

Twitter Trends Manipulation: Credibility Analysis for Twitter Trending

K. Kumaresan, G.S. Rizwana Banu, Dr.E. Baby Anitha,
M. Pavithra, S. Priya, R. Surjith Kumar and S. Keerthana

Abstract--- Information credibility on Twitter has been a topic of interest among researchers in the fields of both computer and social sciences, primarily because of the recent growth of this platform as a tool for information dissemination. Twitter has made it increasingly possible to offer near-real-time transfer of information in a very cost-effective manner. It is now being used as a source of news among a wide array of users around the globe. The beauty of this platform is that it delivers timely content in a tailored manner that makes it possible for users to obtain news regarding their topics of interest. Consequently, the development of techniques that can verify information obtained from Twitter has become a challenging and necessary task. In this paper, we propose a new credibility analysis system for assessing information credibility on Twitter to prevent the proliferation of fake or malicious information. The proposed system consists of four integrated components: a reputation-based component, a credibility classifier engine, a user experience component, and a feature-ranking algorithm. The components operate together in an algorithmic form to analyze and assess the credibility of Twitter tweets and users. We tested the performance of our system on two different datasets from

489,330 unique Twitter accounts. We applied 10-fold cross-validation over four machine learning algorithms. The results reveal that a significant balance between recall and precision was achieved for the tested dataset.

Index Terms--- Credibility, Reputation, Classification, User Experience, Feature-ranking, Twitter.

I. INTRODUCTION

ONLINE social networks, such as Twitter, have grown highly popular in the 21st century, as the numbers of users who are using them on daily basis attest. The fact that users are allowed to express themselves with little to no control is also another very attractive aspect of these platforms [1]. As users are afforded the freedom to publish content with no supervision, the problem of information credibility on social networks has also risen in recent years. For instance, Google Hot Trends ranks the hottest searches that have recently experienced a sudden surge in popularity [2]. Meanwhile, these trends may attract much more attention than before due to their appearance on Google Hot Trends. Research on information credibility is thus the best solution to the problem of how to assess the credibility of information and perhaps mitigate the dissemination of misinformation [3].

Currently, researchers have employed various methodologies in studies on information credibility, Some of them consider the problem to be one of classification that should be solved in an automated fashion using machine learning or graph-based algorithms [4]. Others view it as a cognitived problem requiring human-centric verification [5], [6]. Some authors have looked at how various aspects of social media, such as the effect of the name value and

K. Kumaresan, Assistant Professor, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

G.S. Rizwana Banu, Assistant Professor, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

Dr.E. Baby Anitha, Assistant Professor, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

M. Pavithra, UG Scholar, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

S. Priya, UG Scholar, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

R. Surjith Kumar, UG Scholar, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

S. Keerthana, UG Scholar, Department of Computer Science and Engineering, K.S.R College of Engineering, Tiruchengode, Namakkal.

user connectedness, influence users' judgments concerning credibility, [7], [8].

II. EXISTING SYSTEM

In existing system twitter is used to develop stories, track breaking news, and assess how public opinion is evaluate in the breaking story. There have been many extensive studies related to credibility in OSNs. In this section, various approaches have been highlighted in the area of credibility research, such as automated, human-based, and hybrid approaches. OSNs by their very nature evolve dynamically over time and become very large in size, with various structures that make it difficult to obtain the information needed to discern the credibility of users. The credibility of a user is influenced continuously by various factors, such as changes in the social topography, other users' behavior, preferences, and context. Malicious activities can evade existing spam filters through various means. For example, in Twitter, malicious users can purchase followers or use tools to automatically generate fake ac-counts and post tweets with the same meaning but different words.

III. PROPOSED SYSTEM

In propose system a novel credibility assessment system that maintains complete entity-awareness (tweet, user) in reaching a precise information credibility judgment. This model comprises four integrated components, namely, a reputation-based model, a feature ranking algorithm, a credibility assessment classifiers engine, and a user expertise model. All of these components operate in an algorithmic form to analyze and assess the credibility of the tweets on Twitter. "Using the reputation-based technique, we sought to automatically rank users based on their relevance and expertise on given topics. In our system, an observation is a tweet, and the positive class is credible. In this case, a highly sensitive classifier is more acceptable than precision, because non-credible tweets, if classified as credible, might spread misinformation that goes viral and cause chaos in terms of politics or an emergency. Thus, our

priority being to minimize false positives, we might choose to optimize our model with respect to recall or sensitivity. We validated our system by applying tenfold cross validation with four machine-learning algorithms on two different datasets of Twitter content. Our results show that the system that employed a reputation-based filter approach provide a significant and accurate credibility assessment.

Credibility of the Content

There is a large body of work on the automated-based approach employing machine learning techniques specifically, the supervised learning approach. This approach comprises a decision tree, a support vector machine (SVM). The paper examined automatic ways of assessing credibility via analysis of microblog postings pertaining to trending topics and classification of the posts as either credible or non-credible, using features extracted from the topics. In essence, the texts of posts, external links cited, and the posting behavior of the user were used in the classification.

Credibility of the Source during an Event

Some other researchers have shown the significance of using both content and social structure in finding credible sources. A good example of this approach is a study by Canini et al. in which an experiment was performed to determine the extent to which these factors influence both explicit and implicit judgments of credibility. Other researchers have analyzed not only ways to measure credibility on Twitter but also ways to communicate scores.

Credibility Assessment Systems

The different studies discussed earlier have yielded different results because of the different approaches taken by the authors. In general, the previous studies on this subject show that credibility assessment is possible when different dimensions are considered or different approaches are taken in the analysis. The literature also shows that it is possible to build automated systems for measuring and communicating credibility in online social networks.

Human Perception in Credibility Assessment

The cognitive process involved assesses the consistency of a message, the coherency of the message, the credibility of the source, and the general acceptability of message. The paper presents an algorithm that adopts the collaborative filtering feature of social networks to help users detect false content.

Levels of the Extracted Features

In this section, we divided the extracted features into three levels as follows:

Tweet-Level

Text features include some characteristics related to the content of the tweet such as the length of a message, the number of replies and/or the number of retweets may reflect the importance of the tweet. In addition, is the tweets contains #tags and “@mentions” as well as URLs and number of static and animated emoticons.

Sentiment features calculating the number of positive and negative words, based on a predefined sentiment words list.

User-level

Some of these features are latent and some of them explicitly revealed in user profiles. For example, age, gender, education, political orientation, and even any user preferences are considered as latent attributes. The number of followers, number of friends and the number of retweeted tweets as well as the replies of user’s tweets.

Hybrid-level

Extracting hybrid-level features is the process of aggregating most of the tweet-based features such as the URL fraction of the tweets, the hashtags (#) fraction of the tweets, and the average sentiment score in tweets. The number of duplications, which means that the user may post the same tweets more than once. To assess the credibility of Twitter content, we per-formed extraction of tweet features at three levels: the post (message) level, the user level, and the hybrid level.

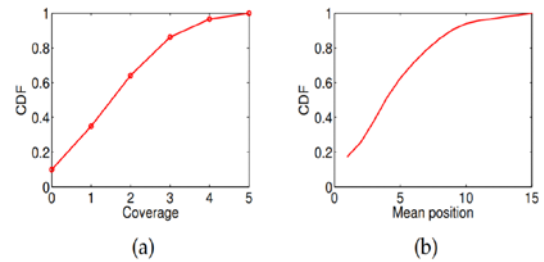


Fig. 1: Coverage and mean position of the sample trends for the public trends. 60% of them rank the common hashtags as the top 5 trends. It suggests that the sample trends of our dataset reflect the public trends.

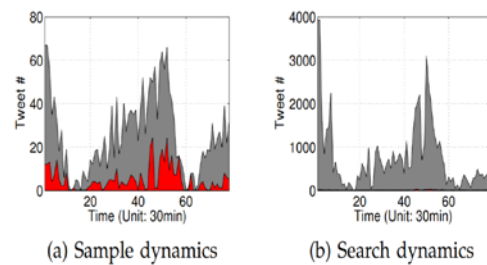


Fig. 2: Sample dynamics, search dynamics, and the intersection of them (red histogram). Relationship between the sample and search dynamics. In other words, the observed dynamics are very likely to be consistent with the general dynamics.

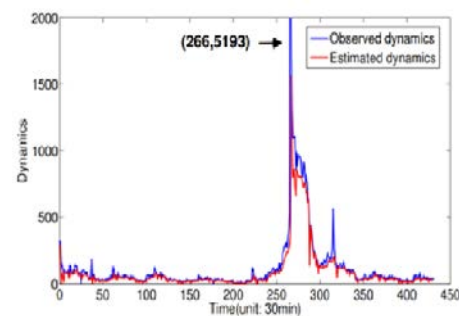


Fig. 3: The observed and estimated dynamics (tweet number) of the meme “Throwback Thursday” to drive the meme to trend beyond the effect of the network. To determine whether a hashtag is a meme, we manually check if the hashtag has been covered by any news media.

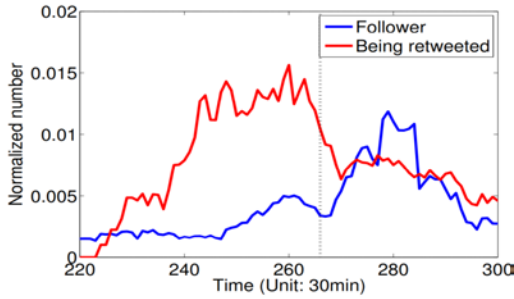


Fig. 4: The normalized number of follower and the normalized number of being retweeted for “Throwback-Thursday” around the spike. Dark dashed line represents the spike.

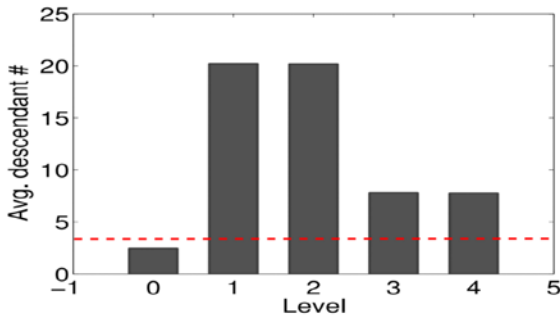


Fig. 5: Average descendant number for different levels from the malicious accounts. Dashed line represents the average descendant number of all accounts.

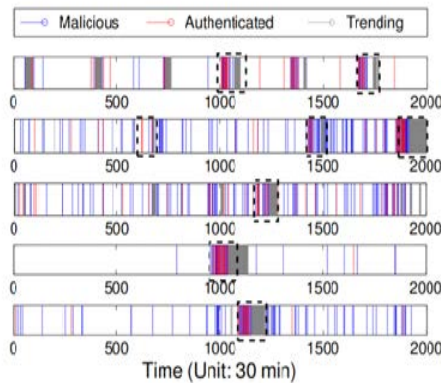


Fig. 6: Malicious account peaks and authenticated account peaks for the topics “tgif,” “wecandateif,” “ifwdate,” “MentionSomeoneHandsome,” and “mentionsomeonebeatiful” (from top to bottom).

IV. CONCLUSION

This paper presents the results of a study of the problem of assessing information credibility on Twitter. The issue of information credibility has come under scrutiny, especially in social networks that are now being used actively as first sources of information. Twitter and other social networks have become widely used in disaster mitigation in cases of high-impact events because they make it possible for relevant parties to obtain important information sufficiently quickly to coordinate countermeasures to such events.

Based on our feature extraction process, we designed an automated classification system that consists of four main components: a reputation-based component, a credibility classifier engine, a user experience component, and a feature rank algorithm. The reputation-based technique helps to filter neglected information before starting the assessment process. The classifier engine component distinguishes between credible and non-credible content. The user expertise component yields ratings of Twitter-user expertise on a specific topic. Finally, the feature rank algorithm helps in selecting the best features, based on the relative importance of each feature.

REFERENCES

- [1] Wall Street Journal (Inside a Twitter Robot Factory), <http://online.wsj.com>
- [2] M. Egele, C. Kruegel and G. Vigna, “COMPA: Detecting Compromised Accounts on Social Networks”, NDSS, 2013.
- [3] M.R. Just, A.N. Crigler, P.T. Metaxas and E. Mustafaraj, “It’s Trending on Twitter’-An Analysis of the Twitter Manipulations in the Massachusetts 2010 Special Senate Election”, APSA 2012 Annual Meeting Paper, 2012.
- [4] M. AlRubaian, M. Al-Qurishi, M. Al-Rakhami, S. M. M. Rahman, and A. Alamri, “A Multi-stage Credibility Analysis Model for Microblogs”, Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France, 2015.
- [5] C. Castillo, M. Mendoza and B. Poblete, “Information credibility on twitter”, Proceedings of the 20th international conference on World wide web, Hyderabad, India, 2011.