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

Anonymous Author(s)
Affiliation
Address
`email`

Xintian Yu
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University

Abstract

Today, with more and more advanced microblogging platforms used, fake news and rumors can spread very fast ,so locating and debunking them has become a crucial problem. Existing works to detect and restrict rumors include using RNN to analyze the text, topic-aware source locating algorithms, scalable online curbing algorithm, etc. In this report, we try to combine them to detect and stop the propagation of fake news and rumors from the source.

1 Introduction

In recent years, social media and online social networking sites have become a major source of false facts, rumors, or misinformation. In this report, a misinformation or rumor is defined as a story or a statement whose truth value is unverified or deliberately false. Examples are like "Eating seafoods with fruits at the same time will cause poisonous arsenic trioxide produced in your body", which was once very popular among the crowd. Or is like "Earthquake will happen in northern China on 2010.6.16", which had a bad influence among the crowd. This incident of a false rumor highlights that automatically predicting the veracity of information on social media is of high practical value. In order to distinguish misinformation and facts, individuals and organizations often rely on common sense and investigation news. Sites like Snopes.com and Factcheck.org[3] are such collaborative efforts. However, since these efforts need many manual verification steps, these websites are not comprehensive and misinformation may be exposed for a long time.

Currently we have much existing work to detect misinformation, debunk them or stop them from propagating among the crowd. Such as using topic-aware SI model [1] to locate the source of misinformation, CURB algorithm to reduce the spread of misinformation [2] ,or using RNN-based algorithm to locate and debunk them [3] .Usually, debunking rumors at an early stage of diffusion is particularly crucial to minimizing their harmful effects.

Each method has its own pros and cons. To combine these methods and their advantages, the needed work is divided into 2 parts:

1. Use crowd-powered fact checking procedure to detect whether the information is a truth or a rumor. One specified method is: Users can mark any story in their feed as wrong information, and if the story receives enough marks, it will be sent to a third party for fact checking. If a third party identifies the story as fake or misinformation, it will be flagged as controversial and may appear lower in the user's subscription.[2]

Submitted to 30th Conference on Neural Information Processing Systems (NIPS 2016). Do not distribute.

2. Locate the source of the story in social network, and stop them from propagating from the source. We may use topic-aware analysis and semantic analysis in this step. Since people prefer to spread the information they are interested in, the information on different topic trends is usually spread along different paths. Therefore, the introduction of topic-aware factors into the source location problem improves the accuracy of the detection, which makes source location in topic-aware social networks a practical tool.[1]

2 Proposed Algorithms

We are here to introduce the algorithms that are used to solve the problems.

2.1 CURB Algorithm

CURB Algorithm[2] is an extensible online algorithm that chooses which story to send for fact checking and when to do so to effectively reduce the spread of false information and provide verifiable guarantees. As is mentioned in part 1, online social networking sites are providing millions of users with suggestions on how to find the wrong information online. However, the above solution process requires careful reasoning and intelligent algorithms. As far as we know, it does not exist yet. So we need to decide which story to send for fact checking, and when to do so.

To simplify the explanation, we first deduced a story solution, introduced an efficient algorithm to implement the solution, and then generalized the solution and efficient algorithm to multiple stories. Then we define an optimal cost function J for the problem, use Bellman's principle of optimality to derive the associated Hamilton-Jacobi-Bellman (HJB) equation[2,4]. Then solve the equation to minimize J. The algorithm is as follows, to specify the meanings of each parameter, please refer to [2].

Algorithm 1 CURB

Input: Parameters $q, \alpha, \beta, p(m|f=1), p(m|f=0), t_f$
Initialization: $N(t) \leftarrow 0, N^f(t) \leftarrow 0, \lambda^e(t) \leftarrow 0$, Update $N^e(t), N^f(t), \lambda^e(t)$
Output: Fact checking time τ

```
τ ← t_f
(t', r, f) ← Next()
while t' < τ do do
    u_0(t) ← u(t)
    N^e(t) ←= N^e(t) + 1; N^f(t) ← N^f(t) + f
    u(t) ← Update(N^e(t), N^f(t), λ^e(t))
    if f = 0 then then
        x ← Uniform(0, 1)
        if u(t)/u_0(t) < x then then
            τ ← Sample(τ, u(t))
        end if
    end if
    if r = 1 then then
        λ^e(t) ← λ^e(t) + g(t − t')
        u(t) ← Update(N^e(t), N^f(t), λ^e(t))
    end if
    κ = Sample(t', max(0, u(t) − u_0(t)))
    τ ← min(τ, κ)
    (t', r, f) ← Next()
end while
return τ
```

2.2 Sample Path Based Solution

Then we need to find the source of misinformation, which needs sample path based heuristic algorithm to simulate the most likely source that triggers the existing infected subgraph. Also, this algorithm is based on the topic-aware SI model (T-SI) referring to [1].

Algorithm 2 Sample Path Based Solution to Source Locating

Initialization: all user set V which contains infected users set I, set $STOP = 0$, infection time $t = 0$
for $v \in V$ do **do**
  for $u \in I$ do **do**
    $t_v^u = N, flag_v^u = 0$
  end for
end for
for $u \in I$ do **do**
  u spreading its ID to its neighbors following the T-SI model
end for
while $STOP = 0$ do **do**
  for $v \in V$ do **do**
    if v received the ID of u for the first time then **then**
      $t_v^u = t, flag_v^u = 1$
      v spreading the ID of u to its neighbors
    end if
    if $flag_v^u = 1$ for all $u \in I$ then **then**
      $STOP = 1$
    end if
  end for
end while
return $\hat{v} = argmin_{v \in A} \sum_{u \in I} t_v^u$ where A is the set of users with $flag_a^i = 1$ when algorithm terminates.

3 Experiments

We have run the algorithm on the datasets that are mentioned in the paper[1][2]. For CURB algorithm, we use data gathered from Twitter and Weibo as reported in previous work[3]. The Twitter dataset contains 28486 posts and reshares from 18880 users for 46 unique stories, which include 7 fake stories and 39 genuine ones. The Weibo dataset contains about 93943 posts and reshares from 88913 users for 156 unique stories, which include 23 fake stories and 133 genuine stories. we set $\gamma = 10^{-4}$ and $\omega = 10^{-5}$ in parameters. This choice of parameters results in approximately 10 to 20 exposures per post (or reshare) and a half-life of approximately 19 hours per post (or reshare). We also have obtained similarly qualified results from other choices of parameters. Here are the results.

p(f=1 | m=1)	0.5	0.7	0.9
Precision	0.214	0.291	0.388

Then, we also have run Sample Path on the real-world datasets, Author collaboration[5]. The result conforms to the experimental results in [1].

Error Distance	0.0	1.0	2.0
Cumulative probability	0.18	0.75	0.97

4 Future Work

Based on what we have done, I'm listing them separately.

1. Try to improve the precision of CURB algorithm on larger datasets. As is introduced, the current algorithm precision is very low, less than 0.5.

2. Try to use RNN-based models to solve the similar problem, and fix our eyes on semantic analysis to grab the more obvious features of rumors or misinformation.

3. Try to get more detailed result and visualize them to compare them with existing methods.

5 Conclusion

In this project, we introduce the CURB algorithm and Sample Path algorithms ,run them on different datasets, get good experiment results, and compared the results to combine their advantages. Though there is still much to do, we still learned much in the project. For example: how to read papers, how to do research. Thanks to the teachers and professors standing by.

References

[1] Wenyu Zang, Peng Zhang, Chuan Zhou and Li Guo (2015) Topic-aware Source Locating in Social Networks Cambridge, Proceedings of the 24th International Conference on World Wide Web, Pages 141-142.

[2] Jooyeon Kim et al. (2018) Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation, WSDM'18, Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Pages 324-332.

[3] Jing Ma et al. (2016) Detecting Rumors from Microblogs with Recurrent Neural Networks, IJCAI'16, Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, Pages 3818-3824.

[4] A. Borodin, G. O. Roberts, J. S. Rosenthal, and P. Tsaparas. (2005). Link analysisranking: algorithms, theory, and experiments. TOIT 5,1(2005), Pages 231–297.

[5]W. Dong, W. Zhang, and C. W. Tan. (2013) Rooting out the rumor culprit from suspects. In Proceedings of IEEE International Symposium on Information Theory (ISIT), pages 2671–2675.