Fact Checking and Information Retrieval

Matthew Lease
School of Information, University of Texas at Austin, USA
ml@utexas.edu

INTRODUCTION. Designating October 2009 as National *Information Literacy*[1] Awareness Month, former U.S. President Barack Obama promoted a key 21st century information challenge: "Though we may know how to find the information we need, we must also know how to evaluate it. Over the past decade, we have seen a crisis of authenticity emerge. We now live in a world where anyone can publish an opinion or perspective, whether true or not, and have that opinion amplified within the information marketplace."

Historically, we have relied upon Information Literacy education to teach key critical reading skills, use of multiple sources, and potential for source bias. However, today's era of information overload presents an historic stress test of traditional information literacy skills. Information tracking and sense-making has become increasingly difficult, along with effort required to consistently and effectively cross-check sources by hand. The rise of misinformation – unwitting or deliberate – has further exacerbated this.

In response, researchers in natural language processing (NLP) and machine learning (ML) have proposed a variety of innovative new models to automatically fact-check claims. However, these works have largely approached fact-checking as a fully-automated task focused whose primary goal it to maximize predictive accuracy. While accurate predictions are important, someone skeptical of online information is likely to be equally skeptical of any fact-checking tool. Thus, a system should also be transparent (and auditable) in how it made a prediction so that a user can understand and trust the model. In addition, an individual's claim assessments will invariably rely at least in part on that person's prior world-views regarding the perceived credibility of claims and sources. A fact-checking system should be open-ended to integrate user beliefs, letting users easily inject their own views and knowledge into the system. Finally, a model should communicate the uncertainty in its predictions while accounting for potential sources of errors, empowering users to conduct their own in-depth inspection and reasoning.

OUR WORK. In 2012, we prototyped[2] a system for searching and browsing *memes* underlying news: similar phrases which spread and evolved across sources [3]. Once detected, these latent memes were revealed to users via generated hypertext, allowing memes to be recognized, interpreted, and explored in context. "Our vision [was] to complement traditional forms of critical literacy education with . . . smarter browsing technology . . . Instead of understanding online narrative through only a single source, we can instead explore how broader community discourse has shaped its development . . . [especially for] campaigns which flood social media with repeated stock phrases while obfuscating their . . . source."

This year, we proposed a graphical modeling approach to fact-checking which augments the efficiency and scalability of automated information retrieval (IR) with transparent, explainable ML [1]. Given a claim as query, the system first finds and retrieves relevant articles from a variety of sources. It then infers the degree to which each article supports or refutes the claim, as well as the reputation of each source. The system then aggregates this body of evidence to predict the veracity of the claim, showing the user precisely which information is being used and the various sources of model uncertainty underlying the overall claim prediction. We also evaluated a hybrid variant in which the system integrates online crowd workers to further improve predictive accuracy.

In our most recent work (under review), we prototyped[3] a *mixed-initiative* design for incorporating user knowledge and beliefs into model predictions. The model's predicted source reputation and stance for each retrieved article is shown to the user and can be revised via simple sliders to reflect user beliefs and/or to correct erroneousness model estimates. The overall claim prediction is then updated visually in real-time as the user interacts with the system. In a user study asking participants to assess claims using variant systems and interaction mechanisms, we found that users tend to highly trust model predictions, benefiting from the model when it is correct, but also (unfortunately) falling victim to its errors. Given the option to interact with these incorrect predictions, however, users were able to do so and improve their own performance, emphasizing the need for interpretable, interactive models.

IR RESEARCH QUESTIONS. What can IR bring to fact checking that NLP and ML do not, and what new questions does fact checking raise for IR? As an explicit or implicit task in IR, fact checking has various implications for both system-centered and user-centered IR: which results should we return, how should we present them, what modes of interaction should we provide, and how should we evaluate success? While assessing the authority of pages for ranking and filtering is not new, fact checking presents a different framing of authority, with ranking and filtering decisions potentially impacting user trust of the system and fears of being manipulated[4]. Beyond topical diversification of search results, how might we diversify political (or other forms of) bias to provide diverse perspectives, especially on controversial topics? How should we rank and evaluate search results based on such diversity?

In terms of personalized search, how do we balance giving users search results matching their existing beliefs vs. challenging those beliefs with alternative viewpoints, and without such challenges

[1] en.wikipedia.org/wiki/Information_literacy
[2] odyssey.ischool.utexas.edu/mb/

[3] fcweb.pythonanywhere.com
[4] fortune.com/2015/08/23/research-google-rig-election

DESIRES 2018, August 2018, Bertinoro, Italy

prompting search engine switching behavior? Just as people choose different news outlets to follow having different political leanings, perhaps a new class of vertical search engines will soon arise which rank and filter search results to match a given audience's views?

While search results have traditionally been evaluated with respect to gain (i.e., how much the user benefits), recent work has explored the idea that, as with any technology, search engines and their results also have the potential to inflict harm on users [2]. How do we frame, measure, and address potential harm of search results including "alternative" facts, be they search result errors or intentional diversification? How do we evaluate traditional information gain alongside not only viewpoint diversification, but also potential varying severity of harm to varying numbers of users?

Ultimately, what is the duty and opportunity for IR in fact checking? How do we effectively present model predictions to benefit users while also conveying uncertainty, supporting mixed-initiative decision making and creating interfaces inviting user exploration?

REFERENCES

[1] An T. Nguyen, Aditya Kharosekar, Matthew Lease, and Byron C. Wallace. 2018. An Interpretable Joint Graphical Model for Fact-Checking from Crowds. In *AAAI*.
[2] Frances A Pogacar, Amira Ghenai, Mark D Smucker, and Charles LA Clarke. 2017. The Positive and Negative Influence of Search Results on People's Decisions about the Efficacy of Medical Treatments. In *ACM ICTIR*. 209–216.
[3] Hohyon Ryu, Matthew Lease, and Nicholas Woodward. 2012. Finding and Exploring Memes in Social Media. In *ACM Hypertext*. 295–304.