7^{th} International Conference on Computational Social Science IC^{2}S^2
July 27-31, 2021, ETH Zürich, Switzerland

Modeling the Spread of Fake News on Twitter

Taichi Murayama*[1], Shoko Wakamiya[1], Eiji Aramaki[1], and Ryota Kobayashi†[2]

[1]Nara Institute of Science and Technology (NAIST), Japan
[2]The University of Tokyo, Japan

Keywords: fake news, point process, modeling, twitter, time series prediction

Extended Abstract

The growing use of mobile devices and the worldwide increase in Internet access have enabled us to share various types of information and to discuss it with other readers. However, it also seems to have become a hotbed of fake news with potentially negative influences on society. In this study, we investigate the question of how fake news spreads on Twitter. The question has practical implications for fake news detection and mitigation. Previous studies mainly focused on the path taken by fake news items as they spread on social networks [1], which clarified the structural aspects of the spread. However, little is known about the temporal or dynamic aspects of how fake news spreads online. Here, we propose a novel mathematical model that describes the spreading of fake news [2].

The proposed model is based on the following two assumptions (Fig. 1):

- Users do not know the falsity of a news item in the early stage, and the fake news spreads as an ordinary news story.
- Users then recognize the falsity of the news item around a correction time t_c. The information that the original news is fake spreads as another news story.

We formulate this assumption by extending the Time-Dependent Hawkes process (TiDeH) [3], a state-of-the-art model for predicting re-sharing dynamics on Twitter. The instantaneous rate of tweets or retweets triggered by a piece of fake news at time t represents $\lambda(t) = \lambda_1(t) + \lambda_2(t)$, where the first term $\lambda_1(t)$ represents the rate of the cascade caused by the original news item and the second term $\lambda_2(t)$ represents the cascade induced by the correction.

Datasets of the spread of fake news based on retweets of the original news post are publicly available [4]. However, the information sharing of fake news can be complex rather than a simple retweet. Thus, we manually compiled two datasets of fake news items from Twitter: 1) Recent Fake News dataset in 2019 and 2) the 2011 Tohoku earthquake dataset. Fake news items were identified using two fact-checking sites (Politifact.com and Snopes.com) and a fake news verification article. We validate the proposed model using the two datasets. Experimental result of predicting the spread of fake news shows that the proposed model is superior to the current state-of-the-art methods for the both datasets. Moreover, we demonstrate that our model

*taichi.murayama.mk1@is.naist.jp
†r-koba@k.u-tokyo.ac.jp

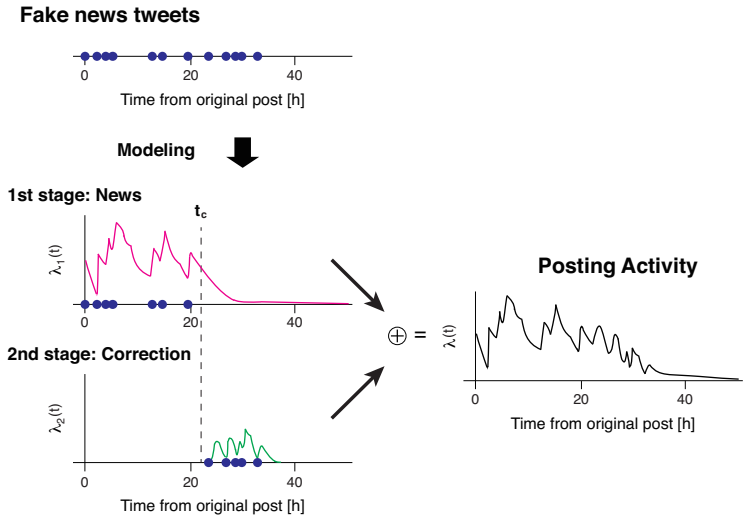

Figure 1: Schematic of the proposed model. We propose a model that describes how posts or re-shares that are related to a fake news item spread on social media (Fake news tweets). Blue circles represent the time stamp of the tweets. The proposed model assumes that the information spread is described as a two-stage process. Initially, a fake news item spreads as a novel news story (1st stage). After a correction time t_c, Twitter users recognize the falsity of the news item. Then, the information that the original news item is false spreads as another news story (2nd stage). The posting activity related to the fake news $\lambda(t)$ (right: black) is given by the summation of the activity of the two stages (left: magenta and green).

appropriately infers the correction time, i.e., the moment when Twitter users start realizing the falsity of the news item.

Summarizing above, we have proposed a two-stage process model for describing the spreading of fake news on Twitter. The proposed model outperforms the state-of-the-art methods for accurately predicting the spread of fake news items and was able to infer the correction time of the news story. The result implies that the change in the perception of the content can be a potential driver for the cascade recurrence. Our model contributes to understanding the dynamics of the spread of fake news on social media. Its ability to extract a compact representation of the spreading pattern could be useful in the detection and mitigation of fake news.

References

[1] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[2] Taichi Murayama, Shoko Wakamiya, Eiji Aramaki, and Ryota Kobayashi. Modeling the spread of fake news on twitter. *PLOS ONE 16(4): e0250419*, 2021.

[3] Ryota Kobayashi and Renaud Lambiotte. Tideh: Time-dependent hawkes process for predicting retweet dynamics. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 191–200, 2016.

[4] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. *Big Data*, 8(3):171–188, 2020.