Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation

Guan-zhen GUO

May 22, 2018

Partners: Sai-sai LU
Instructor: Professor Luo-yi FU

Abstract

Online social media is trying to reduce the spread of fake news through a crowd-powered procedure: when a user considers a post as fake, he can report it through his account. When the post receives a certain number of reports, it will be sent to an authoritative third-party organization for fact checking. Once the post is judged to be fake, it will be stopped spreading immediately and marked as controversial.

The previous works are mainly based on the side of social media to consider which posts to fact check and when to do so can reduce misinformation. On the other hand, in this report, we mainly stand on the position of the third-party organizations, combining the queuing theory in the data network to consider which posts should be detected first to minimize the damage degree of misinformation. According to the above ideas, we propose two schemes with the corresponding algorithm and experimental results based on the dataset gathered from Weibo.

1 Introduction

Fake news and misinformation have dominated the news media around the world since the US presidential election (2016). In recent years, due to the popularity of social media, the spread of fake news is gradually out of control. Users cannot distinguish whether the posts on the online social network are true or not, becoming the victim of misinformation.

In order to curb the spread of misinformation, major online social networking sites, such as Facebook, Twitter and Weibo, are committed to developing a series of measures. In addition to machine learning methods, it is more reliable to use crowd-powered process: when a user considers a post as fake, he can flag it as misinformation through his account. When the post receives a certain number of flags, it will be sent to an authoritative third-party organization for

fact checking. Once the post is identified to be fake, it will be stopped spreading immediately and marked as controversial.

1.1 Related Work

Some of the previous work have used crowd signals to detect fake news in social networks and further reduce misinformation. Two relatively new and representative algorithms are introduced below:

CURB: Kim et al. [1] came up with the process of detecting fake news by leveraging users flagging activity. In particular, they introduce a flexible representation of the above problem using the framework of marked temporal point processes. They develop an algorithm, CURB, to select which posts to fact check and when to do so via solving a novel stochastic optimal control problem.

DETECTIVE: Tschiatschek et al. [2] design a algorithm DETECTIVE, which implements a Bayesian approach for learning about users accuracies over time as well as for performing inference to find which news are fake with high confidence. Their algorithm actively trades off between exploitation (selecting news that directly maximize the objective value) and exploration (selecting news that helps towards learning about users flagging accuracy).

2 Problem Formulation

2.1 Target

As previous work [1][2], our goal is to minimize the spread of misinformation, i.e., how many users end up seeing a fake news before it is blocked. The difference is that they start from the perspective of social media, deciding which should send for fact checking from numerous posts. However, our proposal is mainly based on the viewpoint of the third-party organization, to rank lots of the reports that are sent to fact check, determine the order to accept judgement and minimize the spread of fake news.

2.2 Data Representation

In order to effectively and precisely reduce the harm of misinformation, we take Sina Weibo for example, proposing two schemes to predict the potential hazards of the posts and the credibility of the whistleblowers at two different viewpoints.

SCHEME 1 Sort by the number of visits. Considering that the propagation speed of each post can not be known in advance, we propose SCHEME 1 to predict the potential harm degree of posts. If the number of visits is higher, it represents that the post has affected more users so that the potential

harm is greater. Therefore, examining the posts with high number of visits first can effectively reduce misinformation.

In SCHEME 1, we represent each event as a 2-tuple

$$e_1 := (n_0, c) \tag{1}$$

where n_0 is the number of events in the order of reporting, and $c(t)$ represents the number of visits, which is a function of time t.

SCHEME 2 Sort by the proportion of verified users in whistleblowers. Since Weibo mark those who go through identification as verified, it is believed that their credibility is higher. As the result, the higher the proportion of verified users in whistleblowers, the greater the probability that the posts will be fake. Therefore, examining the posts with high proportion of verified users in whistleblowers first can effectively reduce misinformation.

In SCHEME 2, we also represent each event as a 2-tuple

$$e_2 := (n_0, v) \tag{2}$$

where n_0 is also the number of events in the order of reporting, and $v(t)$ represents the proportion of verified users in whistleblowers, which is a function of time t.

2.3 Queueing Theory

In SCHEME 1 and 2, the number of visits and the proportion of verified users in whistleblowers vary as the increase of waiting time. In order to define the waiting time for fact checking, we introduced queuing theory in data networks [3]. We consider the non-preemptive priority rule in M/G/1 system, whereby a customer undergoing service is allowed to complete service without interruption even if a customer of higher priority arrives in the meantime. When the server becomes free, the first customer of the highest nonempty priority queue enters service. For priority class $1, 2, \ldots, n$, denote

$$\lambda_k = Arrival\ rate\ of\ class\ k$$

$$\overline{X_k} = Average\ service\ time\ of\ class\ k$$

$$\overline{X_k^2} = Second\ moment\ of\ the\ service\ time\ of\ class\ k$$

$$W_k = Average\ queueing\ time\ for\ class\ k$$

$$\rho_k = \lambda_k/\mu_k = System\ utilization\ for\ class\ k$$

$$R = Mean\ residual\ service\ time$$

The formula for the waiting time in queue is

$$W_k = \frac{R}{(1-\rho_1-\cdots-\rho_{k-1})(1-\rho_1-\cdots-\rho_k)} \tag{3}$$

The mean residual service time R can be derived as for the P-K formula

$$R = \frac{1}{2}\sum_{i=1}^{n}\lambda_i \overline{X_i^2} \tag{4}$$

2.4 The Model

According to the data presentation and queuing theory mentioned above, we can introduce mathematical model of SCHEME 1 and 2 respectively.

SCHEME 1 For posts $i = 1, 2, \ldots, n$, $c_i(t)$ indicate the number of visits at time t, and W_i represents the waiting time in the queue to be checked. Therefore, the model of SCHEME 1 is as follow:

$$\min \sum_{i=1}^{n} c_i(W_i) \tag{5}$$

SCHEME 2 For posts $i = 1, 2, \ldots, n$, $v_i(t)$ indicate the proportion of verified users in whistleblowers at time t, and W_i still represents the waiting time in the queue to be checked. Therefore, the model of SCHEME 2 is as follow:

$$\min \sum_{i=1}^{n} v_i(W_i) \tag{6}$$

3 Proposed Algorithm

Algorithm 1: SCHEME 1

Input:
the number of events in the order of reporting n_0
system time t
the number of visits $c(t)$
the time to check an event T
Output:
the number of events in the order of fact checking n

Initialization: $n_0 \leftarrow 0$; $t \leftarrow 0$; $c(t) \leftarrow 0$;
Update $e_1 := (n_0, c)$;
for $k = 1$ *to* n_0 **do**
    Update $c : c \leftarrow c(t)$;
    Direct insertion sort according to the value of $v(t)$;
    $n[k] \leftarrow n_0[1]$;
    delete $e_1[1]$;
    Update $t : t \leftarrow t + T$;
    Update $e_1 : e_1 \leftarrow e_1(t)$;
end
return n

Algorithm 2: SCHEME 2

Input:
the number of events in the order of reporting n_0
system time t
the proportion of verified users in whistleblowers $v(t)$
the time to check an event T
Output:
the number of events in the order of fact checking n

Initialization: $n_0 \leftarrow 0$; $t \leftarrow 0$; $v(t) \leftarrow 0$;
Update $e_2 := (n_0, v)$;
for $k = 1$ *to* n_0 **do**
    Update $v : v \leftarrow v(t)$;
    Direct insertion sort according to the value of $v(t)$;
    $n[k] \leftarrow n_0[1]$;
    delete $e_2[1]$;
    Update $t : t \leftarrow t + T$;
    Update $e_2 : e_2 \leftarrow e_2(t)$;
end
return n

When refreshing the number of visits or the proportion of verified users, the order varies little. Since direct insertion sort algorithm performs better when the data is nearly in order, we use it in above algorithm.

4 Experiments

4.1 Experimental Setup

Our data are collected from Weibo community management center. Among them, we chose the 200 reports of misinformation until May 9th. Since the website sends the post for fact checking as long as there is one user report it, we follow this setting in SCHEME 1 and 2. As a result, SCHEME 2 can only be divided into two cases that the whistleblower is verified user or not. So we also considered the number of the whistleblower's followers, and assumed that the more the number of followers he has, the greater the credibility of the user. For convenience, we assume that all posts enter the queue at the same time $t = 0$.

4.2 Experimental Results

The experimental results of SCHEME 1 and 2 are shown in Figure 1 and 2 respectively. Among them, the ordinate of Figure 1 is the number of visits while the ordinate of Figure 2 is the number of the whistleblower's followers. The original method (FIFO) is introduced for comparison.

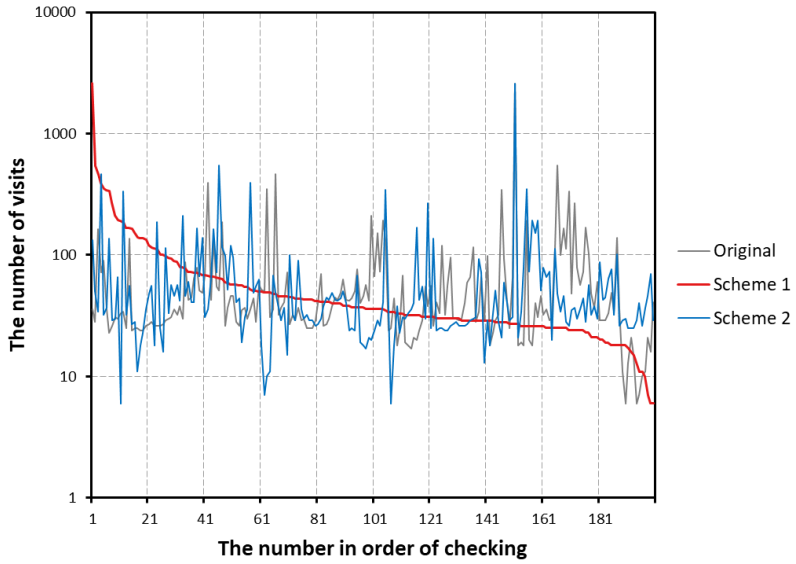

Figure 1: Set the ordinate as the number of visits

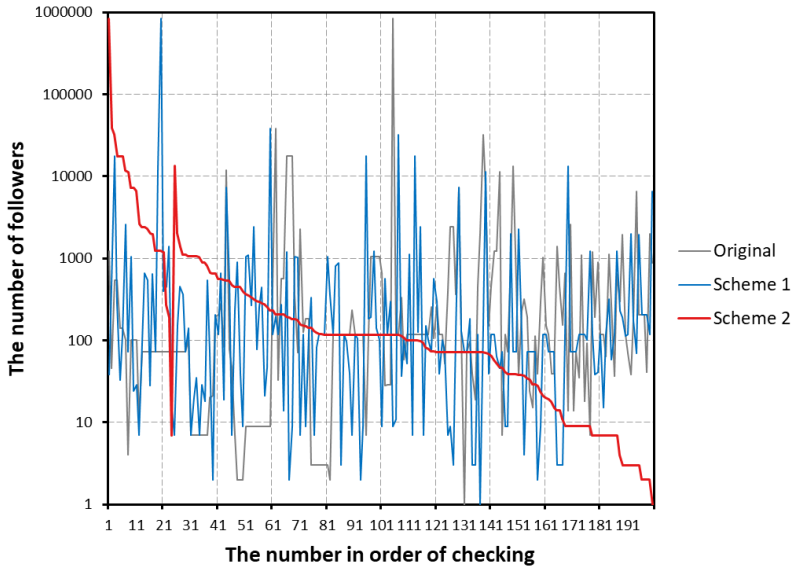

Figure 2: Set the ordinate as the number of the whistleblower's followers

As shown in Figure 1, the fact checking order of SCHEME 1 is the number of visits arranged from high to low, while that of the original method and SCHEME 2 are independent of the number of visits. In Figure 2, the former section of the curve in SCHEME 2 is the situation that the whistleblower is verified, and the latter is the case that the whistleblower is not verified. In the two section, the fact checking order is arranged according to the number of the whistleblower's followers, while the fact checking order of the original method and SCHEME 1 have nothing to do with the number of the whistleblower's followers.

5 Conclusions

In this report, we propose a method to reduce misinformation in social media based on third-party organization. According to the queuing theory in the data network, we come up with two schemes, which estimate the potential harm of the post with the number of visits and the proportion of verified users in whistleblowers, and give the corresponding algorithms. Finally, we use the data gathered from Weibo to carry out experiments on the two schemes, and find that the results of the two schemes have their own advantages and disadvantages.

In the future work, we can consider merging these two schemes and determining their weights so as to minimize misinformation. In addition, in order to comply with the setup of Weibo, we assume in the experiment that a post sent for fact checking as long as there is one user report it. However, due to the high cost of fact checking, it should be set that a post is sent for fact checking only if there is n users report it ($n > 1$). Finally, in our experiment, we assume that all posts enter the checking queue at the same time, which can be improved in future work, such as setting that a post enter the queue according to the Poisson distribution.

References

[1] Jooyeon Kim, Behzad Tabibian, Alice Oh, Bernhard Schölkopf, and Manuel Gomez-Rodriguez. 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*, pages 324–332. ACM, 2018.

[2] Sebastian Tschiatschek, Adish Singla, Manuel Gomez Rodriguez, Arpit Merchant, and Andreas Krause. Fake news detection in social networks via crowd signals. In *Companion of the The Web Conference 2018 on The Web Conference 2018*, pages 517–524. International World Wide Web Conferences Steering Committee, 2018.

[3] Dimitri P Bertsekas and Robert G Gallager. *Data networks*, volume 2.