Research in Progress: Understanding How Emergency Managers Evaluate Crowdsourced Data: A Trust Game-Based Approach

Kathleen A. Moore
Penn State University
College of Information Sciences and Technology
Kam6015@psu.edu

Andrea H. Tapia
Penn State University
College of Information Sciences and Technology
atapia@ist.psu.edu

Christopher Griffin
Penn State University
Applied Research Laboratory
griffinch@psu.edu

ABSTRACT

The use, or barriers to use, of crowdsourced data by emergency managers has been a significant topic of scholarly discussion during the past several years. The single strongest barrier to use has been identified as one of data quality (Tapia, et. al, 2011). We argue that within this environment the Emergency Manager (EM) acts as a decision-maker and evaluator of crowdsourced data. The final judgement on whether to incorporate crowdsourced data into a Crisis response lies with the EM. In this paper we make a brief argument for the role of EM as trustworthy data analyst and then propose a model for capturing the trust-analytical behavior through game theory (Griffin, et. al, 2012). Lastly, we propose a simple computer game, which uses our model through which we will capture EM trust-analytical behavior though a future empirical data collection effort.

Keywords

Trust. Game. Emergency Management.

INTRODUCTION

As more and more of our communication occurs online, our ability to use critical cues such as knowledge of the information source, facial or body language and common references becomes difficult or impossible. We are interested in the malleable elements of trust calibration during complex human-machine-human interactions. It is not enough to know what factors of a microblogged message establish trust in the sender and the information the sender provides. In order to automate the capture, filtering, and processing of that information for the EM, we also need to understand how and when EMs experience a variation of trust and then express it mathematically. Understanding how to mathematically capture trust serves as the foundation by which any future automated process will be built upon. In this paper we make a brief argument for the role of EM as trustworthy data analyst and then propose a model for capturing the trust-analytical behavior through both game theory and semantic content. Lastly, we propose a simple computer game, which uses our model (Griffin, at. Al 2012) through which we will capture EM trust-analytical behavior through a future empirical data collection effort.

Emergency management, broadly defined, is a field where risk is mitigated and avoided (Haddow et al. 2010) and emergency managers are tasked with the job of coordinating the various levels of a local, regional or state system including: police, fire, medical, public works, volunteers and any other group involved in dealing with disaster events (Anon 2007). Each event will have its own learning curve where an emergency manager must take in a great amount of information which is often faulty and incomplete coming from diffuse sources, and perform decision making under stressful conditions (Kowalski-Trakofler et al. 2003). This makes each disaster a case of bounded rationality, where emergency managers decision making is hampered by the stress of time, the cognitive limits each manager possess, the amount and quality of the information at hand.

For the purpose of this paper, the definition of trust is taken from Alpern where trust, on behalf of the trustor or the EM, means the acceptance of a certain amount of risk when lacking full knowledge and lacking the ability to fully control a situation (Alpern 1997). Although not performed on a conscious level, trust is based on perceived measured risk between expectations versus feeling of vulnerability versus imperfect knowledge.

Determining trust, mistrust or distrust in an online environment can be an involved process, time that EMs do not have during a crisis. If an EM is to utilize information from microblogs, trust must become at least a semi-automated process. The research involving trust in technologically mediated environments has had two distinct approaches. The first approach looks at the person supplying the information and the second looks at the information itself. Identifying who is providing information, whether the person is reliable, credible and in a position to know information is extremely valuable to establishing base level trust (Grabner-Kräuter et al. 2006). In assessing the trustworthiness of a person, research has suggested ranking reliability and credibility through reputation based on frequency and quality of past posting activity (Adler & De Alfaro 2006). Further, examining a person's affiliation in a larger social network may also indicate trust (Mendoza et al. 2010). On a personal level, analysis of sentiment, or the implied emotional state of a tweeter has been proven useful in political debate analysis, earthquakes, and during national security incidents (Diakopoulos & Shamma 2010; Qu et al. 2011). From the information side, information in a tweet may be considered credible when linked to a credible source (Starbird et al. 2010), or when it is corroboration through multiple sources (Giacobe et al. 2010). Related to reputation, a tweeter who self-corrects information, or responds to criticism of information may also be deemed credible and reliable (Shklovski et al. 2008). There has been some minor success in determining deception in some text, for example, deceptive messages tend to be longer, more informal and uncertain, more expressive and non-immediate, less complex, and less diverse than truthful messages. This is less valuable in SMS where the micro-blogged data is short and often informal.

MEASURING TRUSTWORTHY DATA, MODELING TRUST BEHAVIOR AND BUILDING A GAME

The decision to use a non-cooperative game to study this phenomenon versus a controlled experiment is based on the decision to study this from a Game Theory perspective. Game Theory provides a system for the analysis of behavior where the consequences of the actors decisions depend on the information provided by others for the EM to act upon (Stirling 2003). The conceptual structure and language attached to Game Theory provides a system to organize, capture and learn from future experience (Myerson 1992). While this could be accomplished in a controlled experiment, in using a game scenario, we hope to create a sense of competitiveness and as such, a sense of urgency, that may mimic, on a basic level, the stress of pressing events facing an EM in a real life crisis.

A Mathematical Measure of Trustworthy Data

In order to understand trustworthy information and trustworthy behavior markers, we must first establish a means to measure these (Griffin, et. al, 2012). We assume the existence of a language $\mathcal{L} = \langle \mathcal{C}, \mathcal{R} \rangle$ consisting of constants (nouns) $\mathcal{C}$ and predicates (simple assertions) $\mathcal{R}$ and no functions. Any arbitary language can be converted to such a language as needed (Bell & Machover 1977). We will also assume we are working in the first order predicate calculus, though extensions to higher-order logics may be permissible. Without loss of generality, we will impose a finite model hypothesis on sentences constructed in our language. Clearly given any model $\mathcal{M}$ that instantiates the language $\mathcal{L}$, the Tarskian definition of truth maybe applied. However, this definition is far too narrow to suit our purposes. To each sentence ω, we wish to associate a rational value $p_\omega|\mathcal{M}$ that is the proportion of truth within the sentence. If ω is as given in:

$$I_i^{\mathcal{M}}(\omega) = \begin{cases} 1 & \text{if } \mathcal{M} \models \varphi_i \\ 0 & \text{otherwise} \end{cases}$$

then;

$$p_\omega|\mathcal{M} = \frac{1}{N}\sum_{i=1}^{M} I_i(\omega)$$

This is the proportion of truth that occurs in the sentence. We assert that this definition gets to the very heart of deception or misinformation insofar as it attempts to capture the notion of "a little true." Deception hinges on the believability of the underlying story being told. A story that is 90% true with 10% falsehood is more likely to be accepted as factual than a story that is 10% true and 90% falsehood, particularly in the presence of additional, corroborating, information. Likewise, a piece of information that is false, but is close to true is more likely to be accepted. Clearly, for any sentence ω with $p_\omega|\mathcal{M} < 1$, ω is a false sentence, but the degree to which it is false is what is measured by $p_\omega|\mathcal{M}$. For the remainder of this paper, we will assume there is a special (potentially unknowable) model $\mathcal{G}$, ground truth, which describes the absolutely true state of world.

A Game Theoretic Model

Consider a simplified world in which there are two players, *Actor (the Emergency Manager),* Player 2 and *Teller(the source of the crowdsourced data),* Player 1. Player 2 will choose to act upon the information received from Player 1. We will assume time to be epochal and without loss of generality, we assume that at any time n both players have a common set of sentences Φ_n. These sentences may be axiomatic (e.g., "the sky is blue in the daytime") or they may be common information shared by the players over the course of the evolution of the situation. At time $n = 0$, there is an original (potentially) empty set of axioms Φ_0 that are introduced.At time n, suppose Player 1 wishes to provide a sentence $\sigma \notin \Phi_n$ to Player 2. The result will be $\Phi_{n+1} = \Phi_n \cup \{\sigma\}$ if Player 2 agrees to act. There are three (gross) possibilities:

1. For every model $\mathcal{M}$ so that $\mathcal{M} \models \Phi_n, \mathcal{M} \models \sigma$. That is, in every possible way the world could be so that all sentences in Φ_n hold simultaneously, it is also true that σ must be true.
2. For every model $\mathcal{M}$ so that $\mathcal{M} \models \Phi_n, \mathcal{M} \models \neg\sigma$. That is, in every possible way the world could be so that all sentences in Φ_n hold simultaneously, it is also true that σ must be false.
3. There are two models $\mathcal{M}_1$ and $\mathcal{M}_2$ so that $\mathcal{M}_1, \mathcal{M}_2 \models \Phi_n$ but $\mathcal{M}_1 \models \sigma$ and $\mathcal{M}_2 \models \neg\sigma$. That is, σ is independent of the set Φ_n.

Since Player 1 is providing the information σ, we assume that Player 1 knows whether $\mathcal{G} \models \sigma$ even if $\mathcal{G}$ is not completely knowable. Furthermore, we assume that Player 1 knows $p_\sigma|\mathcal{G}$. In the absence of this assumption, we can assume that Player 1 can compute a proxy value:

$$p_\sigma|\Phi_n = \frac{\sum_{\mathcal{M}\in\text{Mod}(\Phi_n)} p_\sigma|\mathcal{M}}{|\text{Mod}(\Phi_n)|}$$

This can be approximated, if necessary, by sampling the space of models. An algorithm for such sampling can be obtained from the hypothesis space search algorithm in [9]. In our simplified world, every piece of information is either consistent with what we know, inconsistent or could be consistent, but we can't tell. At any given time n, Player 1 will have a finite (but perhaps large) set Ψ_n of sentences that can be told to Player 2. Thus, the strategy set for Player 1 is Ψ_n while the strategy set for Player 2 is $\mathbb{B} = \{0, 1\}$, where 0 indicates no action is taken and 1 indicates an action is taken. Let Player 2's strategy be $x \in \mathbb{B}$ given $\sigma_n \in \Psi_n$. Then Player 1 receives reward $\pi^{(1)}(x, \sigma_n)$. We will assume that the marginal payoff $\pi^{(1)}(1, \sigma_n)$ is monotonically increasing with $p_{\sigma_n}|\mathcal{G}$ while $\pi^{(1)}(0, \sigma_n)$ is monotonically decreasing with $p_{\sigma_n}|\mathcal{G}$. That is, given Player 2 chooses to act, he obtains a better reward the *more true* σ_n is and if Player 2 chooses not to act, he obtains a higher cost (worse reward) the more true σ_n is. At any time epoch, ignoring global concerns, Player 2's problem is:

$$\max_{x\in\{0,1\}} x \cdot \pi^{(2)}(1, \sigma_n) + (1 - x) \cdot \pi^{(2)}(0, \sigma_n)$$

However, since Player 2 does not know $p_{\sigma_n}|\mathcal{G}$, he may use $p_{\sigma_n}|\Phi_n$ as a proxy. To differentiate this case, we write: $\pi^{(2)}(x, \sigma_n|\Phi_n)$ to denote the computed payoff based on an estimate of the veracity of σ_n from Φ_n, while $\pi^{(2)}(x, \sigma_n)$ is the true payoff.

A Multi-Turn Game Model

We now provide a simplified multi-turn game that describes the interaction between players over time: (i) At time 0, all players have an initial information set Φ_0. (ii) At each time $n \geq 0$, Player 1 chooses a strategy σ_n from Ψ_n and presents it to Player 2. (iii) Player 2 chooses a strategy $x \in \mathbb{B}$. (iv) Player 2 determines whether $\mathcal{G} \models \sigma_n$, Φ_n is updated to Φ_{n+1}. (v) Player 1 gains a reward of $\pi_n^{(1)}(x, \sigma_n)$ and Player 2 gains a reward of $\pi_n^{(2)}(x, \sigma_n)$. (vi) Player 2 decides whether to continue the game or halt play. If play continues, then return to Step 2. Halting play occurs when Player 2 no longer wishes to accept information from Player 1. This occurs in the real-world when an actor decides that a source is untrustworthy and will no longer act on information provided by that source. The objective of the game is to obtain the largest net payoff possible. We can assume a time discounted payoff function to consider arbitrarily long games. That is:

$$R^{(i)} = \sum_{n=0}^{N} \beta^n \pi_n^{(i)}(x_n, \sigma_n)$$ (Griffin et al. 2011)

where $\beta \in (0, 1)$. We denote this game as $\mathfrak{G}(N)$. This game can be presented by a multi-period game tree, where each vertex of the tree is a state indicating which player is to move and each edge indicates the strategic choice of the player in that state. **Definition 1 (Strategy)** *Let $V^{(i)}$ represent the vertices controlled by Player i in the game tree representing the multi-period game. A strategy is a rule that determines the decision made by Player i for each vertex in $V^{(i)}$.* Let $\mathbf{s}_1, \mathbf{s}_2$ be strategies for Player 1 and 2 respectively. Let $R^{(i)}(\mathbf{s}_{(1)}, \mathbf{s}_{(2)})$ be the cumulative payoff for Player i when the players use these strategies. **Definition 2 (Equilibrium)** *A strategy pair $(\mathbf{s}_1^*, \mathbf{s}_2^*)$ is an equilibrium if:* $R^{(i)}(\mathbf{s}_i^*, \mathbf{s}_{-i}^*) \geq R^{(i)}(\mathbf{s}_i, \mathbf{s}_{-i}^*)$

By $(\mathbf{s}_i, \mathbf{s}_{-i})$ we simply mean the strategy for Player i and a corresponding strategy for the alternate player $-i$. **Proposition 1** *For any finite N, there is at least one equilibrium strategy $(\mathbf{s}_1^*, \mathbf{s}_2^*)$ for $\mathfrak{G}(N)$.* Player 1 has a *motivation to deceive* Player 2, if there is at least one vertex in $v \in V^{(2)}$ so that the strategy $\mathbf{s}_1^*$ causes Player 1 to play σ_v with the property that $p_{\sigma_v}|\mathcal{G} < 1$.

Live Play

We would like to use this model to attempt to understand the behavior of individuals in situations where trust is required. In real life, Player 2 will never play to maximize his total payoff since computing this maybe impossible. When Computing $\pi^{(2)}(x, \sigma_v|\Phi_v)$ ($x \in \mathbb{B}$) the wall clock time must be taken into consideration. It may be that to compute the exact value of $\pi^{(2)}(x, \sigma_n|\Phi_v)$ is very computationally intensive because of the nature of Expression 5. However, sub-sampling $\mathrm{Mod}(\Phi_v)$ can lead to performance speedup. The question then becomes, how little time should be allotted to confirming the veracity of a given statement σ_n? To answer this question, we may assign to the teller a level of trust τ_v at each vertex $v \in V^{(1)}$. When τ_v is small, exploration of the space $\mathrm{Mod}(\Phi_v)$ is extensive in an attempt to determine a good approximating value of $\pi^{(2)}(x, \sigma_v|\Phi_v)$. When τ_v is large, exploration of $\mathrm{Mod}(\Phi_v)$ is small because trust is placed in the behavior of Player 1. We can think of $1/\tau_v$ as determining the amount of time we are willing to spend "checking out Player 2's story" as opposed to getting along with the game. Naturally, we can monetize this in the game as well, incurring a cost for each time unit spent computing $\pi^{(2)}(x, \sigma_v|\Phi_v)$. The result will be an incentive to trust whenever it is unlikely that doing so will not increase the regret function (or decrease overall payoff) (Loomes & Sugden 1982).

BUILDING A GAME: EXPERIMENTAL TESTING OF THE MODEL

Above, we have modeled trustworthy data and behavior in a crowdsourced setting. However, we argue that it is the role of emergency manager as analyst and actor, which makes the integration of these data into the crisis response system possible. We also believe that this role is vastly understudied. Therefore, we have built a mechanism by which we can supply crowdsourced data that we know to be either true or false to people playing the role of Emergency Managers and track how the person makes decisions about these data and then decides to act on those data. Understanding the human analyst in this scenario is just as important as automating data authenticity. To accomplish this, we have designed a simple computer game for evaluating our assumptions on behavior in scenarios in which they must evaluate the truth of statements made by individuals (Griffin, et. al. 2012).

In the game, the human player plays the role of an emergency manager. A devastating earthquake has just occurred. The emergency manager is given the information that a living person is trapped beneath some rubble within a 10 X 10 grid. There are no known sources of information in the area of the trapped person, but several cell towers are functioning and several individuals who are located in the area are offering information. The emergency manager is instructed to use this crowdsourced data to determine where to send the search and rescue team. The emergency manager is also told that the trapped person is near a functioning cell tower. The trapped person must be located within two units (by the Manhattan metric) of a cell tower, but no more than five units away. A player can mark a cell as potentially containing the trapped person or not possibly containing the cell. The player can also initiate a rescue. False rescues are penalized in that the emergency manager will waste both resources and time. After a set period the trapped person will die. Emergency managers are given information by five computer players who provide true and false information at varying rates. The information is easy to parse, to allow post-hoc analysis of the optimal strategy for the player. The player advances the game by using a “Done Turn” button, which triggers the computer players to provide his or her next statement. We will track the player’s behavior by recording their moves in the game. This will allow us to analyze strategic thinking after game play is done. We will also analyze responses to questions regarding which of the computer players seemed most trustworthy based on repeated play. By running the game through many iterations, we may determine patterns to allow us to better understand tipping points where trust is established, reinforced, lost, regained, or lost permanently.

CONCLUSIONS

In this paper we have provided some initial thinking on how EMs view crowdsourced data with varying degrees of trust. By understanding the tipping points between trust, mistrust, and distrust in EM decision making behavior, we hope to gain insight into understanding the perception of data versus the actual decision to use data through the EMs attempt to minimize regret. This is with the intended goal of finding mechanisms to support emergency managers as analysts within a crisis response system. In the big picture, we believe our research has the potential to help organizations that respond to disasters, make use of large amounts of citizen-produced data, which in turn may improve the speed, quality and efficiency of emergency response leading to more lives saved.

REFERENCES

1. Alpern, K. (1997) What Do We Want Trust to Be? Some Distinctions on Trust. *Business and Professional Ethics*, 16, 1-3, 29–45.
2. Anonymous. (2007) FEMA Principles of Emergency Management Supplement. Website: http://training.fema.gov/.../emprinciples/...
3. Bell, J. & Machover, M. (1977) *A Course in Mathematical Logic*, Amsterdam, Netherlands: North-Holland.
4. Diakopoulos, N.A. & Shamma, D.A. (2010) Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI ’10*. New York, NY.
5. Giacobe, N.A., Kim, H.-W. & Faraz, A. (2010) Mining social media in extreme events: Lessons learned from the DARPA network challenge. In *2010 IEEE International Conference on Technologies for Homeland Security (HST)*. Waltham, MA.
6. Grabner-Kräuter, S., Kaluscha, E.A. & Fladnitzer, M. (2006) Perspectives of online trust and similar constructs. In *Proceedings of the 8th international conference on Electronic commerce, The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet - ICEC ’06*. New York, NY.
7. Griffin, C. and Moore, K. (2012) A Framework for Modeling Decision Making and Deception with Semantic Information, *IEEE Symposium on Security and Privacy Workshops 2012*, San Francisco, CA.
8. Griffin, Christopher, Testa, K. & Racunas, S. (2011) An algorithm for constructing and searching spaces of alternative hypotheses. *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society*, 41, 3, 772–82.
9. Haddow, G.P., Bullock, J.A. & Coppola, D.P. (2010) *Introduction to Emergency Management* 4th ed., Burlington, MA: Elsevier.
10. Kowalski-Trakofler, K.M., Vaught, C. & Scharf, T. (2003) Judgment and decision making under stress: An overview for emergency managers. *International Journal of Emergency Management*, 1, 3, 278–289.
11. Loomes, G. & Sugden, R. (1982) Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty. *The Economic Journal*, 92, 368, 805–824.
12. Mendoza, M., Poblete, B. & Castillo, C. (2010) Twitter Under Crisis: Can we trust what we RT? SOMA '10 Proceedings of the First Workshop on Social Media Analytics, New York, NY.
13. Myerson, R.B. (1992) On the Value of Game Theory in Social Science. *Rationality and Society*, 4, 1, 62–73.

14. Qu, Y. et al. (2011) Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake. In *Proceedings of the ACM 2011 conference on Computer Supported Cooperative Work - CSCW ’11*. New York, NY.
15. Shklovski, I., Palen, L. & Sutton, J. (2008) Finding community through information and communication technology in disaster response. *Proceedings of the ACM 2008 conference on Computer supported cooperative work CSCW 08,* San Diego, CA.
16. Starbird, K. et al. (2010) Chatter on the Red: What Hazards Threat Reveals About the Social Life of Microblogged Information. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work - CSCW ’10*. New York, NY.
17. Stirling, W.C. (2003) *Satisficing Games and With applications to engineering and computer science*, Cambridge, United Kingdom: Cambridge University Press.
18. Tapia, A.H. et al. (2011) Seeking the Trustworthy Tweet: Can Microblogged Data Fit the Information Needs of Disaster Response and Humanitarian Relief Organizations. In *Proceedings of the 8th International ISCRAM Conference,* Lisbon, Portgual.