Proceedings of the Tenth International AAAI Conference on
Web and Social Media (ICWSM 2016)

Theme-Relevant Truth Discovery on Twitter: An Estimation Theoretic Approach

Dong Wang, Jermaine Marshall, Chao Huang
Department of Computer Science
University of Notre Dame
Notre Dame, Indiana, 46556

Abstract

Twitter has emerged as a new application paradigm of sensing the physical environment by using human as sensors. These human sensed observations are often viewed as binary claims (either true or false). A fundamental challenge on Twitter is how to ascertain the credibility of claims and the reliability of sources without the prior knowledge on either of them beforehand. This challenge is referred to as *truth discovery*. An important limitation exists in the current Twitter-based truth discovery solutions: they did not explore the theme relevance aspect of claims and the correct claims identified by their solutions can be completely irrelevant to the theme of interests. In this paper, we present a new analytical model that explicitly considers the theme relevance feature of claims in the solutions of truth discovery problem on Twitter. The new model solves a bi-dimensional estimation problem to jointly estimate the correctness and theme relevance of claims as well as the reliability and theme awareness of sources. The new model is compared with the discovery solutions in current literature using three real world datasets collected from Twitter during recent disastrous and emergent events: Paris attack, Oregon shooting, and Baltimore riots, all in 2015. The new model was shown to be effective in terms of finding both correct and relevant claims.

Introduction

This paper develops a new analytical model to address the theme-relevant truth discovery problem on Twitter. Twitter has emerged as a new application paradigm of sensing the physical environment by using human as sensors (Wang et al. 2014b). This paradigm is motivated by the massive data dissemination opportunities enabled by online social media and ubiquitous wireless connectivity (Wang, Abdelzaher, and Kaplan 2015). For example, survivors may tweet to document the damage and outage in the aftermath of a disaster or emergency event (Aggarwal and Abdelzaher 2013). These human sensed observations are often viewed as binary claims (either true or false). A fundamental challenge on Twitter is how to ascertain the correctness of claims and the reliability of sources without the prior knowledge on either of them beforehand. This challenge is referred to as *truth discovery*.

Consider an disaster response scenario (e.g., campus shooting, hurricane, or terrorist attack) as an example. People around the prime location are likely to tweet about the current situation of the event (e.g., the shooter's location, damage made by hurricane, and police reactions). It is very challenging to accurately ascertain the correctness of these reports with little or no knowledge about the data sources and the claims they make a priori (Wang et al. 2012). For example, we normally do not know when and where the emergent event will happen and who will get involved and report on Twitter (Wang et al. 2013). Moreover, sources may use the keyword of the emergent event (e.g., hashtag in Twitter) to generate completely irrelevant claims with a purpose of attracting more public attention (Chang 2010). All these complexities make the truth discovery on Twitter a challenging task to accomplish.

Important progress has been made to solve the truth discovery problem in data mining, machine learning and network sensing communities (Wang et al. 2012; 2014b; Ouyang et al. 2015; Yin and Tan 2011; Zhao, Cheng, and Ng 2014; Huang and Wang 2015; Huang, Wang, and Chawla 2015) However, a key limitation exists. In particular, current solutions did not explore the theme relevance aspect of claims and the correct claims identified by their solutions can be completely irrelevant to the theme of interests (Wang et al. 2015). For example, during the Baltimore Riots event 2015, people reported their claims on Twitter that are both relevant and irrelevant to the theme of the riot event (Table 1). It is extremely challenging (if possible) to identify a set of keywords that could perfectly classify claims into theme relevant vs them irrelevant ones, especially with the absent of knowledge on a particular event before it happens. Simply ignoring the theme relevance feature of claims in the truth discovery solutions will generate many irrelevant claims that are useless in the decision making process (Ouyang et al. 2015).

There exists a few technical challenges to incorporate the theme relevance feature of claims into the truth discovery solutions. First, Twitter is an open data contribution platform where the source reliability (the likelihood of a source to report correct claims) and the source theme awareness (the likelihood of a source to report theme relevant claims) are often *unknown* a priori. Second, it is not straightforward to identify a predefined set of keywords (e.g., the hashtags on

Tweet	Theme Relevance
A reporter just asked President Obama about the riots in #Baltimore. Here's his powerful, 15 minute long response.	Relevant
Follow Me Please. Troops deployed for Baltimore riots: Thousands of troops and police officers are.	Relevant
Working in #baltimore today it's good I have a new #tesd episode to make me laugh!	Irrelevant
We have been able to serve food for #liberty64 kids today thanks to generous donors and volunteers at #Baltimore	Irrelevant

Table 1: Theme Relevant and Irrelevant Theme Claims in Baltimore Riots Event, 2015

Twitter) to clearly classify theme relevant claims from the theme irrelevant ones because: (i) the predefined keywords may not necessarily appear in all theme relevant tweets (e.g., different words can be used to describe the same event on Twitter); (ii) theme irrelevant tweets can also contain the predefined keywords (e.g., to obtain public attention).

To address the above challenges, this paper develops a new principled approach that explicitly exploits the *theme relevance* feature of claims in a Twitter-based truth discovery solution. The new approach solves a bi-dimensional estimation problem by modeling the theme relevance feature of claims as a vector of latent variables. In particular, we develop a new Expectation Maximization (EM) based algorithms, theme-relevant EM (TR-EM), to jointly estimate i) correct and theme relevance values of claims and ii) reliability and theme awareness values to sources without knowing either of them a priori. We compared the TR-EM with the current theme-ignorant truth discovery solutions using three real world datasets collected from Twitter during recent disastrous and emergent events: Paris attack, Oregon shooting, and Baltimore riots, all in 2015. The evaluation results showed that the TR-EM scheme effectively identifies both correct and theme relevant claims in the truth discovery results and significantly outperforms other baselines. The results of this paper will enable Twitter-based application to efficiently extract the valuable information (both theme relevant and correct) from massive noisy, conflicting and incomplete data using a new analytical approach.

In summary, our contributions are as follows:

- This paper explicitly exploits both the theme relevance and correctness aspect of claims in solving the truth discovery problem on Twitter.
- We develop a new analytical model that allows us to derive optimal solutions in a bi-dimensional estimation problem that are most consistent with the observed Twitter data.
- We investigate the performance of the TR-EM scheme and other truth discovery solutions through extensive evaluation on three real world Twitter datasets. The evaluation results validate the effectiveness of our new scheme in terms of finding both correct and relevant claims.

Related Work

There exists a good amount of work in data mining on the topics of *fact-finding* that jointly compute the source reliability and claim credibility (Gupta and Han 2011). *Hubs and Authorities* (Kleinberg 1999) proposed a fact-finding model based on linear assumptions to compute scores for sources and claims they asserted. Yin et al. developed an unsupervised fact-finder called *TruthFinder* to perform trust analysis on heterogeneous information networks (Yin, Han, and Yu 2008). Other fact-finders extended these basic frameworks by considering properties or dependencies within claims and sources (Wang et al. 2011; Qi et al. 2013). More recently, new fact-finding algorithms have been designed to address the background knowledge (Pasternack and Roth 2011), multi-valued facts (Zhao et al. 2012), data provenance (Wang et al. 2014a), source uncertainty (Wang and Huang 2015), information collision avoidance (WANG et al. 2008; Wang, Zhao, and Wang 2007), multi-dimensional aspects of the problem (Yu et al. 2014). This paper uses the insights from the above work and develops a new estimation model to explicitly model unreliable human sensors and solve the theme relevant truth discovery problem on Twitter.

Our work is also related with reputation and trust systems that are designed to study the reliability/credibility of sources (e.g., the quality of providers) (Wang and Vassileva 2007; Cabral and Hortacsu 2010). eBay is a homogeneous peer-to-peer based reputation system where participants rate each other after a transaction (Houser and Wooders 2006). Alternatively, Amazon is a heterogeneous on-line review system where sources offer reviews and comments on products they purchased (Farmer and Glass 2010). Recent work has also investigated the consistency of reports to estimate and revise trust scores in reputation systems (Huang, Kanhere, and Hu 2010; Kaplan, Scensoy, and de Mel 2014; Huang, Kanhere, and Hu 2014). However, we normally do not have enough history data to compute the converged reputation scores of sources on Twitter (Wang et al. 2013; 2012). Instead, this paper presents a principled estimation approach that jointly estimates the reliability and theme awareness of sources as well as the correctness and theme relevance of claims based on the data collected from Twitter.

Maximum likelihood estimation (MLE) framework is a widely used technique in the Wireless Sensor Network (WSN) and data fusion communities (Pereira, Lopez-Valcarce, and others 2013; Sheng and Hu 2005; Msechu and Giannakis 2012). For example, Pereira et al. proposed a MLE algorithm for distributed estimation in WSN in based on diffusion (Pereira, Lopez-Valcarce, and others 2013). Sheng et al. developed a MLE method to infer locations of multiple sources by using acoustic signal energy measurements (Sheng and Hu 2005). Eric et al. deisgned a MLE based approach to aggregate the signals from remote sensor nodes to a fusion center without any inter-sensor collaborations (Msechu and Giannakis 2012). However, the above

work primarily focused on the estimation of continuous variables from physical sensor measurements. In contrast, this paper focuses on a set of *binary variables* that represent either true/false and relevant/irrelevant claims from human sensors. The discrete nature of the estimation variables leads to a more challenging optimization problem that has been solved in this paper.

Problem Formulation

In this section, we formulate our theme-relevant truth discovery problem as a bi-dimensional maximum likelihood estimation problem. In particular, we consider a Twitter application scenario where a group of M sources (Twitter users) $S = (S_1, S_2, ..., S_M)$ report a set of N claims $C = C_1, C_2, ..., C_N$. In this paper, we consider two independent features of a claim: (i) theme relevance: whether a claim is related to the theme of interests or not; (ii) correctness: whether a claim is true or false. We let S_u denote the u^{th} source and C_k denote the k^{th} claim. $C_k = O$ and $C_k = \overline{O}$ represent that claim C_k is relevant or irrelevant to the theme of interests respectively. In Twitter-based applications, sources may indicate a claim to be relevant to a certain theme (e.g., using hashtags). Furthermore, $C_k = T$ and $C_k = F$ represent the claim to be true or false respectively. We further define the following terms to be used in our model.

- ST is defined as a $M \times N$ matrix to represent whether a source indicates a claim to be theme relevant or not. It is referred to as the *Source-Theme Matrix*. In ST, $S_uT_k = 1$ when source S_u indicates C_k to be relevant to a theme of interests and $S_uT_k = -1$ when source S_u does not indicate C_k to be theme relevant and $S_uT_k = 0$ if S_u does not report C_k at all.
- SC is defined as a $M \times N$ matrix to represent whether a source reports a claim to be true. It is referred to as the *Source-Claim Matrix*. In SC, $S_uC_k = 1$ if source S_u reports claim C_k to be true and $S_uC_k = 0$ otherwise. We assume that a source will only report the positive status of a claim (e.g., in a smart city application to report potholes on city streets, sources will only generate claims when they observe potholes) (Wang et al. 2012; 2014b).

One key challenge in Twitter-based applications lies in the fact that sources are often unvetted and they may not always report relevant and truthful claims. Hence, we need to explicitly model both the theme awareness and reliability of sources. First, we define the *theme-relevantness* of source S_u as T_u: the probability that a claim C_k is theme relevant given the source S_u indicates it to be. Second, we define the reliability the *reliability* of source S_u as R_u: the probability that a claim is true given that source S_u reports it to be true. Formally, T_u and R_u are defined as follows:

$$\begin{aligned} T_u &= \Pr(C_k = O | S_uT_k = 1) \\ R_u &= \Pr(C_k = T | S_uC_k = 1) \end{aligned} \tag{1}$$

We further define a few conditional probabilities that we will use in our problem formulation. Specifically, we define $H^T_{u,O}$ and $H^F_{u,O}$ as the (unknown) probability that source S_i reports a claim to be theme relevant or not given the claim is indeed theme relevant. Similarly, we define $H^T_{u,\overline{O}}$ and $H^F_{u,\overline{O}}$ as the (unknown) probability that source S_i reports a claim to be theme relevant or not given the claim is indeed theme irrelevant. Formally, $H^T_{u,O}$, $H^F_{u,O}$, $H^T_{u,\overline{O}}$ and $H^F_{u,\overline{O}}$ are defined as:

$$\begin{aligned} H^T_{u,O} &= \Pr(S_uT_k = 1 | C_k = O) \\ H^F_{u,O} &= \Pr(S_uT_k = -1 | C_k = O) \\ H^T_{u,\overline{O}} &= \Pr(S_uT_k = 1 | C_k = \overline{O}) \\ H^F_{u,\overline{O}} &= \Pr(S_uT_k = -1 | C_k = \overline{O}) \end{aligned} \tag{2}$$

In addition, if source S_i is independent, I_u and J_u are defined as the probability that source S_u reports a claim C_k to be true given that claim C_k is indeed true or false. Formally, I_u, J_u are defined as:

$$\begin{aligned} I_u &= \Pr(S_uC_k = 1 | C_k = T) \\ J_u &= \Pr(S_uC_k = 1 | C_k = F) \end{aligned} \tag{3}$$

Notice that sources may report different number of claims, we denote the probability that source S_u reports a claim to be theme relevant as $tp_{u,O}$ (i.e., $tp_{u,O} = \Pr(S_uT_k = 1)$), and denote the probability that source S_u reports a claim to be theme irrelevant as $tp_{u,\overline{O}}$ (i.e., $tp_{u,\overline{O}} = \Pr(S_uT_k = -1)$). Additionally, we denote the probability that source S_u reports a claim to be true by sp_u (i.e., $sp_u = \Pr(S_uC_k = 1)$). We further denote h_O and $h_{\overline{O}}$ as the prior probability that a randomly chosen claim is indeed relevant or irrelevant to the theme of interests respectively (i.e., $h_O = \Pr(C_k = O)$ and $h_{\overline{O}} = \Pr(C_k = \overline{O})$). We denote d as the prior probability that a randomly chosen claim is true (i.e., $d = \Pr(C_k = T)$). Based on the Bayes' theorem, we can obtain the relationship between the items defined above as follows:

$$\begin{aligned} H^T_{u,O} &= \frac{Ta_u \times tp_{u,O}}{h_O}, \quad H^F_{u,O} = \frac{(1 - Ta_u) \times tp_{u,O}}{h_O} \\ H^T_{u,\overline{O}} &= \frac{(1 - Ta_u) \times tp_{u,\overline{O}}}{h_{\overline{O}}}, \quad H^F_{u,\overline{O}} = \frac{Ta_u \times tp_{u,\overline{O}}}{h_{\overline{O}}} \\ I_u &= \frac{Re_u \times sp_u}{d}, \quad J_u = \frac{(1 - Re_u) \times sp_u}{(1 - d)} \end{aligned} \tag{4}$$

Finally, we define two more vectors of hidden variables Υ and Z where Υ indicates the theme relevance of claims and Z indicates the correctness of claims. Specifically, we define an indicator variable r_k for each claim where $r_k = 1$ when claim C_k is theme relevant and $r_k = 0$ when claim C_k is theme irrelevant. Similarly, we define another indicator variable z_k for each claim C_k where $z_k = 1$ when C_k is true and $z_k = 0$ when C_k is false.

Using the above definitions, we formally formulate the theme-relevant truth discovery problem as a multi-dimensional maximum likelihood estimation (MLE) problem: given the Source-theme Matrix ST and the Source-Claim Matrix SC, the objective is to estimate: (i) the theme

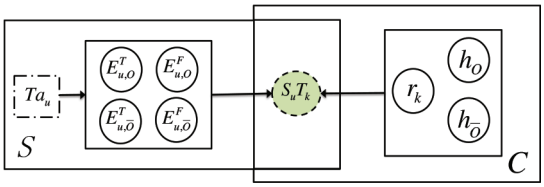


Figure 1: TR-EM Model

relevance and correctness of each claim; (ii) the theme awareness and the reliability of each source. Formally, we compute:

$$\begin{aligned}
&\forall k, 1 \leq k \leq N : \Pr(C_k = O|ST, SC) \\
&\forall k, 1 \leq k \leq N : \Pr(C_k = T|ST, SC) \\
&\forall u, 1 \leq u \leq M : \Pr(C_k = O|S_uT_k = 1) \\
&\forall u, 1 \leq u \leq M : \Pr(C_k = T|S_uC_k = 1)
\end{aligned} \tag{5}$$

Theme Relevance Identification

In this section, we present the theme relevance identification scheme: Theme-Relevance Expectation Maximization (TR-EM). The TR-EM scheme jointly estimates the theme relevance of each claim and the theme awareness of each source.

Deriving the Likelihood Function

Given the terms and variables we defined earlier, the likelihood function $L = (\Theta_{tr}; X, \Upsilon)$ for TR-EM is as follows:

$$\begin{aligned}
&L(\Theta_{tr}; X, \Upsilon) = \Pr(X, \Upsilon|\Theta_{tr}) \\
&= \prod_{k \in C} \Pr(r_k|X_k, \Theta^{(n)}_{tr}) \times \prod_{u \in S} \Psi_{k,u} \times \Pr(r_k)
\end{aligned} \tag{6}$$

where $\Theta_{tr} = (H^T_{1,O}, ..., H^T_{M,O}; H^F_{1,O}, ..., H^F_{M,O}; H^T_{1,\overline{O}}, ..., H^T_{M,\overline{O}}; H^F_{1,\overline{O}}, ..., H^F_{M,\overline{O}}; h_O; h_{\overline{O}})$ is the vector of estimation parameters for the TR-EM scheme. Note that $H^T_{u,O}$, $H^F_{u,O}$, $H^T_{u,\overline{O}}$, $H^F_{u,\overline{O}}$, h_O and $h_{\overline{O}}$ are defined in the previous section. Additionally, $\Psi_{k,u}$ and $\Pr(r_k)$ are defined in Table 2. In the table, $S_uT^O_k = 1$ and $S_uT^{\overline{O}}_k = 0$ when source S_u indicates claim C_k to be theme relevant. $S_uT^O_k = 0$ and $S_uT^{\overline{O}}_k = 1$ when source S_u reports claim C_k but does not indicate it to be theme relevant. $S_uT^O_k = 0$ and $S_uT^{\overline{O}}_k = 0$ when source S_u does not report claim C_k at all. Other notations are defined in the previous section. The model structure is illustrated in Figure 1.

The TR-EM Scheme

Given the above likelihood function, we can derive E and M steps of the proposed TR-EM scheme. First, the E-step is derived as follows:

$$\begin{aligned}
&Q(\Theta_{tr}|\Theta^{(n)}_{tr}) = H_{\Upsilon|X,\Theta^{(n)}_{tr}}[logL(\Theta_{tr}; X, \Upsilon)] \\
&= \sum_{k \in C} \Upsilon(n,k) \times \sum_{u \in S} (log\Psi_{k,u} + log\Pr(r_k))
\end{aligned} \tag{7}$$

Table 2: Notations for TR-EM

$\Psi_{k,u}$	$\Pr(r_k)$	$\Upsilon(n,k)$	Constrains
$H^T_{u,O}$	h_O	$\Upsilon^O(n,k)$	$S_uT^O_k = 1, S_uT^{\overline{O}}_k = 0, r_k = 1$
$H^F_{u,O}$	h_O	$\Upsilon^O(n,k)$	$S_uT^O_k = 0, S_uT^{\overline{O}}_k = 1, r_k = 1$
$H^T_{u,\overline{O}}$	$h_{\overline{O}}$	$1 - \Upsilon^O(n,k)$	$S_uT^O_k = 1, S_uT^{\overline{O}}_k = 0, r_k = 0$
$H^F_{u,\overline{O}}$	$h_{\overline{O}}$	$1 - \Upsilon^O(n,k)$	$S_uT^O_k = 0, S_uT^{\overline{O}}_k = 1, r_k = 0$
$1 - H^T_{u,O} - H^F_{u,O}$	h_O	$\Upsilon^O(n,k)$	$S_uT^O_k = 0, S_uT^{\overline{O}}_k = 0, r_k = 1$
$1 - H^T_{u,\overline{O}} - H^F_{u,\overline{O}}$	$h_{\overline{O}}$	$1 - \Upsilon^O(n,k)$	$S_uT^O_k = 0, S_uT^{\overline{O}}_k = 0, r_k = 0$

where $\Upsilon(n,k)$ is defined in Table 2.

In the above table, $\Upsilon^O(n,k) = \Pr(r_k = O|X_k, \Theta^{(n)}_{tr})$. It represents the conditional probability of the claim C_j to be theme relevant given the observed data X_k and current estimate of Θ_{tr}. $\Upsilon^O(n,k)$ can be further expressed as:

$$\begin{aligned}
\Upsilon^O(n,k) &= \frac{\Pr(r_k = O; X_k, \Theta^{(n)}_{tr})}{\Pr(X_k, \Theta^{(n)}_{tr})} \\
&= \frac{L^O(n,k) \times h_O}{L^O(n,k) \times h_O + L^{\overline{O}}(n,k) \times h_{\overline{O}}}
\end{aligned} \tag{8}$$

where $L^O(n,k)$, $L^{\overline{O}}(n,k)$ are defined as:

$$\begin{aligned}
L^O(n,k) &= \Pr(X_k, \Theta^{(n)}_{tr}|r_k = O) \\
&= \prod_{u=1}^{M} (H^T_{u,O})^{S_uT^O_k} \times (H^F_{u,O})^{S_uT^{\overline{O}}_k} \\
&\quad \times (1 - H^T_{u,O} - H^F_{u,O})^{1 - S_uT^O_k - S_uT^{\overline{O}}_k} \\
L^{\overline{O}}(n,k) &= \Pr(X_k, \Theta^{(n)}_{tr}|r_k = \overline{O}) \\
&= \prod_{u=1}^{M} (H^T_{u,\overline{O}})^{S_uT^O_k} \times (H^F_{u,\overline{O}})^{S_uT^{\overline{O}}_k} \\
&\quad \times (1 - H^T_{u,\overline{O}} - H^F_{u,\overline{O}})^{1 - S_uT^O_k - S_uT^{\overline{O}}_k}
\end{aligned} \tag{9}$$

In the M-step, we set derivatives $\frac{\partial Q}{\partial H^T_{u,O}} = 0$, $\frac{\partial Q}{\partial H^F_{u,O}} = 0$, $\frac{\partial Q}{\partial H^T_{u,\overline{O}}} = 0$, $\frac{\partial Q}{\partial H^F_{u,\overline{O}}} = 0$, $\frac{\partial Q}{\partial h_O} = 0$, $\frac{\partial Q}{\partial h_{\overline{O}}} = 0$. Solving these equations, we get expressions of the optimal $H^T_{u,O}$, $H^F_{u,O}$, $H^T_{u,\overline{O}}$, $H^F_{u,\overline{O}}$, h_O and $h_{\overline{O}}$ as shown in Table 3. In the table, N is the total number of claims in the Source-Theme Matrix. SF^O_u is the set of claims the source S_u indicates to be theme relevant. $SF^{\overline{O}}_u$ is the set of claims the source S_u reports but does not indicate to be theme relevant.

In summary, the input to the TR-EM scheme is the Source-Theme Matrix ST. The output is the maximum likelihood estimation of the theme relevance of claims and the theme awareness of sources. Since we assume the theme relevance feature of a claim is binary, we can classify claims as either theme relevant or theme irrelevant based on the converged value of $\Upsilon^O(n,k)$. The convergence analysis of TR-EM is presented in the next section. Algorithm 1 shows the pseudocode of TR-EM.

Table 3: Optimal Solutions of TR-EM

Notation	Solution	Notation	Solution
$(H^T_{u,O})^*$	$\frac{\sum_{k\in SF^O_u}\Upsilon^O(n,k)}{\sum_{k=1}^N\Upsilon^O(n,k)}$	$(H^F_{u,O})^*$	$\frac{\sum_{k\in SF^{\overline{O}}_u}\Upsilon^O(n,k)}{\sum_{k=1}^N\Upsilon^O(n,k)}$
$(H^T_{u,\overline{O}})^*$	$\frac{\sum_{k\in SF^O_u}\Upsilon^{\overline{O}}(n,k)}{\sum_{k=1}^N\Upsilon^{\overline{O}}(n,k)}$	$(H^F_{u,\overline{O}})^*$	$\frac{\sum_{k\in SF^{\overline{O}}_u}\Upsilon^{\overline{O}}(n,k)}{\sum_{k=1}^N\Upsilon^{\overline{O}}(n,k)}$
h^*_O	$\frac{\sum_{k=1}^N\Upsilon^O(n,k)}{N}$	$h^*_{\overline{O}}$	$\frac{\sum_{k=1}^N\Upsilon^{\overline{O}}(n,k)}{N}$

Algorithm 1 Theme-Relevant EM Scheme (TR-EM)

1: Initialize Θ_{tr} ($H^T_{u,O} = tp_{u,O}$, $H^F_{u,O} = 0.5 \times tp_{u,O}$, $H^T_{u,\overline{O}} = 0.5 \times tp_{u,\overline{O}}$, $H^F_{u,\overline{O}} = tp_{u,\overline{O}}$, $h_O \in (0,1)$, $h_{\overline{O}} \in (0,1)$)
2: $n \leftarrow 0$
3: **repeat**
4: **for** Each $k \in C$ **do**
5: compute $\Pr(r_k = O|X_k, \Theta^{(n)}_{tr})$ based on Equation (8)
6: **end for**
7: **for** Each $u \in S$ **do**
8: compute $\Theta^{(n)}_{tr}$ based on optimal solutions which are presented in Table 3.
9: **end for**
10: $n = n + 1$
11: **until** $\Theta^{(n)}_{tr}$ converges
12: Let $(\Upsilon^O_k)^c =$ converged value of $\Upsilon^O(n,k)$
13: **for** Each $k \in C$ **do**
14: **if** $(\Upsilon^O_k)^c \geq 0.5$ **then**
15: consider C_k as theme relevant
16: **else**
17: consider C_k as theme irrelevant
18: **end if**
19: **end for**
20: **for** Each $u \in S$ **do**
21: calculate T^*_u from converge values of Θ_{tr} based on Equation (4)
22: **end for**
23: Return the MLE on the theme relevance of claims judgment on claim C_k and the theme-awareness T^*_u of S_u.

Evaluation

In this section, we conduct experiments to evaluate TR-EM scheme on three real-world data traces collected in the aftermath of recent emergency and disaster events. We demonstrate the effectiveness of our proposed model on these data traces and compare the performance of our scheme to the state-of-the-art baselines. We first present the experiment settings and data pre-processing steps that were used to prepare the data for evaluation. Then we introduce the state-of-the-art baselines and evaluation metrics we used in evaluation. Finally, we show that the evaluation results demonstrate: (i) TR-EM scheme can identify theme relevant claims more accurately than the compared baselines and (ii) TR-EM can achieve non-trivial performance gains in finding more valuable (i.e., relevant and correct) claims compared to current truth discovery techniques.

Experimental Setups and Evaluation Metrics

Data Traces Statistics In this paper, we evaluate our proposed scheme on three real-world data traces collected from Twitter in the aftermath of recent emergency and disaster events. Twitter has emerged as a new experiment platform where massive observations are uploaded voluntarily from human sensors to document the events happened in the physical world (Wang et al. 2014b). The reported observations on Twitter may be incorrect or irrelevant to the theme of interests due to the open data collection environment and unvetted data sources (Aggarwal and Abdelzaher 2013). However, this noisy nature of Twitter actually provides us a good opportunity to investigate the performance of the TR-EM scheme on real world datasets. In the evaluation, we selected three data traces: (i) Paris Terrorists Attack event that happened on Nov. 13, 2015; (ii) Oregon Umpqua Community College Shooting event that happened on Oct. 1, 2015 and (iii) Baltimore Riots event that happened on April 14, 2015. These data traces were collected through Twitter open search API using query terms and specified geographic regions related to the events. The statistics of the three data traces are summarized in Table 4.

Data Pre-Processing To evaluate our methods in real-world settings, we conducted the following data pre-processing steps: (i) cluster similar tweets into the same cluster to generate claims; (ii) generate the Source-Theme Matrix (*ST* Matrix) and Source-Claim Matrix (*SC* Matrix). After the above pre-processing steps, we obtained all the inputs that are needed for the proposed scheme: *ST* Matrix and *SC* Matrix. The pre-processing steps are summarized as follows:

Clustering: we cluster similar tweets into the same cluster using a clustering algorithm based on K-means and a commonly used distance metric for micro-blog data clustering (i.e., Jaccard distance) (Rosa et al. 2011). We then take each Twitter user as a source and each cluster as a claim in our model described in the Problem Formulation Section.

Source-Theme Matrix and Source-Claim Matrix Generation: we first generate the ST Matrix using the theme indicator (i.e., hashtag: #) from the tweets. In particular, if source S_u reports the claim C_k using a hashtag in the tweet, the corresponding element S_uT_k in *ST* matrix is set to 1. Similarly, if source S_u reports claim C_k without using a hashtag, the corresponding element S_uT_k is set to -1. The element S_uT_k is set to 0 when source S_u did not report claim C_k. Second, we generate the SC Matrix by associating each source with the claims he/she reported. In particular, we set the element S_uC_k in SC matrix to 1 if source S_u generates a tweet that belongs to claim (cluster) C_k and 0 otherwise.

Evaluation Metric In our evaluation, we use the following metrics to evaluate the estimation performance of the TR-EM scheme: *Precision*, *Recall*, *F1-measure* and *Accuracy*. Their definitions are given in Table 5.

In Table 5, TP, TN, FP and FN represents True Positives, True Negatives, False Positives and False Negatives respectively. We will further explain their meanings in the context of experiments carried out in the following subsections.

Table 4: Data Traces Statistics

Data Trace	Paris Attack	Oregon Shooting	Baltimore Riots
Start Date	Nov. 13 2015	Oct. 1 2015	April 14 2015
Time Duration	11 days	6 days	17 days
Location	Paris, France	Umpqua Community College, Oregon	Baltimore, Maryland
# of Tweets	873,760	210,028	952,442
# of Users Tweeted	496,753	122,069	425,552

Table 5: Metric Definitions

Metric	Definition
$Precison$	$\frac{TP}{TP+FP}$
$Recall$	$\frac{TP}{TP+FN}$
$F1-measure$	$\frac{2\times Precison\times Recall}{Precison+Recall}$
$Accuracy$	$\frac{TP+TN}{TP+TN+FP+FN}$

Evaluation of Our Methods

In this subsection, we evaluate the performance of the proposed TR-EM scheme and compare them to the state-of-the-art truth discovery methods.

Evaluation on Theme Relevance Identification We first evaluate the capability of TR-EM scheme to correctly identify the theme relevant claims from noisy Twitter data. We compared the TR-EM with several baselines. The first one is *Voting*: it simply assumes the theme relevance of a claim is reflected by the number of times it is repeated on Twitter: the more repetitions of a claim, the more likely it is relevant to a theme of interests. The second baseline is the *Hashtag*: it considers a claim to be theme relevant if the claim contains the hashtag related to the specified theme. The third baseline is the *Sums* (Kleinberg 1999): it assumes a linear relationship between the source's theme awareness and the claim's theme relevance. The last baseline is the *TruthFinder* (Yin, Han, and Yu 2008): it can estimate the theme relevance of a claim using a heuristic based pesudo-probabilistic model.

In our evaluation, the outputs of the above schemes were manually graded to determine their performance on theme relevant claim identification. Due to man-power limitations, we generated the evaluation set by taking the union of the top 50 relevant claims returned by each scheme to avoid possible sampling bias towards any particular scheme. We collected the ground truth of the evaluation set using the following rubric:

- Theme Relevant Claims: claims that describe a physical or social event which is clearly related with a chosen theme (e.g., Paris Attack, Oregon Shooting or Baltimore Riots in our selected datasets).
- Theme Irrelevant Claims: claims that do not meet the definition of the theme relevant claims.

In our evaluation, the True Positives and True Negatives are the claims that are correctly classified by a particular scheme as theme relevant and irrelevant ones respectively. The False Positives and False Negatives are the irrelevant and relevant claims that are misclassified to each other respectively.

The evaluation results of Paris Attack data trace are shown in Table 6. We can observe that *TR-EM* outperforms the compared baselines in all evaluation metrics. The largest performance gain achieved by *TR-EM* on F1-measure and accuracy over the best performed baseline (i.e., *Hashtag*) are 6% and 9% respectively. The results of Oregon Shooting dataset are presented in Table 7. *TR-EM* continues to outperform all baselines and the largest performance gain achieved by *TR-EM* on F1-measure and accuracy compared to the best performed baseline is 18% and 11% respectively. The results of Baltimore Riots dataset presented in Table 8, similar results are observed.

We also perform the convergence analysis of the TR-EM scheme and the results are presented in Figure 2. We observe the TR-EM scheme converges within a few iterations on all three data traces. The encouraging results from the real world data traces demonstrate the effectiveness of using TR-EM scheme to correctly identify the theme relevant claims from noisy Twitter data.

Estimation Performance on Theme-Relevant Truth Discovery In this subsection, we evaluate the truth discovery performance of TR-EM scheme and compare it with the state-of-the-art truth discovery solutions that ignore the theme relevance feature of claims. The baseline that stays closet to ours is *Regular EM* (Wang et al. 2012), which computes the claims' truthfulness and sources' reliability in an iterative way and has been shown to outperform four fact-finding techniques in identifying truthful claims from social sensing data. The only difference is that Regular EM ignores the theme relevance of claims. Other baselines include *TruthFinder* (Yin, Han, and Yu 2008), *Sums* (Kleinberg 1999) and *Voting* (Pasternack and Roth 2010).

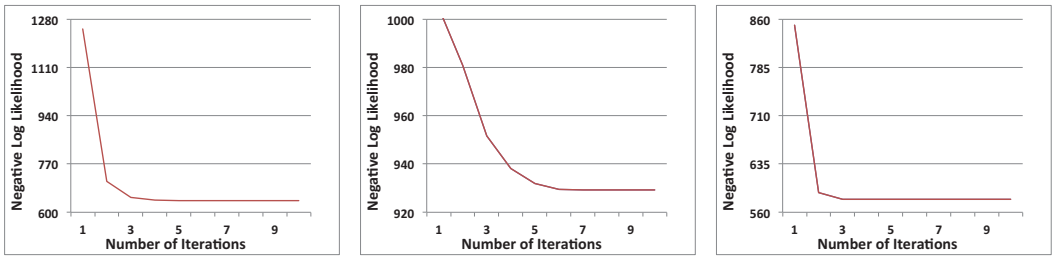
To incorporate both theme relevance and correctness of claims into our evaluation, we generalized the concept of a *correct* claim from the truth discovery problem to a *valuable* claim in the theme-relevant truth discovery problem. In particular, a valuable claim is defined as a claim that is both correct and relevant to the specified theme of interests. The valuable claims are the ones that are eventually useful in the decision making process. Similarly as the theme relevance identification evaluation, we generated the evaluation set by taking the union of the top 50 claims returned by different schemes. We collected the ground truth of the evaluation set

Table 6: Theme Relevance Identification on Paris Attack Dataset

Method	Precision	Recall	F1-measure	Accuracy
TR-EM	**0.7898**	**0.7116**	**0.7354**	**0.7150**
Hashtag	0.725	0.6277	0.6729	0.62230
TruthFinder	0.6422	0.6450	0.6436	0.5588
Sums	0.6456	0.5758	0.6087	0.5428
Voting	0.6689	0.4285	0.5224	0.51604

Table 7: Theme Relevance Identification on Oregon Shooting Dataset

Method	Precision	Recall	F1-measure	Accuracy
TR-EM	**0.7864**	**0.9419**	**0.8571**	**0.7553**
Hashtag	0.73166	0.5155	0.6244	0.5166
TruthFinder	0.7013	0.5967	0.6448	0.6405
Sums	0.7073	0.5388	0.6261	0.4985
Voting	0.6611	0.7287	0.6755	0.6103



(a) Paris Attack Dataset (b) Baltimore Riots Dataset (c) Oregon Shooting Dataset

Figure 2: Convergence Rate of TR-EM

using the following rubric:

- Valuable Claims: Claims that are statements of a physical or social event, which is related to the selected theme (i.e., Paris Attack, Oregon Shooting or Baltimore Riots) and generally observable by multiple independent observers and corroborated by credible sources external to Twitter (e.g., mainstream news media).
- Unconfirmed Claims: Claims that do not satisfy the requirement of valuable claims.

We notice that unconfirmed claims may include the valueless claims and some possibly valuable claims that cannot be independently verified by external sources. Hence, our evaluation provides pessimistic performance bounds on the estimation results by taking the unconfirmed claims as valueless. The True Positives and True Negatives in this experiment are the claims that are correctly classified by a particular scheme as valuable and valueless ones respectively. The False Positives and False Negatives are the valueless and valuable claims that are misclassified to each other respectively.

The evaluation results of Paris Attack dataset are presented in Figure 3. We observe that the proposed scheme (i.e., *TR-EM*) outperform all baselines. Specifically, the

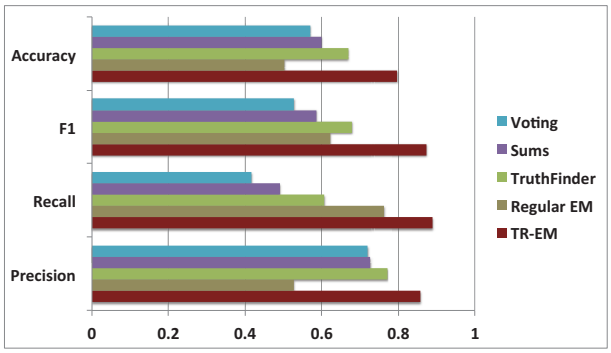


Figure 3: Truth Discovery Results on Paris Attack Dataset

largest performance gain achieved by *TR-EM* compared to the best performed baselines on precision, recall, F1-measure and accuracy is 8%, 12%, 20% and 13% respectively. The results on Oregon Shooting dataset are shown in Figure 4. We observe that our *TR-EM* continues to outperform the compared baselines and the largest performance gain it achieved over the best performed baselines on precision, recall, F1-measure and accuracy is 13%, 11%, 23%

Table 8: Theme Relevance Identification on Baltimore Riots Dataset

Method	Precision	Recall	F1-measure	Accuracy
TR-EM	**0.7489**	**0.8193**	**0.8462**	**0.8595**
Hashtag	0.7097	0.6838	0.7538	0.6419
TruthFinder	0.6194	0.6857	0.6508	0.7163
Sums	0.6376	0.6285	0.6331	0.7190
Voting	0.6290	0.5571	0.5909	0.4680

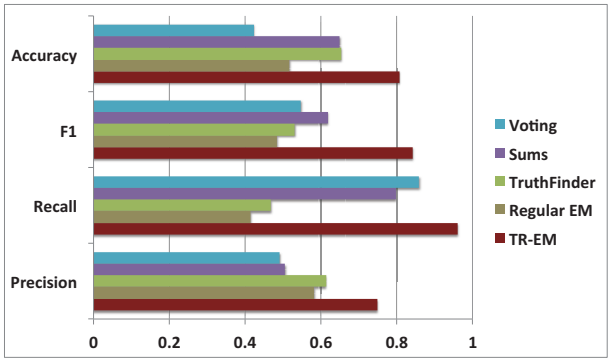


Figure 4: Truth Discovery Results on Oregon Shooting Dataset

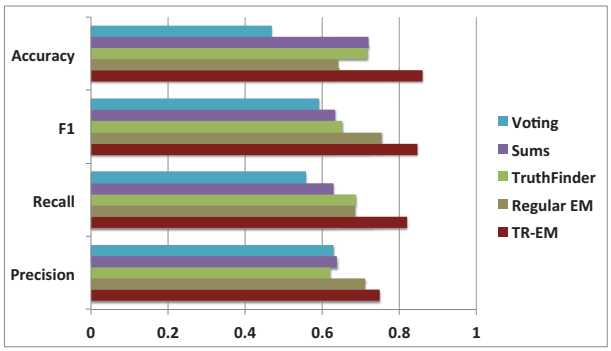


Figure 5: Truth Discovery Results on Baltimore Riots Dataset

and 15% respectively. The results on Baltimore Riots are presented in Figure 5. We observe consistent performance improvements achieved by the *TR-EM* compared to other baselines. The performance improvements of *TR-EM* are achieved by explicitly considering the *theme relevance* feature of claims on Twitter, a main challenge addressed by this paper.

Conclusion

This paper develops a new principled approach to solve the theme-relevant truth discovery problem on Twitter. The framework explicitly incorporates the theme relevance feature of claims into the truth discovery solutions. The proposed approach jointly estimates the theme awareness and reliability of sources as well as the theme relevance and truthfulness of claims using expectation maximization schemes. We evaluated our solution (i.e., TR-EM scheme) using three real world datasets collected from Twitter. The results demonstrated that our solution achieved significant performance gains in correctly identifying theme relevant and correct claims compared to the state-of-the-art baselines. The results of the paper is important because it lays out a solid analytical foundation to explore the topic relevance feature of claims on Twitter-based applications based on a rigorous analytical foundation.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1447795.

References

Aggarwal, C. C., and Abdelzaher, T. 2013. Social sensing. In *Managing and Mining Sensor Data*. Springer. 237–297.

Cabral, L., and Hortacsu, A. 2010. The dynamics of seller reputation: Evidence from eBay. *The Journal of Industrial Economics* 58(1):54–78.

Chang, H.-C. 2010. A new perspective on twitter hashtag use: diffusion of innovation theory. *Proceedings of the American Society for Information Science and Technology* 47(1):1–4.

Farmer, R., and Glass, B. 2010. *Building web reputation systems*. " O'Reilly Media, Inc.".

Gupta, M., and Han, J. 2011. Heterogeneous network-based trust analysis: a survey. *ACM SIGKDD Explorations Newsletter* 13(1):54–71.

Houser, D., and Wooders, J. 2006. Reputation in auctions: Theory, and evidence from eBay. *Journal of Economics & Management Strategy* 15(2):353–369.

Huang, C., and Wang, D. 2015. Spatial-temporal aware truth finding in big data social sensing applications. In *Trustcom/BigDataSE/ISPA, 2015 IEEE*, volume 2, 72–79. IEEE.

Huang, K. L.; Kanhere, S. S.; and Hu, W. 2010. Are you contributing trustworthy data?: the case for a reputation system in participatory sensing. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, 14–22. ACM.

Huang, K. L.; Kanhere, S. S.; and Hu, W. 2014. On the need for a reputation system in mobile phone based sensing. *Ad Hoc Networks* 12:130–149.

Huang, C.; Wang, D.; and Chawla, N. 2015. Towards time-sensitive truth discovery in social sensing applications. In *Mobile Ad Hoc and Sensor Systems (MASS), 2015 IEEE 12th International Conference on*, 154–162. IEEE.

Kaplan, L.; Scensoy, M.; and de Mel, G. 2014. Trust estimation and fusion of uncertain information by exploiting consistency. In *Information Fusion (FUSION), 2014 17th International Conference on*, 1–8. IEEE.

Kleinberg, J. M. 1999. Authoritative sources in a hyperlinked environment. *Journal of the ACM* 46(5):604–632.

Msechu, E. J., and Giannakis, G. B. 2012. Sensor-centric data reduction for estimation with wsns via censoring and quantization. *Signal Processing, IEEE Transactions on* 60(1):400–414.

Ouyang, R. W.; Kaplan, L.; Martin, P.; Toniolo, A.; Srivastava, M.; and Norman, T. J. 2015. Debiasing crowdsourced quantitative characteristics in local businesses and services. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 190–201. ACM.

Pasternack, J., and Roth, D. 2010. Knowing what to believe (when you already know something). In *International Conference on Computational Linguistics (COLING)*.

Pasternack, J., and Roth, D. 2011. Generalized fact-finding (poster paper). In *World Wide Web Conference (WWW'11)*.

Pereira, S. S.; Lopez-Valcarce, R.; et al. 2013. A diffusion-based em algorithm for distributed estimation in unreliable sensor networks. *Signal Processing Letters, IEEE* 20(6):595–598.

Qi, G.-J.; Aggarwal, C. C.; Han, J.; and Huang, T. 2013. Mining collective intelligence in diverse groups. In *Proceedings of the 22nd international conference on World Wide Web*, 1041–1052. International World Wide Web Conferences Steering Committee.

Rosa, K. D.; Shah, R.; Lin, B.; Gershman, A.; and Frederking, R. 2011. Topical clustering of tweets. *Proceedings of the ACM SIGIR: SWSM*.

Sheng, X., and Hu, Y.-H. 2005. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *Signal Processing, IEEE Transactions on* 53(1):44–53.

Wang, D.; Abdelzaher, T.; and Kaplan, L. 2015. *Social Sensing: Building Reliable Systems on Unreliable Data*. Morgan Kaufmann.

Wang, D., and Huang, C. 2015. Confidence-aware truth estimation in social sensing applications. In *Sensing, Communication, and Networking (SECON), 2015 12th Annual IEEE International Conference on*, 336–344. IEEE.

Wang, Y., and Vassileva, J. 2007. A review on trust and reputation for web service selection. In *Distributed Computing Systems Workshops, 2007. ICDCSW'07. 27th International Conference on*, 25–25. IEEE.

WANG, J.-w.; WANG, D.; TIMO, K.; and ZHAO, Y.-p. 2008. A novel anti-collision protocol in multiple readers rfid sensor networks [j]. *Chinese Journal of Sensors and Actuators* 8:026.

Wang, D.; Abdelzaher, T.; Ahmadi, H.; Pasternack, J.; Roth, D.; Gupta, M.; Han, J.; Fatemieh, O.; and Le, H. 2011. On bayesian interpretation of fact-finding in information networks. In *14th International Conference on Information Fusion (Fusion 2011)*.

Wang, D.; Kaplan, L.; Le, H.; and Abdelzaher, T. 2012. On truth discovery in social sensing: A maximum likelihood estimation approach. In *The 11th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN 12)*.

Wang, D.; Abdelzaher, T.; Kaplan, L.; and Aggarwal, C. C. 2013. Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications. In *The 33rd International Conference on Distributed Computing Systems (ICDCS'13)*.

Wang, D.; Al Amin, M. T.; Abdelzaher, T.; Roth, D.; Voss, C. R.; Kaplan, L. M.; Tratz, S.; Laoudi, J.; and Briesch, D. 2014a. Provenance-assisted classification in social networks. *Selected Topics in Signal Processing, IEEE Journal of* 8(4):624–637.

Wang, D.; Amin, M. T.; Li, S.; Abdelzaher, T.; Kaplan, L.; Gu, S.; Pan, C.; Liu, H.; Aggarwal, C. C.; Ganti, R.; et al. 2014b. Using humans as sensors: an estimation-theoretic perspective. In *Proceedings of the 13th international symposium on Information processing in sensor networks*, 35–46. IEEE Press.

Wang, S.; Su, L.; Li, S.; Hu, S.; Amin, T.; Wang, H.; Yao, S.; Kaplan, L.; and Abdelzaher, T. 2015. Scalable social sensing of interdependent phenomena. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 202–213. ACM.

Wang, J.; Zhao, Y.; and Wang, D. 2007. A novel fast anti-collision algorithm for rfid systems. In *Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on*, 2044–2047. IEEE.

Yin, X., and Tan, W. 2011. Semi-supervised truth discovery. In *WWW*. New York, NY, USA: ACM.

Yin, X.; Han, J.; and Yu, P. S. 2008. Truth discovery with multiple conflicting information providers on the web. *IEEE Trans. on Knowl. and Data Eng.* 20:796–808.

Yu, D.; Huang, H.; Cassidy, T.; Ji, H.; Wang, C.; Zhi, S.; Han, J.; Voss, C.; and Magdon-Ismail, M. . 2014. The wisdom of minority: Unsupervised slot filling validation based on multi-dimensional truth-finding. In *The 25th International Conference on Computational Linguistics (COLING)*.

Zhao, B.; Rubinstein, B. I. P.; Gemmell, J.; and Han, J. 2012. A bayesian approach to discovering truth from conflicting sources for data integration. *Proc. VLDB Endow.* 5(6):550–561.

Zhao, Z.; Cheng, J.; and Ng, W. 2014. Truth discovery in data streams: A single-pass probabilistic approach. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 1589–1598. ACM.