Principal Areas of Exploration: News, trust, and information quality

Fact-checking Matters: The Value of Crowdsourcing for Rating Article Veracity

PI: Dr. Gita V. Johar, Vice Dean for Diversity, Equity, and Inclusion, Meyer Feldberg Professor of Business, Marketing, Columbia Business School, Columbia University, 3022 Broadway, New York, NY 10027, Office: +01 212-854- 3480, Email: gvj1@gsb.columbia.edu

Coauthor: Yu Ding, PhD Student, Marketing, Columbia Business School, Columbia University, 3022 Broadway, New York, NY 10027, Email: yu.ding@columbia.edu

ABSTRACT

This research proposes a method to leverage the input of the general population (crowdsourcing), algorithms (supervised learning), and experts (third-party fact-checkers) to rate the veracity of the information on news media. We propose the use of similarity judgments to facilitate unbiased participant responses.

BACKGROUND

Misleading information in the news media is of key interest to individual readers, publication managers, and policymakers. In recent years, both mass media and social media have been rated as more biased than ever before (Pew Research 2018; Garimella and Weber 2017). Information can be misleading if it contains a mix of true and false statements, if it is presented in a slanted way, or if it implies conclusions that are not necessarily supported by the facts (Hastak and Mazis 2011; Ecker et al. 2014; Shu et al. 2017; European Commission 2018).

Research and industry reports have relied on individual readers' self-reported beliefs and opinions to measure media trustworthiness (Pennycook and Rand 2019; Reuters 2017). For instance, readers are usually asked to answer questions such as "Does [certain news media company] provide accurate and reliable news?" or "How trustworthy is [certain news media company]?" At the article level, we see fact-checking websites such as Snopes or Fackcheck.com that employ researchers to investigate claim veracity. For scientific articles, organizations have leveraged expert opinion and published article credibility scores based on scientists' opinions (e.g., ScienceFeedback.co evaluates articles in the climate and health domains). These credibility scores go beyond fact-checking and incorporate "the quality of the logic and reasoning to analyze and expose logical flaws and rhetorical manipulation" (ScienceFeedback.co). While these efforts are truly laudable, the problem that these methodologies face is that there is a limit to the number of articles that fact-checking researchers or scientists can verify or rate. We propose leveraging the power of individuals to help with these efforts while ensuring that their ratings are valid.

A critical problem with asking individuals to rate the veracity of articles is that they are invariably biased by their prior opinions and ideology (Kunda 1990, Kahan 2012). We propose that an article similarity question rather than a belief question can surmount this problem and that crowdsourcing responses to this question can vastly increase the number of articles that are rated for veracity. A starting assumption is that there is a corpus of articles that has been rated for veracity in a valid way (e.g., by scientists), so that our approach can use these ratings as a starting point. Our initial research focus is on scientific topics such as climate change and articles related to healthcare (e.g., Covid-19 treatment or vaccine).

RESEARCH DESIGN

We propose asking individuals to judge the similarity of unrated articles with articles rated for veracity by experts, and to use this similarity judgment as well as the expert rating as input to provide a veracity rating for unrated articles. By leveraging the power of the crowd, we can vastly scale up the number of articles that are rated. Before this methodology can be implemented, we need to test certain hypotheses. First, we expect that because <u>similarity judgments are constructed rather than retrieved</u> (Tversky 1977), they are less likely to be biased by readers' ideology and more likely to reach high consensus among different reader populations. This hypothesis has been supported by our preliminary work that has also identified characteristics of individuals that can help improve similarity judgments. <u>Second, we suggest that combining similarity judgments and experts' veracity ratings can provide valid ratings.</u> We propose to validate our approach by comparing our predictions with expert ratings of a holdout sample. This work is ongoing. <u>Third, we propose that individuals can provide more accurate similarity judgments than machine learning algorithms because of the nuanced nature of articles.</u> We propose to test this hypothesis by comparing the predictive power of crowdsourced similarity judgments versus that of machine learning algorithms such as LSA, Word2Vec, and GloVe (Naili, Chaibi, and Ghezala 2017).

PROPOSED RESEARCH PLAN

(1) ***Conduct a large-scale survey to confirm that political ideology and similarity judgments are uncorrelated.*** We plan to conduct a pre-registered large-scale experiment to verify that ideology and similarity judgments are uncorrelated for other scientific topics (e.g., in the healthcare domain, articles about COVID-19).
(2) ***Validate predicted veracity scores against expert ratings.*** We plan to test the validity of ratings inferred from similarity judgments against actual expert ratings. In this set of studies, we will use a set of articles that have been rated by experts and compare our predicted ratings with actual ratings.
(3) ***Compare the predictive power of crowdsourced similarity ratings with similarity ratings of machine learning algorithms.*** We will model the value of human similarity ratings as an addition or an alternative to machine learning models, acknowledging the benefits and drawbacks of each approach.
(4) ***Improve the predictive model by <mark>adding semantic markers of veracity</mark>.*** Research has identified semantic markers of information veracity, such as the use of extreme language and anger (Li et al. 2014; Guo et al. 2019). We plan to employ different methods of supervised learning to extract these markers and combine them with crowdsourced similarity ratings to build a tool that can improve the validity of veracity ratings.
(5) ***Develop a platform for individual readers to perform similarity ratings.*** As a final step, we plan to help create a platform that can generate articles within a topic area and present them to individuals for similarity ratings. All steps of the model can be automated to generate veracity ratings based on similarity judgments and other information included in the final model.

CONTRIBUTION

We believe this project will benefit society by involving readers as contributors to the article's veracity rating system and thus increase citizen participation in our democracy. In the long run, participation in crowdsourced veracity ratings could increase media literacy. Consumers are also likely to trust these ratings because they are from the crowd rather than a source that could be viewed as being biased.

We see this project as topical, sustainable, and cross-disciplinary. Besides publishing our findings in top-tier general interest journals, we also hope to work with partners to scale up the effort and build the technology to leverage and use crowdsourced similarity judgments as input to veracity ratings.

APPENDIX 1 - PRELIMINARY FINDINGS 验证假设：无偏见.

Study 1 examined our hypothesis that similarity judgments are unbiased by prior ideology. Participants were instructed to read three news articles that either supported the idea of climate change (Article 1), contained balanced arguments (Article 2), or disagreed with climate change arguments (Article 3). Half of the participants were asked to make evaluation judgments (how true the argument is in each article; how believable the argument is in each article; how trustworthy the source is for each article) and the other half of the participants were asked to make similarity judgments (how similar the arguments are between the two articles; how much the arguments align in the two articles; how likely it is that the two articles are from the same source).

Results revealed that responses to evaluation questions are significantly affected by readers' political ideology, with correlations varying from 0.73 to 0.83. For example, more liberal participants were more likely to rate the article that is against climate change false (Article 1), $b = -.37$, $p = .002$; but rate the article that shows evidence of climate change true (Article 3), $b = .45$, $p < .001$. Importantly, similarity judgment ratings are not affected by political ideology, $ps > .88$. That is, both conservative and liberal participants rated the similarities between articles in the same way regardless of their political ideology. Participants who scored higher on the Cognitive Reflection Task (CRT) were better able to distinguish articles with opposing arguments, $b = -.56$, $p = .002$. However, the CRT score did not affect the relationship between political ideology and responses to evaluation questions. Study 2 replicated the findings in Study 1 with a similar experimental design and another set of news articles on climate change with a different subtopic about Arctic Sea Ice (the articles in Study 1 were about Polar Bears).

Political ideology is correlated with true/false judgments on climate change articles, but uncorrelated with similarity judgments on climate change articles (Study 1)
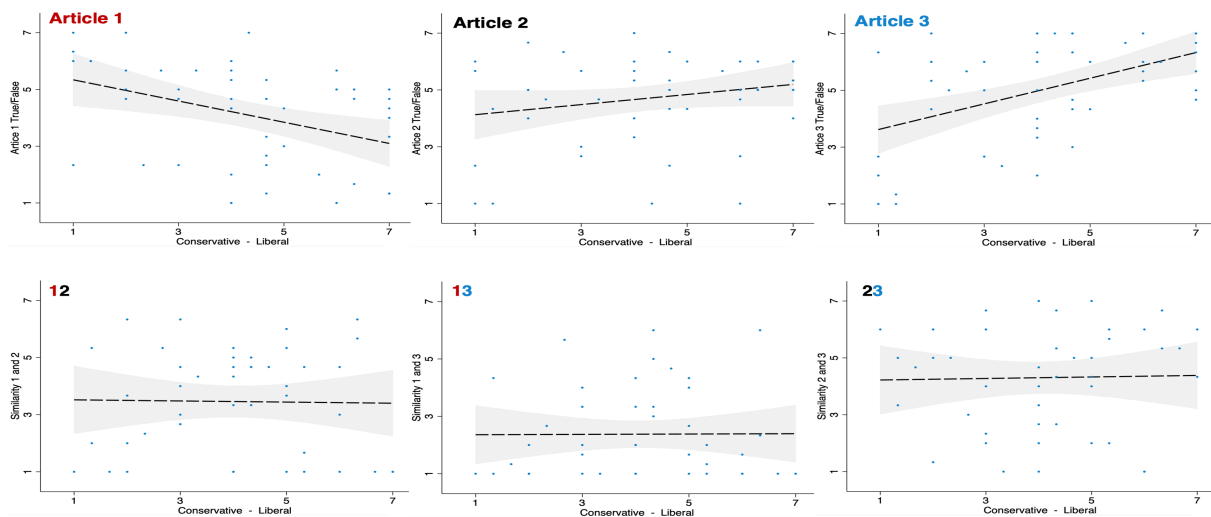
APPENDIX 2 - REFERENCES

Ecker, U. K., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied*, *20*(4), 323-335.

European Commission. Directorate-General for Communication Networks, Content and Technology. (2018). *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation*. Publications office of the European Union.

Garimella, V. R. K., & Weber, I. (2017, May). A long-term analysis of polarization on Twitter. In Eleventh International AAAI Conference on Web and Social Media.

Green, P. E., Wind, Y., & Carroll, J. D. (1973). Multiattribute decisions in marketing: A measurement approach (Vol. 973). Hinsdale, IL: Dryden Press.

Guo, C., Cao, J., Zhang, X., Shu, K., & Liu, H. (2019). DEAN: Learning Dual Emotion for Fake News Detection on Social Media. *arXiv preprint arXiv:1903.01728*.

Hastak, M., & Mazis, M. B. (2011). Deception by implication: A typology of truthful but misleading advertising and labeling claims. *Journal of Public Policy & Marketing*, *30*(2), 157-167.

Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision Making*, *8*, 407-24.

Li, J., Ott, M., Cardie, C., & Hovy, E. (2014). Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1566-1576).

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74.

Naili, M., Chaibi, A. H., & Ghezala, H. H. B. (2017). Comparative study of word embedding methods in topic segmentation. *Procedia Computer Science*, *112*, 340-349.

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences*, *116*(7), 2521-2526.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, *19*(1), 22-36.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An Identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213-224.

Reuters. (2017). Reuters institute digital news report 2017. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf

Pew Research Center. (2018). Publics globally want unbiased news coverage but are divided on whether their news media deliver. Retrieved from https://www.pewresearch.org/global/2018/01/11/publics-globally-want-unbiased-news-coverage-but-are-divided-on-whether-their-news-media-deliver/