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

卢赛赛　515030910081

1. INTRODUCTION:

In recent years, social media and online social networking sites have become a major disseminator of false facts, urban legends, fake news, or, more generally, misinformation. There are various motivations for generating and spreading fake news, for instance, making political gains, harming the reputation of businesses, as clickbait for increasing advertising revenue, and for seeking attention. In this context, there are growing concerns that misinformation on these platforms has fueled the emergence of a post-truth society, where debate is perniciously framed by the repeated assertion of talking points to which factual rebuttals by the media or independent experts are ignored.

Social media sites and online social networks, for example Facebook and Twitter, have faced scrutiny for being unable to curb the spread of fake news. In an effort to curb the spread of misinformation, major online social networking sites, such as Facebook, Twitter, Weibo or Wechat, are (considering) resorting to the crowd. In particular, 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.

2. RELATED WORK

2.1. CURB[1]: 1)、they consider all users to be equally reliable and estimate the flagging accuracy of the population of users from historical data. 2)、they model the actual propagation dynamics as a continuous-time dynamical system with jumps and arrive at an algorithm by casting the problem as an optimal control problem. 3)、they use continuous time and consider an overall budget for their algorithm.

2.2. DETECTIVE[2]: 1)、they learn about the flagging accuracy of individual users in an online setting; 2)、their algorithms are agnostic to the actual propagation dynamics of news in the network. 3)、they use discrete epochs with a fixed budget per epoch (i.e., the number of news that can be sent to an expert for reviewing);

2.3. Our approach: 1)、The same points are that our target to minimize the spread of misinformation and the crowd-powered procedure. 2)、The first difference point is that their methods is from the perspective of social media, while our method is from the perspective of the third-party organization.3)、In addition, we adopt different measures. They decide which post should send for fact checking, while we rank the reports sent for fact checking in the order of their importance so that we can decide which post to accept judgement first.

3. METHOD

3.1. Scheme 1: It’ s to sort the posts by the number of visits, that is how many users have read the posts. By the process, we can predict the potential harm degree of the posts. If the number of visits is higher, there is more users affected by the post, so that the potential harm is higher. Therefore, checking the posts with high number of visits first can effectively reduce misinformation.

$$e_1 \coloneqq (n_0,\ c)$$

the number of events in the order of reporting

$c(t)$ represents the number of visits, which is a function of time t

3.2. Scheme 2: It’ s to sort the posts by the proportion of verified users in whistleblowers. The whistleblowers mean the users who report a post. As the result, we can assume that the verified users have higher credibility. Therefore, if the proportion of verified users in whistleblowers is higher, the posts are more likely to be fake. Therefore, checking the posts with high proportion of verified users in whistleblowers first can effectively reduce misinformation.

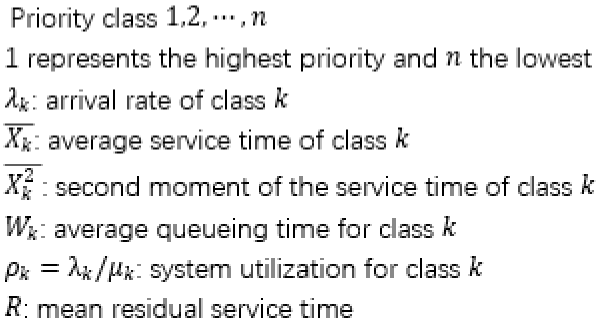
$$e_2 \coloneqq (n_0,\ v)$$

the number of events in the order of reporting

$v(t)$ represents the proportion of veried users in whistleblowers, which is a function of time t

4. MODEL

4.1. Queueing theory: We consider the non-preemptive priority rule in M/G/1 system. When the server becomes free, the first customer of the highest nonempty priority queue enters service.



Priority class $1,2,\cdots,n$

1 represents the highest priority and n the lowest

λ_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:

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

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

$$\mathrm{R} = \frac{1}{2}\sum_{i=1}^{n}\lambda_i\overline{X_\iota}$$

4.2. Scheme 1: For posts i=1, 2, ⋯, n, c_i (t) indicates 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:

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

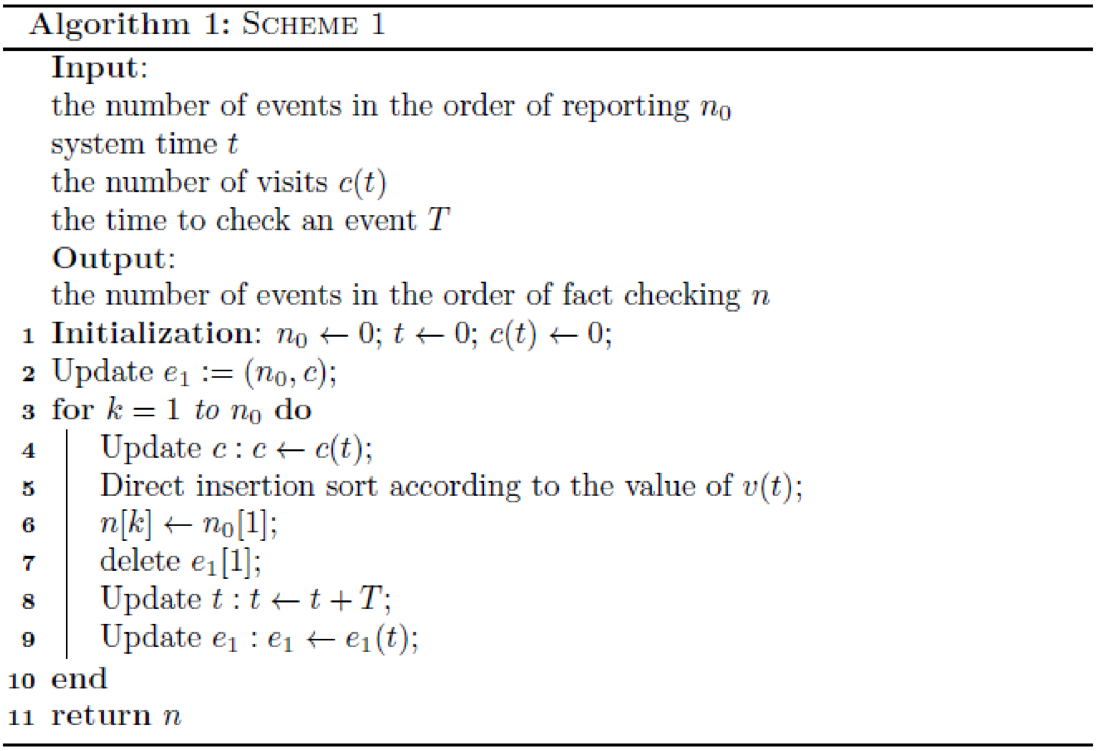
4.3. Scheme 2: For posts i=1, 2, ⋯, n, v_i (t) indicates the proportion of verified users in whistleblowers at time t, and W_i represents the waiting time in the queue to be checked. Therefore, the model of Scheme 2 is as follow:

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

5. ALGORITHM

5.1. Algorithm 1:

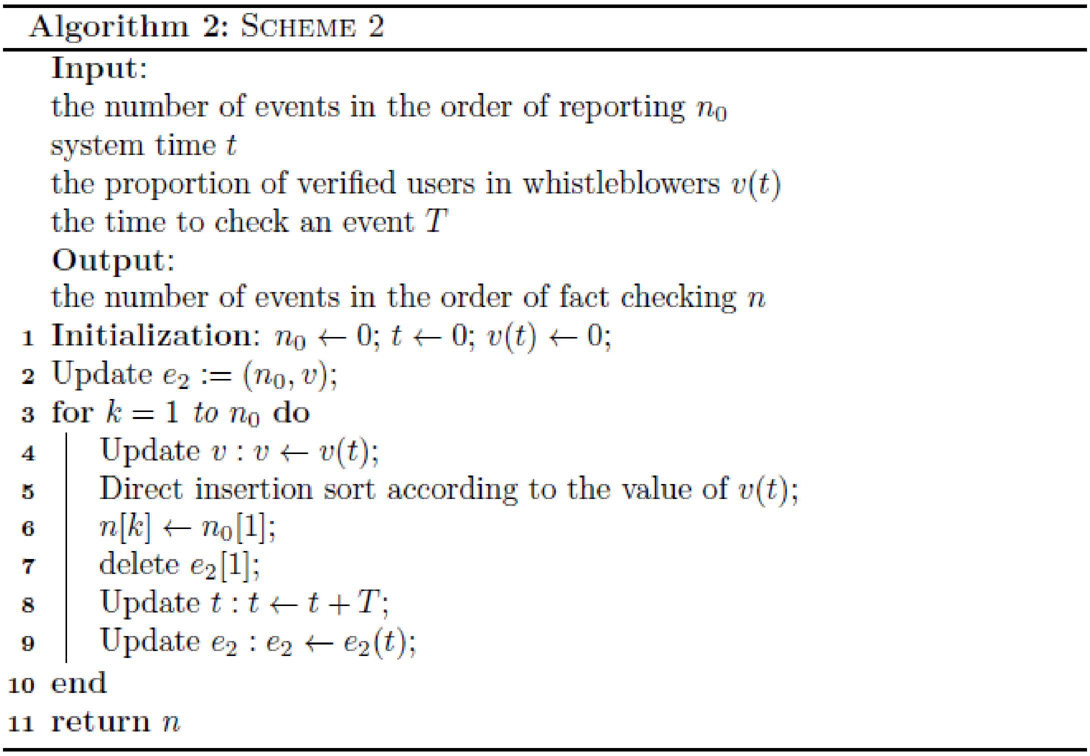
We first update the number of visits and prioritize according to the number of visits to user stories. Then, the story of high priority is output. Then update the story for a new round of processing.



```
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 $n$
```

5.2. Algorithm 2:

We first prioritize users based on the proportion of V users, and then process the highest priority stories and output them. then, Update the story for a new round of processing.



```
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 $n$
```

6. EXPERIMENTS

6.1. Algorithm 1: we think that the same reported incidents are more frequent in a short time, and that the impact can be reduced as early as possible, therefore, we have proposed the first scheme. As you can see from the Figure 1a, compared with the scheme two and the scheme that the stories reported are not processed, the scheme one gives priority to the story of more frequent visits, which reduces the impact of diffusion.

6.2. Algorithm 2:

we think that users with V have higher reliability, thus the story is more likely to be false and should be dealt with as soon as possible. Figure 1b shows that the same reported events, with the high credibility of the plus V users, will have a greater impact on the false message as soon as possible, and scheme two can reduce the impact. Scheme two is mainly for fear that some people are reporting in disorder. In fact, this news is true. But if a lot of plus V users think it is fake, it is more likely to be misled.

Therefore, the consideration of the scheme one is the diffusion rate. the consideration of the scheme two is whether the story is really misleading.

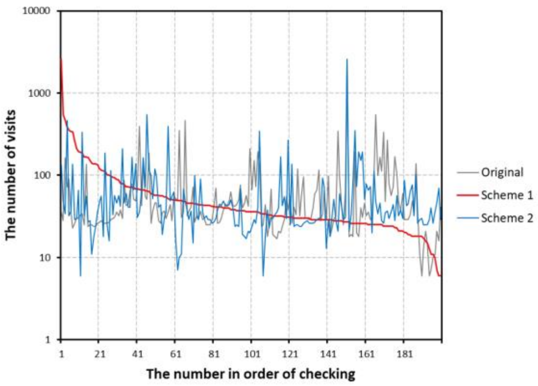


Figure 1a

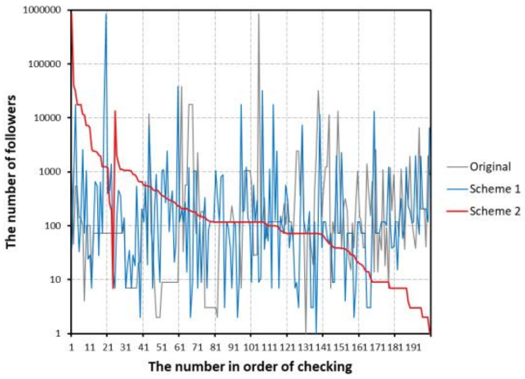


Figure 1b

7. Conclusions

In this paper, I have introduced two algorithm, that leverages the crowd to detect and prevent the spread of misinformation in online social networking sites. I experimented with one real-world dataset gathered from Weibo and showed that our algorithm may be able to effectively reduce the effect and spread of misinformation.

There are many interesting directions for future work. At first, we planned to update the database to obtain the real time data which can improve the practicality of our scheme, but we don’ t realize it. In addition, the priority should be further improved: when the number of data reaches a certain value (for example, 100), the detection is performed. Among them, plus V users can have more reporting weights according to their V level. That it’ s to say, an ordinary user represents one ticket while the plus V user represents two or more tickets. We considered that stories are independent and the probability that a story is misinformation given that a user did(not) flag a story is equal for all stories.

References

[1] Jooyeon Kim∗1, Behzad Tabibian2,3, Alice Oh1, Bernhard Schölkopf2, and Manuel Gomez-Rodriguez3: *Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation*

[2] Sebastian Tschiatschek∗, Adish Singla, Manuel Gomez Rodriguez: *Fake News Detection in Social Networks via Crowd Signals*

[3] Liang Wu, Huan Liu: *Tracing Fake-News Footprints: Characterizing Social Media Messages by How They Propagate*