ControCurator: Understanding Controversy Using Collective Intelligence

B. Timmermans, L. Aroyo and T. Kuhn, Vrije Universiteit Amsterdam
K. Beelen, E. Kanoulas and B. van de Velde, University of Amsterdam
G. van Eerten, CrowdyNews

1. INTRODUCTION

There are many issues in the world that people do not agree on, such as Global Warming [Cook et al. 2013], Anti-Vaccination [Kata 2010] and Gun Control [Spitzer 2015]. Having opposing opinions on such topics can lead to heated discussions, making them appear controversial. Such opinions are often expressed through news articles and social media. There are increasing calls for methods to detect and monitor these online discussions on different topics. Existing methods focus on using sentiment analysis and Wikipedia for identifying controversy [Dori-Hacohen and Allan 2015]. The problem with this is that it relies on a well structured and existing debate, which may not always be the case. Take for instance news reporting during large disasters, in which case the structure of a discussion is not yet clear and may change rapidly. Adding to this is that there is currently no agreed upon definition as to what exactly defines controversy. It is only agreed that controversy arises when there is a large debate by people with opposing viewpoints, but we do not yet understand which are the characteristic aspects and how they can be measured. In this paper we use the collective intelligence of the crowd in order to gain a better understanding of controversy by evaluating the aspects that have impact on it.

2. RELATED WORK

Controversies exist as public debates as they often stem from scientific and political debates. Furthermore, they touch on issues that concern large segments of society, and thus involve many different actors such as the media, politicians and scientists. In medicine, debates on stem-cell research [Clarke 1990] and abortion [Kelley et al. 1993] are examples in case. Another recurrent characteristic of controversy is the unsolvable nature, and henceforth, their persistence over time. This is described by [Dalgalarrondo and Urfalino 2002] as "a prolonged public disagreement in which a series of opposed arguments are exchanged." The dispute, therefore, gives rise to articulate pro and con sides who draw on distinct values and belief systems. Controversies are not limited to a rational argument about the "facts", they mostly invite strong emotion responses, precisely because they bear on the values actors hold dear [Horst 2010].

Recent work on OpinioNetIT [Awadallah et al. 2012] attempts to computationally reconstruct public debates as an exchange of pro and con statements using person-opinion-topic triples. In [Garimella et al. 2016] the controversy of a topic is measured by building conversation graphs using a set of Twitter retweets on a given hashtag. Another approach by [Popescu and Pennacchiotti 2010] uses Twitter to measure the controversy of events. Their model principally relies on are linguistic, structural and sentiment features. Besides Twitter, Wikipedia has proven useful for modeling controversy on historic data. An example of this is Contropedia, where the metadata associated with Wikipedia pages such as the presence of edits and reverts were used [Borra et al. 2015]. Another method is proposed by [Dori-Hacohen and Allan 2015] to detect controversial pages on the Web through mapping them to their closest Wikipedia pages.

Collective Intelligence 2017.

Only a few articles explicitly tackled the problem of detecting controversy in news articles. [Lourentzou et al. 2015] measured which sentences trigger the largest responses in terms of tweets in order to locate the most controversial points in media coverage. [Choi et al. 2010] identified controversial topics by looking at which ones tend to invoke conflicting sentiment, and [Mejova et al. 2014] analyzed news using a crowdsourced lexicon that comprises frequent content words for which participants were asked to judge their controversy. Our work differs from each of these methods in that we try to measure controversy through all the aspects that indicate it, resulting in a more reliable measure. Further more, we combine both social media, news articles and web pages so that a more complete structure of a debate can be measured. As this structure may also be unknown and may continuously change, the state of the art approaches would not be effective for measuring controversy.

3. TEMPO CONTROVERSY MODEL

From the literature we have identified the following aspects as characteristic aspects of a controversy. We combine these in our TEMPO controversy model:

—**Time-persistence:** The debate is persistent over longer stretches of time.

—**Emotion:** The viewpoints in a debate are typically described with the expression of strong emotions.

—**Multitude of actors:** The debate typically consists of a multitude of participating actors.

—**Polarity:** The viewpoints in a debate are typically polarized and not uniform or scattered.

—**Openness:** The debate is openly exposed to the public

The TEMPO model describes controversy as an *emotional and polarized public debate by a multitude of actors persisting over time*. Following this definition, our model can be used in order to facilitate the process of detecting controversy in news articles and social media. It captures the different aspects of controversy and their relations and dependencies. We show how all these different aspects can be extracted, quantified, and integrated with existing approaches on Web content harvesting, text mining, and network analysis.

In order to identify that a topic or a document is *controversial* according to this model, we need to be able to identify a number of opinions and arguments on a common topic and measure it based on the five criteria above. Importantly, all these aspects can nowadays be assessed and quantified based on a variety of techniques and data sources. Emotion can be measured and categorized with *sentiment analysis* techniques, polarity can be detected through the *sentiment distribution* across actors, time-persistence can be measured with *dynamic network models*. Further, the openness of a controversy and the generality of the web allows us to use web content crawling to retrieve pertinent data on all dimensions. Finally, popularity can be identified with *social network analysis* and *named entity extraction* across all previously mentioned aspects of controversy.

4. EXPERIMENTAL SETUP

We evaluated the controversy aspects through a crowdsourcing experiment using the CrowdFlower[1] platform. The collected annotations from this experiment were evaluated using the CrowdTruth methodology [Aroyo and Welty 2014] for measuring the quality of the annotations, the annotators, and the annotated articles. The relevance of each of the aspects was collected by asking the annotators whether they applied to the main topic of a given newspaper article. For this, we used a collection of 5 048 Guardian newspaper articles that were retrieved through the Guardian news API. In order to save cost and focus on the main topic of an article only the first two paragraphs of each article were used. In

[1] http://crowdflower.com

Table I. : Controversy aspect weights

Controversy Aspect	Regression Coefficient	Pearson Correlation	Judgments	Articles	Aspect Clarity
Time-persistence	0.825	0.411	44.1 %	29.0 %	0.882
Emotion	1.717	0.517	55.0 %	40.3 %	0.900
Multiple actors	0.080	0.270	62.1 %	58.5 %	0.886
Polarity	1.310	0.414	56.8 %	50.4 %	0.885
Openness	0.790	0.320	70.9 %	73.0 %	0.914

an initial pilot we used 100 articles to test the use of a five point likert-scale answers versus "yes/no/I don't know" type answers, and additionally whether showing five comments would help annotators identify whether the topic in an article is controversial. In a second pilot we evaluated with the same dataset whether rephrasing of the aspects and adding the time-persistence would make the identification more clear.

5. RESULTS

The results of the first pilot showed that for both settings when showing the article comments the number of annotators that select "I don't know" option is significantly smaller (p-value = 0.003). Additionally, we found that the "yes/no/I don't know" setup always finished faster. Although this difference is not significant (p-value = 0.0519), it may indicate that annotators were more willing to perform this task. Based on this we conclude that the variant *with comments and yes-no answers* gave the best performance in terms of speed and annotation quality. The results of the second pilot showed the rephrasing of the questions improved the identification as the number of people that selected the "I don't know" option dropped from 15% to 3% with p=0.0001.

In the main experiment 5048 articles were annotated by 1 659 annotators resulting in 31 888 annotations. This dataset is available for download at the CrowdTruth data repository[2]. The evaluation of the controversy aspects was a two-fold: first the Pearson correlation coefficients were measured in order to identify how strong an aspect correlated with controversy in each judgment. Second, linear regression was applied to learn the regression coefficient between all of the aspects combined and the controversy score for a judgment. This value indicates the weight of an aspect with respect to the other aspects. As can be seen in Table I, the emotion aspect of an article was found to be the strongest indicator for controversy using both measures, while the multitude of actors was the weakest. The openness was said to be present most in 70.9% of the annotations, was annotated with a majority in 73% of the articles, and was found to be the most clearly represented aspect.

6. CONCLUSIONS

In this study we identified five aspects of controversy: the time-persistence, emotion, multiple actors, polarity and openness. Using crowdsourcing, annotations were gathered on the relevance of these aspects to 5 048 Guardian articles. The results indicate that each of these aspects is a positive indicator of controversy, but also that there is a clear difference in their signal strength. Most notably, the emotion was found to be the highest indicator. Though, all the measured controversy aspects were found to positively correlate with controversy. These results suggest that the TEMPO controversy model is accurate and useful for modeling controversy in news articles.

ACKNOWLEDGEMENTS

This publication was supported by the Dutch national program COMMIT/.

[2]http://data.crowdtruth.org

REFERENCES

Lora Aroyo and Chris Welty. 2014. The three sides of crowdtruth. *Journal of Human Computation* 1 (2014), 31–34.

Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. 2012. Opinions network for politically controversial topics. In *Proceedings of the first edition workshop on Politics, elections and data*. ACM, 15–22.

Erik Borra, Esther Weltevrede, Paolo Ciuccarelli, Andreas Kaltenbrunner, David Laniado, Giovanni Magni, Michele Mauri, Richard Rogers, and Tommaso Venturini. 2015. Societal Controversies in Wikipedia Articles. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 193–196. DOI:http://dx.doi.org/10.1145/2702123.2702436

Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng. 2010. Identifying controversial issues and their sub-topics in news articles. In *Pacific-Asia Workshop on Intelligence and Security Informatics*. Springer, 140–153.

Adele E Clarke. 1990. Controversy and the development of reproductive sciences. *Social Problems* 37, 1 (1990), 18–37.

John Cook, Dana Nuccitelli, Sarah A Green, Mark Richardson, Bärbel Winkler, Rob Painting, Robert Way, Peter Jacobs, and Andrew Skuce. 2013. Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental research letters* 8, 2 (2013), 024024.

Sébastien Dalgalarrondo and Philippe Urfalino. 2002. Tragic choice, controversy, and public decision-making: the case in France of random selection of AIDS patients for treatment (" lot-drawing"). *Revue française de sociologie* (2002), 3–40.

Shiri Dori-Hacohen and James Allan. 2015. Automated controversy detection on the web. In *European Conference on Information Retrieval*. Springer, 423–434.

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying controversy in social media. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. ACM, 33–42.

Maja Horst. 2010. Collective closure? Public debate as the solution to controversies about science and technology. *Acta Sociologica* 53, 3 (2010), 195–211.

Anna Kata. 2010. A postmodern Pandora's box: anti-vaccination misinformation on the Internet. *Vaccine* 28, 7 (2010), 1709–1716.

Jonathan Kelley, Mariah DR Evans, and Bruce Headey. 1993. Moral reasoning and political conflict: The abortion controversy. *British Journal of Sociology* (1993), 589–612.

Ismini Lourentzou, Graham Dyer, Abhishek Sharma, and ChengXiang Zhai. 2015. Hotspots of news articles: Joint mining of news text & social media to discover controversial points in news. In *Big Data (Big Data), 2015 IEEE International Conference on*. IEEE, 2948–2950.

Yelena Mejova, Amy X Zhang, Nicholas Diakopoulos, and Carlos Castillo. 2014. Controversy and sentiment in online news. *arXiv preprint arXiv:1409.8152* (2014).

Ana-Maria Popescu and Marco Pennacchiotti. 2010. Detecting controversial events from twitter. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 1873–1876.

Robert J Spitzer. 2015. *Politics of gun control*. Routledge.