



A topic-focused trust model for Twitter



Liang Zhao¹, Ting Hua¹, Chang-Tien Lu, Ing-Ray Chen*

Department of Computer Science, Virginia Tech, United States

ARTICLE INFO

Article history:

Received 14 November 2014

Revised 18 July 2015

Accepted 2 August 2015

Available online 8 August 2015

Keywords:

Trust management

Social networks

Twitter

Trustworthiness

Credibility

ABSTRACT

Twitter is a crucial platform to get access to breaking news and timely information. However, due to questionable provenance, uncontrollable broadcasting, and unstructured languages in tweets, Twitter is hardly a trustworthy source of breaking news. In this paper, we propose a novel topic-focused trust model to assess trustworthiness of users and tweets in Twitter. Unlike traditional graph-based trust ranking approaches in the literature, our method is scalable and can consider heterogeneous contextual properties to rate topic-focused tweets and users. We demonstrate the effectiveness of our topic-focused trustworthiness estimation method with extensive experiments using real Twitter data in Latin America.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

As one of the most popular social messaging tools, Twitter is experiencing a tremendous growth. The number of users is over 200 million as of 2013, contributing over 200 million of tweets every day [1]. The posts in Twitter can be about any domain and any topic in the world, ranging from daily conversations to socially crucial issues. Thanks to the 140 character limitation of length, “timeliness” and “brevity” become the most distinguishing features of tweets. This empowers the freshness of the Twitter posts which usually beat traditional breaking news broadcasting media. Therefore, Twitter is becoming a promising information source to get the most timely knowledge and news around us [2]. Since different users may favor information of different topics, how to identify credible tweets belonging to the specific topics according to users’ interests is of great importance. This paper is particularly concerned with the issue of how to treat Twitter as a news channel and use our proposed trust model to identify trustworthy tweets/users.

Despite the advantages of timeliness, Twitter suffers from the fact that it is hardly a trustworthy news resource. First, tweets are usually posted by individual users instead of news authorities. The trustworthiness of tweets or users is hard to be ascertained. Second, the spread of tweets in Twitter is through social networks instead of formal news broadcasting like traditional media. In Twitter, the trustworthiness of tweets/users can only be estimated through indirect

means, such as the number of followers of a user or a tweet, and the number of retweets of a tweet. This is potentially problematic and can even foster the spread of rumors, because a malicious user can easily forge followers or retweets. Finally, the noisy nature of tweets (largely due to unstructured languages and abbreviations) further hinders accuracy of trustworthiness assessment. Tweets are often written in a casual style, without following standard grammatical rules. For example there is no verb in the tweet “Pretty bad day ioi waiting for it to go by already”. New abbreviations and slangs are emerging each day, such as TMB (tweet me back) and abt (about). These noises make it difficult to understand tweets and to properly assess their trustworthiness.

Considering the social impact of information trustworthiness in Twitter, currently there is significant interest on trustworthiness evaluation of tweets or users [3,4]. A thread of works focused on the evaluation of credibility of tweets by inspecting the contextual contents of tweets [5–10]. Typically, key features indicating the quality/credibility, such as the length and the language style, are chosen as the features to train a classifier using tweets manually labeled as credible and incredible. Another thread of works focused on investigating the trustworthiness of the users by considering the underlying social network structure of Twitter through the numbers of followers and retweets and the social relationships between users [11–14].

We observe a number of deficiencies in the works cited above and aim to devise an effective trustworthiness estimation method to remove these deficiencies:

1. Most of current work focused on evaluating the credibility of general tweets. Credibility evaluation for topic-focused tweets of users’ interest is of significantly practical use, yet hasn’t been well studied. Supervised learning method is often applied

* Corresponding author. Tel.: +(703) 538-8376; fax: +(704) 538-8348.

E-mail addresses: liangz8@vt.edu (L. Zhao), tingh88@vt.edu (T. Hua), ctlu@vt.edu (C.-T. Lu), irchen@vt.edu (I.-R. Chen).

¹ These two authors contributed equally to this work

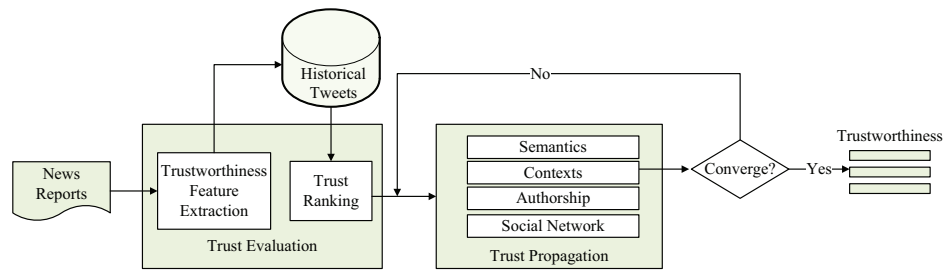


Fig. 1. System architecture.

- to identify the tweets of specific domains; however, it is not scalable to manually label credible and incredible tweets for supervised learning. To build a training dataset for supervised learning, current technologies require extensive human effort to label tweets. Moreover, labeling of tweets in the training dataset must be updated periodically. There is a need to automatically rate tweets dynamically for scalability. In our work, we do not use supervised learning so there is no need building a training set. Instead, we automatically rate topic-focused tweets by means of a novel similarity-based trust evaluation mechanism.
- Prior works treat tweets as independent of each other. Tweets are typically classified by a feature vector while the relationships between tweets are neglected. In Twitter, however, one must consider the relationships (e.g., replying, retweeting, authorship, and semantic context) among tweets as these are strong indicators to trustworthiness. For example, the tweets posted by the same untrustworthy user tend to be less trustworthy. In our work, we consider the social and contextual relationships between users/tweets for trustworthiness estimation dynamically by means of a novel iterative trust propagation algorithm.
 - Prior works are based on a social graph trust model [4] with which the credibility of a user is determined by its surrounding neighbors, e.g., how many social connections a user has. However, the social graph model is often constructed without considering the possibility that the edges in the graph can be artificially manufactured by a malicious user. One example is political astroturf, where political campaigns fake as spontaneous “grassroots” that are actually carried out by a malicious plotter or a conspiracy organization [15]. Our work is also based on social graphs. However, we do not use the social graph for directly inferring tweet trustworthiness. Rather, we rate topic-focused tweets by means of a novel similarity-based trust evaluation mechanism and then use the social and contextual relationships described by a social graph for trust propagation dynamically to achieve trust accuracy.
 - Prior works consider that trust is context independent, i.e., trust is deterministic in any situation and any context. However, in reality, trust is context dependent. A node may be trustworthy in one context, but not in another context. For example, a doctor is not as trustworthy when talking about laws, compared with medicine. In our work, we consider textual, spatial and temporal contextual features as we estimate trustworthiness of one user/tweet against another user/tweet.

In this paper, we propose a novel method to estimate the user/data trustworthiness in Twitter. Our method first accurately identifies topic-focused trustworthy tweets, and then updates the user/data trustworthiness through iterative trust propagation. To address the scalability issue, we apply our similarity-based trust evaluation method with contextual heterogeneous properties to rate users/tweets against trustworthy users/tweets (say from authorities) without the need of human efforts in labeling credible tweets for supervised learning. As shown in Fig. 1, our system consists of two main components: topic-focused similarity-based trust evaluation

and trust propagation. The first module rates users/tweets against trustworthy users/tweets for the initial trustworthiness scores, and then the second module further propagates trustworthiness scores among tweets. Our contributions are as follows:

- Untreated in the literature, we assess trustworthiness of users/tweets by a novel topic-focused trustworthiness estimation method. We propose a new design notion of similarity-based trust evaluation by which a candidate tweet is considered trustworthy if it is non-conflictingly similar in contextual properties against trustworthy tweets or trustworthy news reports from broadcasting stations. Twitter data are noisy and pointless. However, we can “infer” trust from trustworthy news reports to noisy tweets if there is a sufficient context similarity between news reports and tweets, considering textual, spatial, and temporal contextual properties. Our method is scalable and can consider heterogeneous contextual properties to rate topic-focused tweets/users.
- We propose a novel trust propagation algorithm which iteratively re-estimates the trustworthiness of users/tweets, by jointly considering their social and contextual relationships in a Twitter social graph. The theoretical proof of convergence is demonstrated.
- We demonstrate the scalability of our topic-focused trustworthiness estimation method with raw tweet data (Latin America civil unrest tweets) without the need to manually label credible and incredible tweets in a training set for supervised learning.

2. Related work

In this section we survey the state of the art in user and tweet trustworthiness assessment in Twitter. Existing approaches in general can be categorized into two types, namely, feature-based trust ranking [5–9,11,12,16–18,35–37], and social graph based trust ranking [10,13–15,19–22,38,39]. We discuss them in Sections 2.1 and 2.2, respectively. In particular, we survey the subject area of tweet trustworthiness in [5–10,16]; user trustworthiness in [11–14,35–37]; rumor and misinformation propagation in [15,21,22,38,39]; supervised learning based on classification in [5,6,8,35,37]; and unsupervised learning based on clustering in [17–20].

2.1. Feature-based trust ranking

Existing works in this category in general classify tweets related to a target topic based on credibility “features” of tweets and then apply supervised learning to classify if a tweet is credible. Gupta et al. [8] provided a SVM-rank based system TweetCred to assign a credibility score to tweets in a user’s timeline. Ravikumar et al. [11] studied features that affected user perception. Shariff et al. [16] identified eight features that cannot be automatically identified from tweets, but are perceived by users as important when judging information credibility. Weerkamp and de Rijke [6] used several credibility indicators and divided them into post-level (e.g., spelling, timeliness and document length) and blog-level (e.g., regularity, expertise, and comments). Based on these credibility indicators, they proposed a series

Download English Version:

<https://daneshyari.com/en/article/449941>

Download Persian Version:

<https://daneshyari.com/article/449941>

[Daneshyari.com](https://daneshyari.com)