

6.804/6.864 HW2 TASK B: Credibility Evaluation

The internet and social media platforms have made sharing information incredibly easy, but with this has come the increasing spread of false or misleading information. Recently, many factors — like the design of social media sites and political polarization — have exacerbated the misinformation problem. As we have seen, the prevalence of online misinformation can have dire consequences on civil discourse and democratic processes. When it comes to topics like vaccine safety or climate change, the quality of information the public is consuming can have a big impact on public health and policy.

Increasingly, it seems like it's not enough to just ask readers to evaluate the credibility of news or other information on their own. Indeed, studies have found that people are not very good at doing this.¹ Online, social, and search platforms are attempting to find interventions that can help stop the spread of misinformation. For example, several crowd-sourced fact-checking initiatives try to have people rate research claims in an article and rate it according to how truthful it is.²

However, manual crowdsourcing credibility assessments can be difficult to scale. Some companies, like Facebook, also employ a machine learning model to filter articles to those that the model thinks might have misinformation, and then has reviewers check that smaller pool of articles.³ Models to assess credibility could also be used to filter articles on people's news feeds or to display warnings next to articles with low predicted credibility.

These models are often trained on data about existing articles, annotated by how credible or trustworthy they are. In this part of the assignment, pretend you're on the development team of a new social media network that wants to limit the spread of misinformation. You're collecting a dataset like this to study signs of misinformation and to research models that could detect it. You decide to focus on climate change as a topic where misinformation is prevalent, and collect a set of articles scraped from different news sites. You have a budget to hire annotators to label them, but you need a labeling scheme and instructions to give them.

Here are links to some example articles from your dataset:

- <https://www.nytimes.com/2018/09/10/climate/united-nations-climate-change.html>
- <https://blogs.scientificamerican.com/observations/more-recycling-wont-solve-plastic-pollution/>
- <https://www.westernjournal.com/report-poor-management-forests-global-warming-blame-widespread-wildfires/>
- <https://dailycaller.com/2017/11/29/study-satellites-show-no-acceleration-in-global-warming-for-23-years/>
- <https://www.iflscience.com/policy/epa-kill-off-obama-signature-plan-fight-climate-change/>

¹ Wineburg, Sam and McGrew, Sarah and Breakstone, Joel and Ortega, Teresa. (2016). Evaluating Information: The Cornerstone of Civic Online Reasoning. Stanford Digital Repository. Available at: <http://purl.stanford.edu/fv751yt5934>

² E.g., see <https://eufactcheck.eu/>, <https://fullfact.org/>

³ Henry Silverman. 2019. Helping Fact-Checkers Identify False Claims Faster - About Facebook.

MIT OpenCourseWare
<https://ocw.mit.edu>

RES.TLL-008 Social and Ethical Responsibilities of Computing (SERC)
Fall 2021

For information about citing these materials or our Terms of Use, visit:
<https://ocw.mit.edu/terms>