All the Wiser: Fake News Intervention Using User Reading Preferences

Kuan-Chieh Lo*
kclo7898@iis.sinica.edu.tw
Institute of Information Science,
Academia Sinica
Taipei, Taiwan

Shih-Chieh Dai*
sjdai@iis.sinica.edu.tw
Institute of Information Science,
Academia Sinica
Taipei, Taiwan

Aiping Xiong
axx29@psu.edu
Pennsylvania State University
University Park, PA, USA

Jing Jiang
jingjiang@smu.edu.sg
Singapore Management University
Singapore

Lun-Wei Ku
lwku@iis.sinica.edu.tw
Institute of Information Science,
Academia Sinica
Taipei, Taiwan

ABSTRACT

To address the increasingly significant issue of fake news, we develop a news reading platform in which we propose an implicit approach to reduce people's belief in fake news. Specifically, we leverage reinforcement learning to learn an intervention module on top of a recommender system (RS) such that the module is activated to replace RS to recommend news toward the verification once users touch the fake news. To examine the effect of the proposed method, we conduct a comprehensive evaluation with 89 human subjects and check the effective rate of change in belief but without their other limitations. Moreover, 84% participants indicate the proposed platform can help them defeat fake news. The demo video is available on youtube[1].

KEYWORDS

Fake News Intervention; Web Application; Human-Subject Experiment

ACM Reference Format:

Kuan-Chieh Lo, Shih-Chieh Dai, Aiping Xiong, Jing Jiang, and Lun-Wei Ku. 2021. All the Wiser: Fake News Intervention Using User Reading Preferences. In *Proceedings of the Fourteenth ACM International Conference on Web Search and Data Mining (WSDM '21), March 8–12, 2021, Virtual Event, Israel.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3437963.3441696

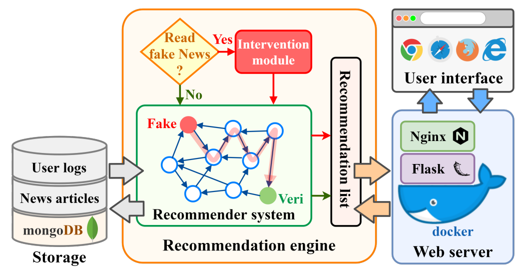


Figure 1: The news platform framework, which consists of four components: 1) storage, 2) recommendation engine, 3) web server, and 4) user interface

1 INTRODUCTION

With the development of the Internet, people consume news articles more through online platforms than through traditional newspapers or magazines. Nevertheless, the lack of information scrutiny on those platforms has increased the rise of misinformation (i.e., fake news). Moreover, fake news has permeated major events and has influenced various aspects of modern life [1, 12, 14]. To address the issues of fake news, fact-checking has been proposed and implemented by the government [5], industry [11], and academia [15]. However, such a solution ends up being a trade-off between the benefit of accumulating more reports of fake news and the risk of exposing more users to potential fake news. Because the work of fact-checking by experts or crowd source is labor-intensive and time-consuming, another popular approach is to adopt machine learning algorithms to detect fake news in an automatic manner [18]. Nevertheless, fake news detection is still at the experimental stage. Given the effort on automatic detection and fact-checking, effectively delivering detection or fact-checking results such that people reduce their belief in fake news is still a challenge [16]. The most common way to deliver results is to warn online users using warning tags. However, the effect of such warnings is limited [7, 13, 16]. Furthermore, some studies indicate somewhat negative impacts of warning tags, such as implied truth [14] and warning habituation [2]. Alternatively, some studies present verified news articles to participants when they read fake news simultaneously [17, 18]. Nevertheless, the results thereof reveal an adverse effect: participants whose attitude is congruent with the fake news *increase* their belief in the fake news; in other words, it backfires [8].

In light of the issues mentioned above, we propose a new intervention module to expose users to the verified news in an implicit manner (see Fig. 1). Instead of showing the verified news and fake news to users simultaneously, we take into account each user's

*Both authors contributed equally to this research.
[1]https://youtu.be/wKI6nuXu_SM

ACM ISBN 978-1-4503-8297-7/21/03.
https://doi.org/10.1145/3437963.3441696

reading preference provided by the recommendation system and guide the user to read the verified news more frequently with some delay. The novel platform has the potential to address the above-mentioned limitations of current delivery methods because 1) it shows no explicit warning label; 2) it encourages readers to learn to detect fake news instead of feeding them with instant verification in the hope—however unrealistic—of correcting their belief; and 3) it incurs a relatively small cost in terms of human effort. In this paper, we demonstrate the effect of the proposed module through a human-subject experiment, in which the module is applied in the context of a news reading platform.

2 RELATED WORK

Most previous studies focus on fact-checking. We divide them by methodology into three strands. The first strand is checking the news using human experts. For example, FactCheck,[2] PolitiFact,[3] and Snopes[4], are famous fact-checking platforms which employ groups of experts to verify news. However, this method is labor-intensive and requires readers to check the fact-checking results on these websites, which means readers are expected to proactively initiate the verification process. The second strand is leveraging crowd-sourcing to label news. Wang et al. [19] exploit the user's report as a weak label, and then use reinforcement learning to select a better label. Yet, as mentioned in Section 1, this method ends up exposing more users to fake news articles in return for gathering more fake news reports from users. The third strand uses machine learning to verify news. Hassan et al. [9] propose ClaimBuster, which utilizes an end-to-end model to identify the truth of news or information. Della Vedova et al. [6] construct a Facebook messenger chatbot that uses machine learning to verify facts. Users check the truthfulness of posts and information from Facebook. Once the chatbot receives input from the user, it provides verified facts. Still, such method relies on users to initiate the verification process. To this end, we propose a system that avoids these disadvantages. For example, our method does not require human effort to label fake news; we guide the user to read verified news instead of directly showing them to the user.

3 SYSTEM DESCRIPTION

Fig. 1 shows the proposed news platform. The details of this framework and the provided functions are described in the following subsections.

3.1 News Platform

We developed a web platform such that users can read up-to-date news articles by interacting with the proposed system. Fig. 2 illustrates the user interface (UI) of the system. We provide a simplified web UI composed of essential factors, including news headlines, news contents, and recommendation lists only for demonstration, but not extraneous elements such as advertisements.

We used Python as the programming language to develop news platform backend. The frontend interface was developed using HTML, javascript, and the CSS framework Bootstrap. To integrate

[2]https://www.factcheck.org/
[3]https://www.politifact.com/
[4]https://www.snopes.com/

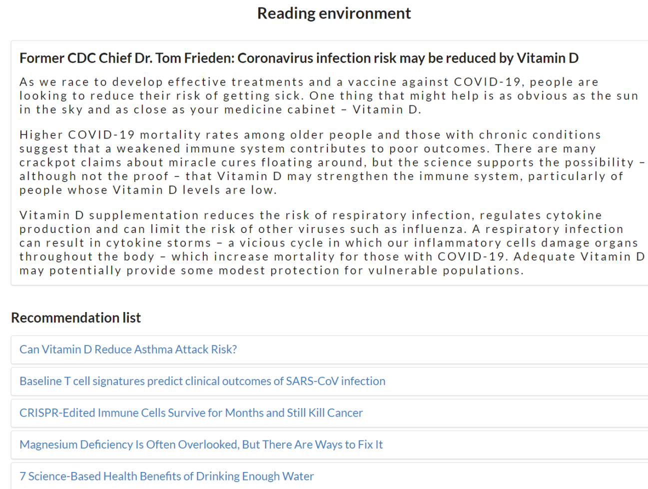


Figure 2: User interface of news platform

the backend and frontend, we used Flask as the framework. MongoDB is then selected to store the news articles and user information. Finally we deployed the platform using Docker and Nginx as the webserver.

3.2 Recommender System (RS)

Since almost all existing news platforms use recommendation algorithms to push news articles according to user preferences, we adopt a content-based recommender system (RS) [3] on the news platform to demonstrate how a RS co-works with the intervention module. The RS recommends news articles to the user according to the user's browsing history and the consumed news corpuses.

3.3 Fake news Intervention Module

We implement the fake news intervention module on top of the RS. We first leverage the result of the RS to construct a graph. The green rectangle in Fig. 1 illustrates the constructed graph, where the intervention module is activated when any candidate fake news article is touched. The working path consists of a predefined number of steps which guide the user to be exposed to as many articles of verified news as possible. Once the intervention module has been triggered, it replaces the RS to provide a recommendation list towards the verified news. Note that the recommendation graph is built based on recommendation results from the RS. Thus, once users' updated news consumption interests are included in the graph, the intervention module reflects their preferences in its recommendation in real time.

For such a guiding process, a model is required to determine which path in the recommendation graph should be selected. We formulate this decision process in each step as the search from a question to its answer, in which the question is the fake news and the answer is the verified news. Drawing from [4, 20], we adopt a reinforcement learning (RL) model to learn how to guide the user from fake news to verified news.

Specifically, once the user clicks a news article, the RS recommendation results are sent to the intervention module as part of the candidates for the next recommendation. Next, the triggered node

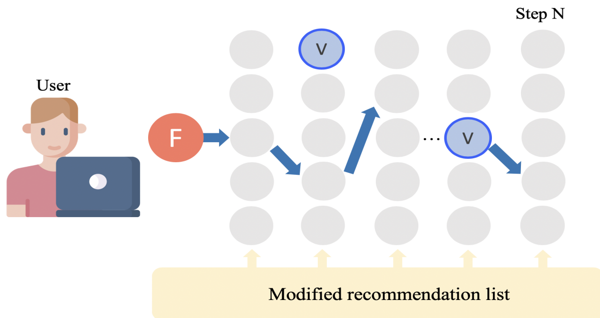


Figure 3: System workflow. The red dot depicts a piece of fake news and the blue dots represent two pieces of verified news related to the fake news. The gray dots show other news articles unrelated to the fake or verified news. The blue arrows illustrate the user's browsing path. N represents the number of reading steps.

and these next-step candidates are aligned to the recommendation graph for the intervention module to take over the recommendation task from the triggered node. Meanwhile, the recommendations suggested by the intervention module in the following steps are presented to the users.

The intervention module is dismissed when it reaches a predefined number of reading steps, e.g., 10 in the proposed platform. Take the user in Fig. 3 as an example. The user triggers the intervention module by clicking the fake news, after which the module begins guiding the user towards the verified news. Note that we do not need to know the exact verified news, as the module learns to reach it to the maximum possibility. From Fig. 3, we see that the user is exposed to the verified news in step 2 without clicking and in step $N-1$ by clicking through. In the recommendation list, users read the titles of the verified news, constituting a higher exposure rate that affords more possibilities to change user beliefs.

The intervention process is completed at the predefined step N, after which control is returned to the RS. Through this process, the influence of fake news is diluted without explicit correction.

3.4 Informed Trajectory Generation

We utilize D3.js[5] to visualize user history. If the viewed news articles are labeled as fake news or verified news, the node is represented by red and blue nodes, respectively. Users browse their viewed news articles in this page by selecting the reading time period. Our informed trajectory generator visualizes the browsing history and statistics about how many and in what order the known fake and verified news articles are read by the user. Thus, users can check their browsing history qualitatively and quantitatively and hence understand their own reading patterns from the aspect of fake and verified news.

4 EVALUATION

To evaluate the effectiveness of the proposed system, we conducted a human-subject experiment with 172 participants recruited on Facebook. Using a within-subject design, we compared the proposed method with a vanilla RS, which means the RS without fake news intervention module, and an RS with warning tags. In each condition, participants' ability to identify fake news was evaluated across three phases: pre-test, post-test, and one day after intervention. In each phase, we presented four pieces of fake and four pieces of real news. Participants were asked to evaluate the veracity of each piece of news article using a four point Likert scale ("1" means "I think this is fake news", "2" means "I am not sure, but I feel something is wrong", "3" means "It seems real, but I am not certain", "4" means "I think this is real news"). We observe the percentage of fake news articles that were not identified in the pre-test but correctly identified in the post-test. Results show that the proposed platform achieves the same rate 29.4% as using the warning tags, which has known limitations.

At the end of each condition, we also asked participants to indicate whether the reading environment is helpful for them to detect fake news using a three-point Likert scale ("1" means "I think the reading environment is not helpful","2" means "I think the reading environment has a little help", "3" means "I think the reading environment is helpful"). After participants completed all three conditions, we asked them to describe the criteria that they used to determine the veracity of the news articles with an open-ended question. Since participants' reading of the news articles was critical to our study, we used an attention check mechanism [10] in each phase. In the end, 89 participants passed all attention check questions.

5 ANALYSIS

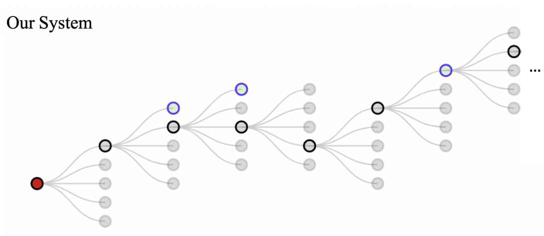
To explore the difference between the proposed platform and a RS, we conducted a coarse-grained analysis from the following three aspects: 1) Are the reading paths changed in the proposed platform? 2) Did participants feel that the three conditions are helpful in mitigating fake news? and 3) Which criteria, e.g., headline, content, or both, were used by participants to evaluate the veracity of fake news?

To understand the difference between the reading path of our system and RS, we picked one of the reading path samples of the proposed system and the RS, respectively. We then visualized each in Fig. 4. From Fig. 4a, we observe that the system with the fake news intervention module guides the user to read the verified news (the light blue dot with the blue circle in the reading path). In contrast, Fig. 4b shows that the RS recommends the related news article initially but other topics later. Thus, users lose their chance to read the verified news article. Furthermore, the result shows that the RS recommends verified news and fake news to the user at the same time, which can backfire, exposing the user to more fake news.

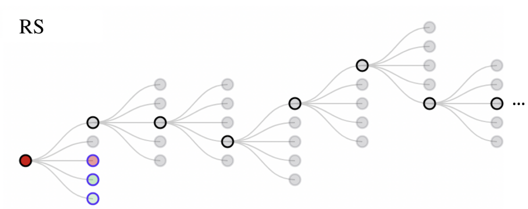
Across the three conditions–the proposed method, the RS, and the RS with warning tags–84% of the participants indicated that our proposed method can help them defeat the fake news, which numerically outperformed both the RS (75%) and the RS with warning tags (80%).

According to the participants' responses to the open-ended question, only eight of them determined the veracity of news articles by using the headlines, thirty participants evaluated the veracity by

[5] https://d3js.org



(a) Sample in proposed system



(b) Sample in RS reading environment

Figure 4: Reading path visualizations. Dark red nodes on the left in (a) and (b) are the starting point: a fake news article. The light red node is a fake news article, and the blue nodes are the verification corresponding to the current fake event. Users are exposed to verified news articles in steps 2, 3, and 6, and the user reads one in step 6 in the proposed system. The RS recommends verified news articles to the user only in step 1, but the user reads neither one.

the content of news articles, and the rest 51 participants judged the veracity by both of the news' content and headline. Such results indicate that most readers infer the news' veracity by reading the content, suggesting that our goal of providing verified news articles to mitigate the impact of misinformation is promising.

Altogether, the results provide preliminary evidence that the proposed system can guide users to read more verified news compared with a RS.

6 CONCLUSION

In this paper, we propose an innovative fake news intervention module and develop a platform to make the module co-work with the recommendation system. We conduct a human-subject experiment measuring the effectiveness of the platform in addressing fake news. The user study show the effectiveness of the module. Moreover, users pointed out the rationale of considering the news content for determining the veracity. In the future we plan to conduct large scale user study where more fake news topics are included.

ACKNOWLEDGEMENT

This research is supported by Ministry of Science and Technology, Taiwan, under Grant no. 109-2221-E-001-015-, and the National Research Foundation, Singapore under its International Research Centres in Singapore Funding Initiative.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31, 2 (2017), 211–36.

[2] Cristian Bravo-Lillo, Lorrie Cranor, Saranga Komanduri, Stuart Schechter, and Manya Sleeper. 2014. Harder to ignore? Revisiting pop-up fatigue and approaches to prevent it. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 105–111.

[3] Chia-Wei Chen, Sheng-Chuan Chou, Chang-You Tai, and Lun-Wei Ku. 2019. Phrase-Guided Attention Web Article Recommendation for Next Clicks and Views. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 315–324.

[4] Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alexander J. Smola, and Andrew McCallum. 2018. Go for a Walk and Arrive at the Answer: Reasoning Over Paths in Knowledge Bases using Reinforcement Learning. *CoRR* abs/1711.05851 (2018). arXiv:1711.05851 http://arxiv.org/abs/1711.05851

[5] Madeleine de Cock Buning. 2018. *A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation*. Publications Office of the European Union.

[6] Marco L Della Vedova, Eugenio Tacchini, Stefano Moret, Gabriele Ballarin, Massimo DiPierro, and Luca de Alfaro. 2018. Automatic online fake news detection combining content and social signals. In *2018 22nd Conference of Open Innovations Association (FRUCT)*. IEEE, 272–279.

[7] Mingkun Gao, Ziang Xiao, Karrie Karahalios, and Wai-Tat Fu. 2018. To label or not to label: The effect of stance and credibility labels on readers' selection and perception of news articles. *ACM CHI* 2, CSCW (2018).

[8] R Kelly Garrett and Brian E Weeks. 2013. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. 1047–1058.

[9] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. 2017. ClaimBuster: The First-Ever End-to-End Fact-Checking System. *Proc. VLDB Endow.* 10, 12 (Aug. 2017), 1945–1948. https://doi.org/10.14778/3137765.3137815

[10] David J Hauser and Norbert Schwarz. 2016. Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods* 48, 1 (2016), 400–407.

[11] Jooyeon Kim, Behzad Tabibian, Alice Oh, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 2018. Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (Marina Del Rey, CA, USA) *(WSDM '18)*. Association for Computing Machinery, New York, NY, USA, 324–332. https://doi.org/10.1145/3159652.3159734

[12] Adam Kucharski. 2016. Study epidemiology of fake news. *Nature* 540, 7634 (2016), 525–525.

[13] Gordon Pennycook and Tyrone Cannon. 2018. Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology General* (06 2018). https://doi.org/10.2139/ssrn.2958246

[14] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. 2020. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science* 31, 7 (2020), 770–780.

[15] Gordon Pennycook and David G Rand. 2019. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* 116, 7 (2019), 2521–2526.

[16] Haeseung Seo, Aiping Xiong, and Dongwon Lee. 2019. Trust It or Not: Effects of Machine-Learning Warnings in Helping Individuals Mitigate Misinformation. In *Proceedings of the 10th ACM Conference on Web Science* (Boston, Massachusetts, USA) *(WebSci '19)*. ACM, New York, NY, USA, 265–274. https://doi.org/10.1145/3292522.3326012

[17] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 3 (2019), 1–42.

[18] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.* 19, 1 (Sept. 2017), 22–36. https://doi.org/10.1145/3137597.3137600

[19] Yaqing Wang, Weifeng Yang, Fenglong Ma, Jin Xu, Bin Zhong, Qiang Deng, and Jing Gao. 2020. Weak Supervision for Fake News Detection via Reinforcement Learning. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 516–523. https://aaai.org/ojs/index.php/AAAI/article/view/5389

[20] Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. 2019. Reinforcement Knowledge Graph Reasoning for Explainable Recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 285–294. https://doi.org/10.1145/3331184.3331203