Combating Disinformation on Social Media and Its Challenges: A Computational Perspective

Kai Shu

Department of Computer Science
Illinois Institute of Technology
kshu@iit.edu

Abstract

The use of social media has accelerated information sharing and instantaneous communications. The low barrier to entering social media enables more users to participate and keeps them engaged longer, incentivizing individuals with a hidden agenda to spread disinformation online to manipulate information and sway opinion. Disinformation, such as fake news, hoaxes, and conspiracy theories, has increasingly become a hindrance to the functioning of online social media as an effective channel for trustworthy information. Therefore, it is imperative to understand disinformation and systematically investigate how to improve resistance against it. This article highlights relevant theories and recent advancements of detecting disinformation from a computational perspective, and urges the need for future interdisciplinary research.

Introduction

Social media has become the leading platform for individuals to communicate. In particular, social media is immensely popular for news dissemination, information sharing, and event participation due to its capacity to rapidly spread information at scale. While this massive capacity can promote social trust and enhance social connectivity, it can also facilitate the rampant propagation of disinformation. Such rampant disinformation often leverages the trust and social connectivity of social media users, spreads manipulated information to foment hatred, and inflicts damages on individuals or groups. With disinformation growing at unprecedented volumes on social media, disinformation is now viewed as one of the greatest threats to democracy, justice, public trust, freedom of expression, journalism, and economic growth. Hence, there is a pressing need to tackle digital disinformation for social good.

However, detecting disinformation and fake news with computational approaches poses unique challenges that make it nontrivial (Shu et al. 2017; Shu and Liu 2019). First, the data challenge has been a major roadblock because the content of fake news and disinformation is rather diverse in terms of topics, styles and media platforms; and fake news attempts to distort truth with diverse linguistic styles while simultaneously mocking true news. Thus, obtaining annotated fake news data is non-scalable, and data-specific embedding methods are not sufficient for fake news detection with little labeled data. Second, the evolving challenge of fake news and disinformation is another obstacle in this task—fake news is usually related to newly emerging, time-critical events, which may not have been properly verified by existing knowledge bases due to the lack of corroborating evidence or claims. Third, the explainability challenge is concerned with the development of machine learning algorithms for disinformation that are explainable. Existing disinformation detection techniques are often machine learning black boxes that provide little or no explanation on the detection process. Explainability ensures that the developed algorithms are transparent, ethically responsible and trustworthy. However, deriving algorithmic explanations useful for domain experts and enhancing transparency by understanding and incorporating prior expert knowledge to prediction models has been challenging.

In this surveyed research, we look into the recent advancements of computational approaches to detect disinformation and fake news on social media. The techniques developed combine interdisciplinary theories and computational algorithms to help policymakers and social media users address disinformation. In particular, these methods are attempting diverse challenging scenarios, towards detecting fake news more effectively, with explainability, at an early stage, and across domains. We further discuss the open issues and future directions along the line of combating disinformation from a computational perspective, and the pressing need for interdisciplinary research.

Acknowledgements

This material is supported by the Cisco Faculty Research Award, and the seed grant of the Active Computational Thinking (ACT) Center at Illinois Institute of Technology.

References

Shu, K.; and Liu, H. 2019. Detecting fake news on social media. *Synthesis lectures on data mining and knowledge discovery*, 11(3): 1–129.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1): 22–36.