Information Verification during Natural Disasters

Abdulfatai Popoola,
Dmytro Krasnoshtan,
Attila-Peter Toth
Masdar Institute of Science
and Technology
Abu Dhabi, UAE

Victor Naroditskiy
Electronics and Computer
Science
University of Southampton
Southampton, United Kingdom

Carlos Castillo,
Patrick Meier
Qatar Computing Research
Institute
Doha, Qatar

Iyad Rahwan
Masdar Institute of Science
and Technology
Abu Dhabi, UAE

ABSTRACT

Large amounts of unverified and at times contradictory information often appear on social media following natural disasters. Timely verification of this information can be crucial to saving lives and for coordinating relief efforts. Our goal is to enable this verification by developing an online platform that involves ordinary citizens in the evidence gathering and evaluation process. The output of this platform will provide reliable information to humanitarian organizations, journalists, and decision makers involved in relief efforts.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human information processing*

General Terms

Human Factors

Keywords

Crowdsourcing, Data Verification

1. INTRODUCTION

Emergency relief is not possible without information about the affected areas. Recent advances in Information and Communication Technologies enabled the rapid collection of relevant information from people living in and near disaster areas. However, as powerful as the social web and mobile technologies have been in collecting disaster-related information, their usefulness for disaster response depends on the quality of the collected information.

Reliability and authenticity issues are the single biggest challenge for the use of social media by media organizations and humanitarian agencies. While disaster-affected populations are increasingly the source of vital user-generated content, many humanitarian agencies are hesitant to fully leverage this important source due to concerns over these issues. The same is true for major media companies and newsrooms.

Several methods have been developed to verify social media content generated during humanitarian crises. The British Broadcasting Corporation's (BBC) User-Generated Content (UGC) Hub in London focuses on contrasting sources and interviewing individuals who provide this content. They have also developed a series of metrics to ascertain the credibility of user-generated content. For example, they consider the number of followers that a given Twitter user has, whether they are followed by any reputable sources, what the Twitter user in question has tweeted previously and how long that Twitter account has been active for. *Storyful*, a Twitter-based news company, focuses on verifying the authenticity of photographic evidence and YouTube videos by considering the time of day (shadows), weather and accents spoken as well as any landmarks that they can identify and confirm through other sources, such as Google Earth.

Existing approaches require in-house resources for processing information, and cannot keep up with the volume of data generated during natural disasters. Furthermore, there are no readily available ways to share reliability findings among different parties working independently.

We thus propose a platform that overcomes these limitations by opening evidence collection to the public, targeting the people who are most likely to possess relevant evidence, and providing tools for evidence sharing, evaluation, and prioritization. This is line with the growing usage of *crowdsourcing* for information quality tasks, including methods to further verify and validate the assessments provided by crowdsourcing workers [6].

Ordinary citizens participating in the proposed **Verily**[1] platform will produce a body of evidence for the use by relief participants. The platform will enable humanitarian and governmental organizations to quickly evaluate which information meets their standards of trustworthiness based on the evidence presented. This will lead to a faster and more reliable integration of crucial disaster information into coordination of relief efforts.

[1]http://www.veri.ly/

2. THE RATIONALE: TIME-CRITICAL SOCIAL MOBILIZATION

The Economist recently published an insightful article[2] entitled "Six Degrees of Mobilisation: To what extent can social networking make it easier to find people and solve real-world problems?" The notion, six degrees of separation, comes from Stanley Milgram's experiment in the 1960s which found that there were, on average, six degrees of separation between any two people in the US [7]. Last year, Facebook found that users on the social network were separated by an average of 4.7 hops. The Economist thus asks the following, fascinating question:

> Can this be used to solve real-world problems, by taking advantage of the talents and connections of one's friends, and their friends? That is the aim of a new field known as social mobilization, which treats the population as a distributed knowledge resource which can be tapped using modern technology.

The article specifically refers to DARPA's Red Balloon Challenge [12]. Ten large weather balloons were secretly positioned across the continental United States. The challenge was to be the first to identify the exact locations of these balloons in order to win the $40,000 prize. The winning team from MIT found all ten weather balloons in just 8 hours and 36 minutes without ever leaving their laptops. How did they pull this off?

They won by using social media, crowdsourcing and a technique they refer to as a recursive incentive mechanism. The team recruited thousands of volunteers using social media and told them that they would be financially rewarded for finding the correct location of all ten balloons before any other group. In other words, they pledged to give away the $40K prize they would receive if their volunteers found the ten weather balloons.

They promised $2,000 per balloon to the first person to find the correct coordinates, $1,000 to the person who recruited that balloon finder to the team, $500 to whoever invited the inviter, $250 to whoever invited that person, and so on. This is the recursive incentive mechanism at work. Note that teams from other universities deliberately tried to sabotage MIT's efforts by planting false leads on the balloons. Still, the Boston team managed to work around the misinformation campaigns to win the challenge in under 9 hours.

Similar to information gathering in the Red Balloon Challenge, information verification requires the involvement of a large number of people. Making sense of large amounts of fast-changing and contradictory information about a broad geographic area can only be accomplished if enough participants are mobilized. It also calls for participation of people who have physical access to supporting evidence. As we describe in the next section, our platform, called Verily, will encourage mobilization of the people who are geographically close to the disaster and who are qualified to evaluate evidence. That is, instead of finding the location of weather balloons, we will use time-critical social mobilization to crowdsource the collection of and evaluation of evidence in order to determine whether certain claims are true.

3. VERILY: A VERIFICATION PLATFORM

Verily is a web-based platform designed for rapid collection and assessment of information generated during natural disasters (e.g., social media reports from New Yorkers during hurricane Sandy and reports posted on the Ushahidi map during the earthquake in Haiti [5]). The departure point for the Verily platform is the posting of a verification request. This request is structured as a yes or no, event-based question. For example: Has the Brooklyn Bridge been damaged by Hurricane Sandy? The posting of the request subsequently triggers the collection of evidence to assess whether or not a given event has actually happened. This assessment is performed through evaluation of collected evidence which is aggregated into a judgment.

Participants with access to evidence need to be mobilized in order to collect relevant evidence. Verily uses lessons from recent crowdsourcing projects and social mobilization experiments to design incentives for attracting such participants. Following the Red Balloon Challenge [8] and the Tag Challenge [9], Verily provides explicit incentives for referrals. In the case of Verily, incentives are not monetary, but take the form of virtual points. We discuss below why points may be desirable. Similar to the Red Balloon Challenge the points are awarded using a recursive mechanism for the volunteers recruited by a user. The referral points are awarded only if a recruit has made a useful contribution. In particular, a contribution is useful if it leads to proving or disproving a claim. The system only rewards those participants (and recursively their recruiters) who provided evidence supporting the final decision, or voted in favor of this decision. This ensures points are awarded only for useful contributions and encourages referrers to invite relevant people.

The points are desirable for multiple reasons. Participants may want points just for the gamification value.[3] The points also ensure the participants that their helpful contributions are acknowledged. Furthermore, points can be used to provide more explicit motivation such as public recognition (e.g., being listed as a top helper on the project's website), a way to recommend oneself to the humanitarian community with the view of becoming a trusted ambassador, lottery to win a non-monetary prize, receiving new equipment to carry out tasks more effectively, getting an invitation to a dinner with key project people, etc. Another purpose of points is to provide a measure of reputation, which is crucial for evaluating evidence.[4]

Evaluation of evidence is performed through voting. Specifically, each piece of evidence submitted to the platform is displayed for evaluation: participants can click to vote for or against the piece of evidence and enter supporting comments (See Figure 1). Votes are weighted by voters' reputation, which is given by the points earned by the participant. Collection and aggregation of votes will be based on recent
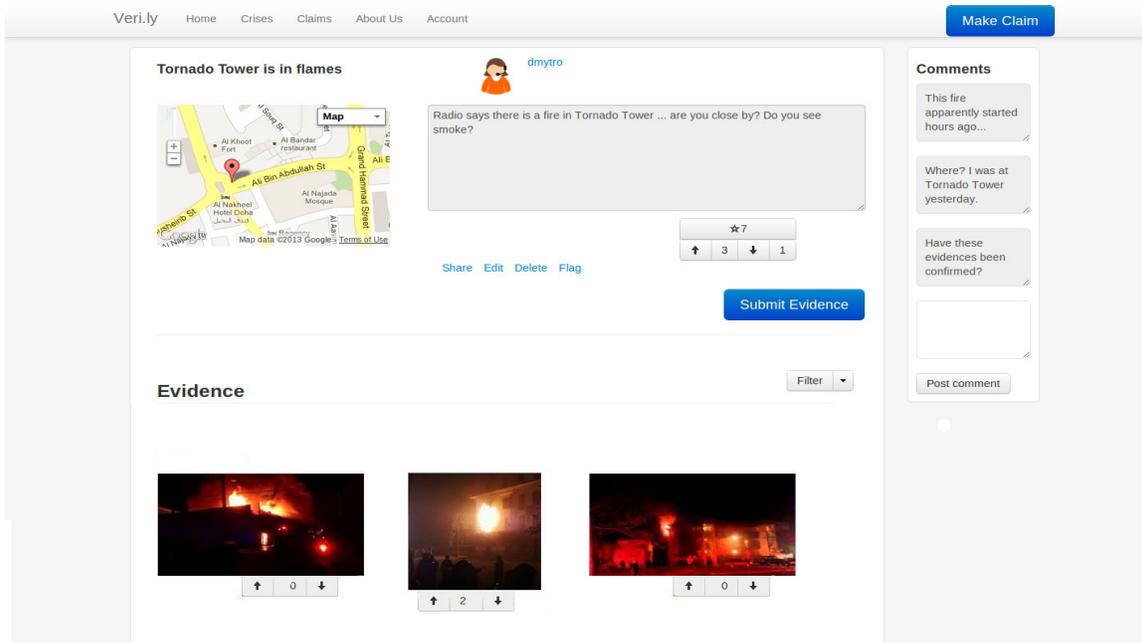
Figure 1: The claim page showing a sample information request.

research on crowdsourced opinion aggregation [10, 2, 11] and voting mechanisms [1].

Points provide a measure of trustworthiness for participants who have a history of interactions with the platform. However, most of the participants during a disaster will be new, and the platform needs to make sure malicious or irrelevant reports are detected. New members who joined through a direct referral or were vouched for by an existing trusted participant are endowed with a higher initial reputation. For all participants, the platform undertakes automated checks for malicious behavior using information about referrals, time stamps and IP addresses of contributions, and patterns in individual and group behavior. In particular, the checks will attempt to recognize collusive attempts to sabotage the system. In case of high degree of malicious behavior, Verily will limit the types and number of tasks new participants can perform until they provide enough contributions that are judged useful by established participants. In case of individual malicious behavior, contributions of new participants (e.g., evaluation or submission of evidence) will not take effect until reviewed by a trusted participant.

Participants will have access to a dashboard with a variety of relevant information about the task and trustworthiness of participants who contributed each piece of information. Consider it a "meta-data and information-fusion cheat sheet" for participating volunteers to accelerate their ability to verify information and evaluate submitted evidence.

4. VERILY SCENARIOS

Consider the following Scenario 1. An earthquake has just struck the country of Chile. According to breaking news reports from the mainstream media, the immediate extent of the damage is unclear. Several Twitter hashtags are already appearing and contradictory reports are circulating that a specific bridge in Santiago has been destroyed. High resolution satellite imagery is not yet available to independently confirm the status of the bridge. So the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) in Chile posts a verification request on Verily, a platform designed to crowdsource the collection of evidence during sudden-onset natural disasters. Within minutes, two pictures of the destroyed bridge appear on Verily. Minutes later, dozens of tweets reporting the destroyed bridge are also posted on Verily, including one with a link to a short 20 second video of the bridge, half of which has collapsed into the water. Based on Verily guidelines borrowed from the BBC's User-Generated Content Hub, the video appears authentic. Given the aggregated evidence, UN OCHA concludes that the bridge has indeed been destroyed. They pass this information on to the nearest hospital and proceed to consider other bridges to carry out their disaster response operations.

Reputation of the participants who submitted evidence that helped proving the bridge had collapsed is increased, while any reports showing that the bridge was intact, result in a decreased reputation for participants who submitted them or voted them up.

Consider the following Scenario 2. The worst flash floods in decades have just struck the capital city Seoul. There are unconfirmed reports that the water levels are still rising in the neighborhood of Gangnam even though the rains have stopped. Juni, a journalist with Seoul TV Channel 4 is headed to the area with his camera crew to investigate. Meanwhile, a verification request is posted on Verily: "Is it true that the water levels are still rising in Gangnam neighborhood?" Brian, a new member of Verily who lives in Seoul sees that a verification request has been posted in his country of residence. He turns his TV on just as Juni is speaking live, reporting that the waters are starting to recede. So Brian simply posts this information on Verily, noting that Juni on Seoul TV Channel 4 has just confirmed that the wa-

ter levels are beginning to drop in Gangnam. This report is then confirmed by another user who posts a link to the same information posted in the online edition of a national newspaper. At the same time, a photo of Gangnam submerged in water is posted by Bob as evidence that water levels are still rising. Soon the picture is voted down as irrelevant by John who posts a link to the web site where the picture had been available for over 2 years. The task is marked as resolved, and Brian and John receive a reputation boost, while Bob's reputation decreases, and he is flagged as a potentially malicious user. Reputation of the person who invited Bob and of the people Bob invited is decreased as well.

5. DISCUSSION

Accurate information is crucial for providing relief following natural disasters. Social media and specialized information-gathering platforms such as Ushahidi have proved to be invaluable for collecting information. Distinguishing which of the collected information is accurate is a necessary step for acting effectively based on it. To this end, we propose crowdsourcing evidence collection and subsequent evaluation of information generated during natural disasters. Our online platform Verily is being developed to enable and coordinate these efforts.

Our motivating example is the Red Balloon Challenge where a recursive mechanism was used to provide monetary incentives. While payments are a natural motivator in labor markets, the role of payments to stimulate volunteer activities is questionable. In particular, some experiments indicate that payments "crowd out" the volunteering spirit [4]. Furthermore, in contexts with a strong humanitarian goal such as disaster relief, being paid may be viewed negatively as profiting from sufferings of others. Of course, there is also the question of who will provide funds for monetary incentives should they be used. A final reason against using monetary incentives is that monetary rewards may make the system a target for attacks by people trying to cheat in order to receive a higher reward. Due to these reasons, we chose to stay away from monetary incentives in the current version of Verily. Having said this, we acknowledge that monetary incentives can play an important role. A practical way of implementing monetary incentives in developing countries is through transferring airtime on mobile phones [3]. In the future, we will investigate whether monetary incentives can be effectively incorporated in natural disaster scenarios.

6. REFERENCES

[1] Y. C. Andrew Mao, Ariel D. Procaccia. Better human computation through principled voting. Working Paper, 2013.

[2] Y. Bachrach, T. Graepel, G. Kasneci, M. Kosinski, and J. Van Gael. Crowd iq: aggregating opinions to boost performance. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 1*, AAMAS '12, pages 535–542, Richland, SC, 2012. International Foundation for Autonomous Agents and Multiagent Systems.

[3] N. Eagle. txteagle: Mobile crowdsourcing. *Internationalization, Design and Global Development*, pages 447–456, 2009.

[4] B. S. Frey and L. Goette. Does pay motivate volunteers? Unpublished Manuscript, 1999.

[5] J. Heinzelman and C. Waters. Crowdsourcing crisis information in disaster-affected haiti. 2010.

[6] C. J. Ho and K. T. Chen. On formal models for social verification. In *Proceedings of the ACM SIGKDD Workshop on Human Computation*, HCOMP '09, pages 62–69, New York, NY, USA, 2009. ACM.

[7] S. Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

[8] G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland. Time-critical social mobilization. *Science*, 334(6055):509–512, 2011.

[9] I. Rahwan, S. Dsouza, A. Rutherford, V. Naroditskiy, J. McInerney, M. Venanzi, N. Jennings, and M. Cebrian. Global manhunt pushes the limits of social mobilization. *Computer*, 2013.

[10] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy. Learning from crowds. *The Journal of Machine Learning Research*, 11:1297–1322, 2010.

[11] E. Simpson, S. Roberts, I. Psorakis, and A. Smith. Dynamic bayesian combination of multiple imperfect classifiers. *arXiv preprint arXiv:1206.1831*, 2012.

[12] J. Tang, M. Cebrian, N. Giacobe, H. Kim, T. Kim, and D. Wickert. Reflecting on the DARPA red balloon challenge. *Communications of the ACM*, 54(4):78–85, 2011.