Truth Finding in Response to Misinformation

Ebrahim Bagheri
Ryerson University
bagheri@ryerson.ca

Misinformation has become a prevalent phenomenon that is impacting the general public across the world on a wide variety of issues, e.g., misinformation campaigns have linked vaccination to autism, or have downplayed the potential impact of the coronavirus pandemic in order to minimize its impact on economic drivers. The 2019 Cigi-Ipsos survey reported that 86% of the participants had fallen for fake news at least once (https://bit.ly/2y3rjEs). Vosoughi et al. [1] further show that it is not only about impact on an individual level but that falsehood has deeper and faster penetration rates compared to the truth and spreads wildly. A misinformed public not only suffers harm from receiving and acting upon misinformation, but also loses trust in domain authority and in democratic institutions. Research has shown that immediate response to misinformation is required to minimize its impact—otherwise, repeated exposure to misinformation will lead to the public's acceptance of the incorrect information as fact [2].

Existing literature on computational approaches for handling misinformation have primarily been focused on the detection, tracking and propagation of misinformation on online social platforms [3]. From the detection point of view, a host of machine learning-based classification models, ranging from those using features derived from psychological and sociological literature to fully end-to-end deep learning models, have been introduced. These models focus on determining whether one or a collection of online documents are (un)intentionally conveying factually incorrect information [4]. From the perspective of tracking and propagation, researchers have explored techniques that identify the source of misinformation, how misinformation propagates in a social structure through information diffusion patterns, and the ways in which influence networks impact misinformation [5].

In our approach towards building computational models, we chose to focus on techniques to enable the automated retrieval of correct information in response to misinformation, offering positive societal impact by reducing the harm misinformation engenders to the wellbeing of the public. From a theoretical perspective, the abundance of misinformation on social networks and their associated fact-checking initiatives have created a suitable sandbox with appropriate organic datasets for information retrieval and social analytics research, making it an ideal platform to explore new tasks in social information retrieval.

As such, the objective of our research is not to detect or track misinformation but rather to (1) *identify* factually correct information that is directly relevant to a given misinformation campaign, (2) *assess* it from the perspective of coverage, convincingness and trustworthiness, and (3) *present* it to the general public in way that maximizes the likelihood of their adoption. In other words, in contrast to the existing literature, our work is engaged with the development of computational techniques that effectively address misinformation claims once the misinformation has been identified. This work will simplify truth-finding endeavors in order to minimize the impact of misinformation.

LIST OF REFERENCES

[1] Vosoughi S, Roy D, and Aral S. (2018) "The spread of true and false news online." Science 359.6380 (2018): 1146-1151.

[2] Crozier W, Strange D. (2019). “Correcting the misinformation effect.” *Applied Cognitive Psychology* 33(4). 585-95.
[3] Shu K, Wang S, Lee D, Liu H. (2020). “Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements.” arXiv preprint arXiv:2001.00623.
[4] Zhang X, Ghorbani AA. (2020). “An overview of online fake news: Characterization, detection, and discussion. *Information Processing & Management* 1;57(2). 102-125.
[5] Shu K, Bernard HR, Liu H. (2019). “Studying fake news via network analysis: detection and mitigation.” In *Computational Social Network Analysis and Mining* (Springer). 43-65.