has improved the prediction of decision making in process of loan granting [12]. As we have observed relationships between discussion's sentiment and its decision outcome, it will be interesting to explore the predictive power of an AfD discussion's sentiment on the outcome of the discussion.

References

1. Bird, S., Klein, E., Loper, E.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. OReilly Media Inc., Sebastopol (2009)
2. Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., Potts, C.: A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078 (2013)
3. Hutto, C.J., Gilbert, E.E.: VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Eighth International Conference on Weblogs and Social Media (ICWSM-2014), Ann Arbor, MI, June 2014
4. Lam, S.T.K., Riedl, J.: Is Wikipedia growing a longer tail? In: Proceedings of the ACM 2009 International Conference on Supporting Group Work, pp. 105–114 (2009)
5. Lam, S.K., Karim, J., Riedl, J.: The effects of group composition on decision quality in a social production community. In: Proceedings of the 16th ACM International Conference on Supporting Group Work, pp. 55–64 (2010)
6. Laniado, D., Kaltenbrunner, A., Castillo, C., Morell, M.F.: Emotions and dialogue in a peer-production community: the case of Wikipedia. In: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration, p. 9. ACM (2012)
7. Ribeiro, F.N., Araújo, M., Gonçalves, P., Gonçalves, M.A., Benevenuto, F.: SentiBench-a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci. **5** (1), 1–29 (2016)
8. Sigillo, A.E., Sicafuse, L.L.: The influence of media and community sentiment on policy decision-making. In: Miller, M., Blumenthal, J., Chamberlain, J. (eds.) Handbook of Community Sentiment, pp. 29–42. Springer, New York (2015). https://doi.org/10.1007/978-1-4939-1899-7_2
9. Taraborelli, D., Ciampaglia, G.L.: Beyond notability. Collective deliberation on content inclusion in Wikipedia. In: Proceedings of the 4th IEEE International Conference on Self-Adaptive and Self-Organizing Systems (SASO), pp. 122–125 (2010)
10. Wang, L., Cardie, C.: A piece of my mind: a sentiment analysis approach for online dispute detection. arXiv preprint arXiv:1606.05704 (2016)
11. Xiao, L., Askin, N.: What influences online deliberation? A Wikipedia study. J. Am. Soc. Inf. Sci. Technol. **65**(5), 898–910 (2014)
12. Zhang, D., Xu, W., Zhu, Y., Zhang, X.: Can sentiment analysis help mimic decision-making process of loan granting? A novel credit risk evaluation approach using GMKL model. In: 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 949–958. IEEE, January 2015