Determining Controversy in Wikipedia through Twitter Data

Muhammad Zeeshan Jhandir[1], Tenvir Ali[1], Byung-Won On[2], Ingyu Lee[3], Gyu Sang Choi[*1]
[1]Department of Information and Communication Engineering Yeungnam University, Gyeongbuk, Republic of Korea;
[2]Department of Statistics and Computer Science, Kunsan National University, Republic of Korea;
[3]Sorrel College of Business, Troy University, Troy, AL 36082

Summary

In day to day life, people discuss things sometimes they agree and sometimes disagree. The later in many cases lead to conflicts and with the passage of time give rise to controversies among people, communities, nations and even reaches to the global level. Wikipedia for having "*anyone can edit*" policy, has to offer much for raising conflicts and ultimately the controversies. Previously used techniques for Wikipedia controversy detection, either lacked human annotated data set or used a negligible annotated dataset, compared to the size of Wikipedia. In this paper, we have proposed a novel method of Wikipedia controversy detection that can evaluate and identify the controversies based on twitter sentiment data about the same topic as that of Wikipedia article. To the best of our knowledge, this is a pioneer attempt to use the current version of Wikipedia page, instead of all revision histories. In addition to this, we have resolved the problem of human annotated data set limitations by employing the various social media data (currently twitter only) to automatically get annotation of human based opinion about a specific edit pair.

Introduction

Most conflicts lead to controversy, as the time passes on. If an immediate attention is not given to rising controversies in the society, they may lead to unrest, hatred, and the eruption of violent activities among the conflicting groups. Apart from this, controversial contents, talks and literature may confuse the readers to reach the true and authentic facts about certain things. Controversy identification is very important especially in the current scenario of electronic sources for information (web-search engines, web pages, wikis, social networking sites etc.). Analyzing the web contents and especially Wikipedia is very important as it is being the ultimate source of information.

Wikipedia for having "*anyone can edit*" policy, has to offer much for raising conflicts and ultimately the controversies. Researcher community has diverted towards detection and evaluation of Wikipedia controversial contents since over the last decade.

Machine learning methods were introduced in [1], using parameters like the number of revisions, the number of unique authors page length, etc. Mutual reinforcement principle was introduced in [2], stating that conflicting content is more controversial if page's controversy is low. Bipolarities in edit graphs was discovered by [3]. Revert statistics were used in [4]. Controversy by user feedback was measured by [5], while fine-grained controversy detection is carried in [6].

All the above methods either used statistical data of historical revisions such as deleting, inserting or modifying the tokens in Wikipedia articles. Although the revision data is a good source of getting the conflict among editors over the specific text of the article, however, the common reader always gets interacted with current version or revision of the article. Hence he is likely to be interested in knowing what controversy exists in the article in the contents he is viewing currently rather than being interested in knowing the historical edit wars among the authors or conflicts of interest among them.

We have proposed a twitter based model for evaluation of identifying controversies in Wikipedia articles. The details would come in coming sections. Our contributions in this paper are given below.

1. We proposed a novel method of Wikipedia controversy detection by using the sentiment data from social media such as twitter for deciding controversy of Wikipedia contents at sentence level, Figure 1 gives an overview of the whole system and Figure 2 gives overview of the process flow of Wikipedia controversy detection, based on opinionated (sentiment scored) text from twitter.
2. In addition to this, we have resolved the problem of human annotated data set limitations by employing the various social media data to automatically get annotation of human based opinion about a specific Wikipedia content.
3. Previously people have used historical data of Wikipedia article revisions, we have implied our model on the current versions of the pages.
4. The method is generic in nature and can be used on any topic from any category of the Wikipedia.

The rest of organization of the paper is, as section 2 gives detail of the proposed method, while section 3 is about results and evaluations, followed by section 4 with conclusive remarks and future work.

1. Methodology

We proposed a novel method for detection of controversy in Wikipedia by using human opinionated text collected from twitter. For this purpose, we have to extract data from twitter and Wikipedia. On the first step, we randomly selected 10 controversial topics from Wikipedia list of controversial issues* such as abortion, same-sex marriage, atheism etc. and 10 topics from the Wikipedia featured articles list† such as bacteria, virus, cabbage, etc. Later we used these topics to extract the tweets from the twitter by using Search and Stream APIs. For querying the twitter we used all possible alternates of each topic since in Wikipedia. Mostly, the similar or look alike topics are redirected to the same page, for example, same-sex marriage, gay marriage, marriage equality etc. are redirected to same single page "Same-sex

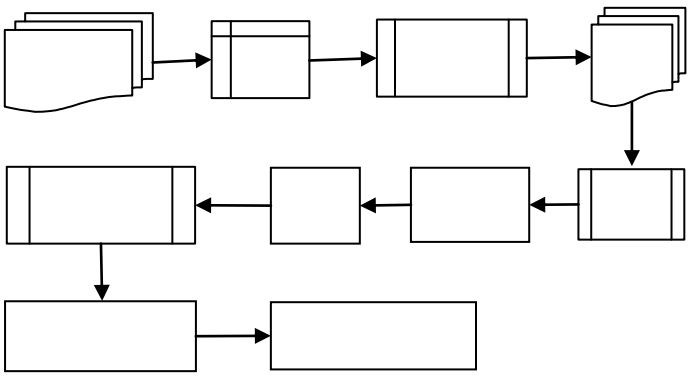


Figure 1: *Process flow of twitter based Wikipedia articles controversy evaluation*

* https://en.wikipedia.org/wiki/Wikipedia:List_of_controversial_issues accessed as on January 01, 2017

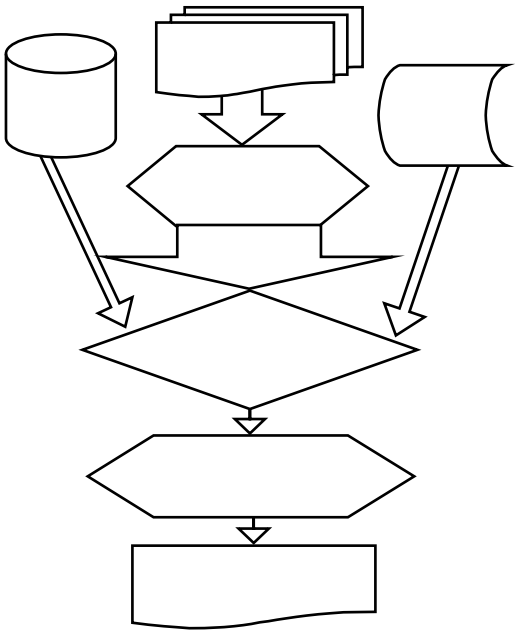
† https://en.wikipedia.org/wiki/Wikipedia:Featured_articles accessed as on January 01, 2017

Sr#	Tweets	Score
1	I'm guessing that the opposition is more likely about "harm" to children in gay marriage, not the sole purpose of breeding.	-0.25
2	You need to look up what marriage is, because gay marriage is not redefining it, considering the history	+0.3
3	absolutely. Same with feminist movement, gay marriage, etc. really just to undermine the institutions that exist	0.2
4	if those against allowing gay marriage base their position on breeding, then let us also disallow marriage between heterosexual nonbreeders	-0.7
5	extreme conservatives love to tout out the slippery slope that gay marriage would bring were it put in place. the one we're on is far scarier	+0.2

Table 1: *Example tweets sentences with positive sentiment score (GREEN), negative sentiment score (RED) and neutral sentiment score (BLUE)*

Evaluation Metrics	F1 Score	Precision	Recall	Accuracy
Obtained results	0.57	0.69	0.57	0.78

Table 2: *Evaluation results for the performance of proposed system*



marriage".

After obtaining tweets data about Wikipedia topics we carried out the cleaning and filtration process. We cleaned tweets by removing URLs, all non-English tweets and also slangs like hmmmm, hahaha, Zzzzzz, etc. Since it is useless to obtain any sentiment from single or two words reply like ok, holy shit, etc., hence we removed any tweets of unigrams or bigrams. Cleaned tweets were tokenized, normalized and tagged for part of speech (POS-tagging). Then the sentiment score of each tweet was calculated based on sysnet scores obtained from the SentiWordNet corpus. In Table 1 sample tweets about gay-marriages along with final scores are shown.

As mentioned earlier, to find the controversy of Wikipedia articles we have focused the current version of the article instead of the traditional approach in which whole historical versions of the article are taken into account. Our approach is based on the intuitions.

1. An ordinary user, when visits an article is usually presented with a current version of the article and mostly he is not interested to know what happened in past to the article.
2. By using the historical revisions of the article we at maximum can establish that which portion of the article is the focus of conflict but we cannot tell whether it is showing any controversy or not, until unless some human opinion is taken. The steps involved are,

1. After getting a Wikipedia article, we split it into sentences and put single Wikipedia sentence (WPS) on each line, as in Figure 4.
2. The second step is to use the individual sentence as a query to the human opinion sentences, HOS (tweet corpus) and find out semantically similar sentences to the queried sentences. For semantic similarity measure, we adopted the method presented by [12].
3. After having the similar sentences we calculate their average sentiment score and assign that score to the WPS. For example for WPS1 we got 5 HOS out of 100 HOSs with sentiment scores {0.32,-0.2,-0.21,-0.25, 0.78} then the average score will be {0.088} which means the WPS sentence is neutral with score 0.088. Likewise we assign scores to each WPS sentences and finally, we average the score of all sentences and thus get the final score of Wikipedia article itself. With either signed scores positive or negative we consider the article as controversial and non-controversial if the score is unsigned. In simple words the voting is influenced by sentences with a high score as compared to sentences with low scores, i.e. if there are 2 tweets with a combined score of -0.9 as compared to 5 sentences with combined score of +0.7, the net score will determine the subjectivity score of the Wikipedia sentence.

The model was trained on a tweets and Wikipedia articles sentences data for all the selected topics. We split the data by 80% to 20% for training and testing respectively, giving equal proportion to the two classes of Wikipedia topics (articles sentences and related tweets).

2. Results and Discussion

The results of the system are quite promising. Table 2 shows the overall evaluation results of the proposed model for various evaluation matrices, such as precision, recall, accuracy and F_1 score.

Since the model is unique in its nature, hence we could not get any base model for the comparison purpose. All other models for Wikipedia controversy detection use the historical revision data of each article in order to measure controversy, whereas we used the current version of Wikipedia pages that a common reader interacts with.

3. Conclusion and Future Work

We proposed a Wikipedia controversy detection method that is capable of producing self-annotations, the model is the first ever attempt of identifying controversial contents in Wikipedia pages. It is independent of historical revision of the article. Although the data set size used is very small and results of evaluation matrices may reduce by induction of larger volume of data, but still being a pioneering effort towards the controversy detection on the current version of the page the results are still good enough. A further, the technique is useful for situations where human annotated data is absent. To induce more accuracy in the annotation we will add data from the other social media as well, such as Facebook, Debatepedia, Pinterest, and others. The comments and post related to the Wikipedia topic will be searched and processed. Further, we will increase the data size in future experiments.

References

[1]. Kittur A., Suh B., Pendleton B., Chi E., He says, she says: conflict and coordination in Wikipedia, Proceedings of the SIGCHI conference on Human factors in computing systems San Jose, California, USA (pp., 453 – 462), 2007.

[2]. Vuong, B.-Q., Lim, E.-P., Sun, A., Le, M.-T., & Lauw, H. W. On ranking controversies in Wikipedia: models and evaluation. WSDM '08: Proceedings of the international conference on Web search and web data mining. New York, NY, USA: ACM, (pp. 171-182) 2008.

[3]. Yasseri, T., Sumi, R., Rung, A., Kornai, A. and Kertész, J. 'Dynamics of conflicts in Wikipedia', PLoS ONE, 7(6), p. e38869. 2012.

[4]. Rad, H. S., & Barbosa, D. Identifying Controversial Wikipedia Articles Using Editor Collaboration Networks. *ACM TIST*, *6*(1), 5:1-5:24, 2015.

[5]. Jankowski-Lorek, M., Nielek, R., Wierzbicki, A., & Zielinski, K. (2014). Predicting Controversy of Wikipedia Articles Using the Article Feedback Tool. In S. Yang, K. Lerman, J. She, & M. Atzmueller (Ed.), *SocialCom* (pp. 22:1-22:7). ACM

[6]. Bykau, S., Korn, F., Srivastava, D., & Velegrakis, Y. (2015). Fine-grained controversy detection in Wikipedia. In J. Gehrke, W. Lehner, K. Shim, S. K. Cha, & G. M. Lohman (Ed.), ICDE (pp. 1573-1584). IEEE.

[7]. M. T. Pilehvar, D. Jurgens, and R. Navigli. Align, Disambiguate and Walk: A Unified Approach for Measuring Semantic Similarity. Proc. of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, August 4-9, 2013, pp. 1341-1351.