Research Statement

Wei GAO
School of Computing and Information Systems,
Singapore Management University
Tel: (65) 6826-1369; Email: weigao@smu.edu.sg
30/12/2020

Background

Generally speaking, I'm part of the research community that intersects Artificial Intelligence and Data Science. In this community, researchers basically look to deliver innovations in problem-solving for providing theoretical and practical solutions in business, industry and public sectors by intensively utilizing machine learning and data-driven approaches. My primary research interest lies in Natural Language Processing (NLP) and Social Computing (SC), which is concerned about fundamental and applied technologies for developing effective methods to find, collect, process, represent and predictively analyze the target information.

Nowadays, social media outlets have been ubiquitous and featured as a kind of noisy, contentious and viralable source of information. This is due to the fact that social media platforms are like wild west, which are carriers of crowded content posted without necessary moderation and allowing undomesticated flow of posts to circulate and get populated, and no matter whether the content per se is real or not. This has created tricky challenges both socially and technically, but meanwhile it also opens ample opportunities for problem solving from the socio-technical point of view which may benefit many downstream applications built on top of social media. I believe that such a wild nature of social media context prompts to break a new horizon in the broader data-centric AI research and some of its relevant sub-fields, especially NLP and social network analysis. With the continuous proliferation of social media, my research is strongly motivated to develop new methods and viable technologies to invigorate the area of Social NLP based on principled approaches.

Research Areas

Rumor and misinformation detection: The popularity of social media platforms such as Twitter, Facebook, etc. has pushed information dissemination and sharing to a new height thanks to phenomenal real-time citizen journalism. However, without systematic effort of moderation, large volume of false or unverified information can spread quickly and pollute our online community via various malicious activities like spreading rumors, fabricating news reports, conducting unfair social advertising or political campaigns. Massive spread of misinformation brings individuals and society an increasing danger of devastating repercussions, which has rendered the daunting “post-truth-era” phenomena and posed unprecedented challenges to the contemporary mass media ecosystem.

Public survey shows that only 4% of common adults can correctly identify whether a news story is true or fake. Humans are naturally not good at identifying misinformation due to the inherent cognitive biases. Fact-checking is the professional act of verifying claims in news reports to determine the veracity and truthfulness of the asserted statements. Some popular fact-check websites such as snopes.com and politifact.com conduct post-hoc verification on the veracity of circulated news events originated from social media posts and news reports. However, this non-trivial process requires tremendous amount of time devoted to manual investigation and analysis, and moreover it is prone to low efficiency and poor coverage due to the complexity of topic to check and is incompetent with the speedy generation and diffusion of misinformation.

In the past, my research work has been focused on automated approaches to rumor and misinformation detection [1, 2, 3, 4, 5, 6, 7, 9] relying on training supervised classifiers, for which past events or claims are gathered and labelled as trustworthy or fake. For strengthening event representation and classification effectiveness against misinformation, useful features based on linguistic, temporal, network and propagation structures are learned and combined in various novel manners with social media posts that are relevant to the concerned events. For example, to avoid painstakingly detailed feature engineering effort [1], we tried to exploit deep neural models to learn the high-level rumor-indicative representations with recurrent neural networks [2], and extended this sequential model into a range of structured models based on tree kernel [3], tree-structured recursive neural networks [4,5] and more recently tree transformers [6], considering the propagation structure and user responses sparked by source posts. Besides, we studied to reinforce stance detection and rumor detection jointly in a unified neural multi-task learning model [7], considering the task-invariant features shared to signal the veracity of a claim and the stances of its responsive posts.

However, it is found that fact-checking is inherently difficult and oftentimes controversial due to the complexity of the task [8]. General research on information credibility evaluation is still premature to deal with the wild nature in social media context. For example, the methods which tried to classify the veracity of news based on assessing and comparing user reactions, viewpoints or stances can suffer from low recall since news may not spark enough enquiry posts. Also, the responsive information could be noisy, ambiguous, propagandistic or astroturfed, thus are largely unreliable for deducing a solid credibility assessment [9].

With multi-modality and multi-facet characteristic of online data, I believe it is necessary to dig more thoroughly the key surface and innate factors rooted in the multi-modal content, users and network structures and leverage them all for the early detection of misinformation across different social and web media platforms. Beyond just detection, it is also helpful to study and understand the user perception and behaviors with different types of misinformation online, and further propose pre-emptive solutions to prevent misinformation from going viral, which is crucial for effective treatment and intervention to combat their spreading.

To this end, I firstly plan to attack the detection of online misinformation with a multi-modal NLP and learning approach combined with human intelligence through crowdsourcing. This is important and technically interesting to the realization of fast, high-confidence and interpretable cross-checking of the credibility of target information. Secondly, instead of just finding out who have actively engaged in spreading rumors [10], I plan to develop effective and reliable intervention methods to complement the detection of misinformation based on user perception and behavior studies for the early red-flagging of malicious and vulnerable users. This may be attempted with the aid of profiling techniques, such as learning user representations that could accurately encode stance-specific user embeddings through one's online language use and behaviors rooted from their relatively stable world views. My work going to this direction is based on the assumption that the "infodemic" of misinformation is basically originated from the good match of the "value" of specific information (like the spike glycoprotein of coronavirus) and the value bias of the accepting users (like the receptor protein of human cell). In general, this is concerned about the hard questions of how to accurately encode the information and user in the right dimension and how to measure the strength of their potential cohesion and relevance. I would think that determining such information-user cohesiveness by matching their innate values is deeper and more essential than the common semantic matching in language understanding. How this can be approached with exiting semantic technologies to improve the understanding of language and user is a profound challenge of data-centric AI.

Based on the proposed technical solutions to misinformation, I'm targeting to gain useful insights towards the commonly concerned questions, such as 1) Do different types of misinformation (e.g., topics, modalities, actor types) have different characteristics or features? Which types of misinformation go viral or have larger spread? Do they have certain traits, e.g., emotions, messaging strategy, user base? 2) Further, when does a piece of misinformation begin to propagate? How likely will it become viral? And if so, to what extent? 3) Who are the actors (e.g., individuals, organizations, and websites) responsible for the creation of misinformation, the spread of misinformation and debunking of falsehoods? How likely does a user become any one of such actors? …

Given the new development of major world events and crises in recent years such as the COVID pandemic that have enriched the soil for misinformation to flourish, we can speculate that with the increasingly stressed societal fault lines the problem of online misinformation will get more serious in the foreseeable future both domestically and globally. Overall, this research area requires effective fusion of different computational methods and cross-disciplinary thinking for meaningful fact-checking and intervention at various levels, which will constitute a coherent and integrated research line of social NLP for dealing with the challenges.

Selected Publications and Outputs

[1] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.F. Wong. Detect rumors using time series of social context information on microblogging websites. CIKM 2015.

[2] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.F. Wong, and Meeyoung Cha. Detecting rumors from microblogs with recurrent neural networks. IJCAI 2016.

[3] J. Ma, W. Gao, and K.F. Wong. Detect rumors in microblog posts using propagation structure via kernel learning. ACL 2017.

[4] J. Ma, W. Gao, and K.F. Wong. Rumor detection on Twitter with tree-structured recursive neural networks. ACL 2018.

[5] J. Ma, W. Gao, S. Joty, and K.F. Wong. An attention-based rumor detection model with tree-structured recursive neural networks. ACM Transactions on Intelligent Systems and Technology, Vol. 11, Issue 4, Article No. 42, June 2020.

[6] J. Ma and W. Gao. Debunking rumors on Twitter with tree transformer. COLING 2020.

[7] J. Ma, W. Gao, and K.F. Wong. Detect rumor and stance jointly by neural multi-task learning. WWW2018 Companion.

[8] J. Ma, W. Gao, S. Joty, and K.F. Wong. Sentence-level evidence embedding for claim verification with hierarchical attention networks. ACL 2019.

[9] J. Ma, W. Gao, and K.F. Wong. Detect rumors on Twitter by promoting information campaigns with generative adversarial learning. WWW 2019.

[10] B. Rath, W. Gao, J. Ma, and J. Srivastava. From retweet to believability: Utilizing trust to identify rumor spreaders on Twitter. ASONAM 2017.