Volume 3, Issue 5, May 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Check On-the-Ground Twitterers Enhancing Single Twitterer(Tweets)

**R Harish***
*Computer Science*
*&Bharath university, India.*

***Abstract— In this paper "Check On-the-Ground Twitterers Enhancing Single Twitterer (Tweets)"describes an efficient method to identify the local users of an political campaign in Twitter. The reason to identify these users is that in a Political campaign, local users have a lot of valuable first-hand information to share with other users. And people are more likely to retweet accounts of those who were local to the event. The method is based on some previous work suggesting that certain characteristics of crowd's behavior could be used for identifying people tweeting from the ground during a mass disruption event, because crowd acts to individuals who are on the ground and who are not differently. Then mainly we should take the previous result check the on the ground twitterers and retweets ,follower growth ,listed growth for single twitterer.We should calculate ratio of single twitterer and combine with on- the-ground twitterers by using SVM.***

***Keywords— Micro blogging, Crisis informatics, Twitter political disruption, support vector machine***

## I. INTRODUCTION

Starting not long after its inception, the microblogging platform also consistently been appropriated for use during mass disruption events by those affected (Messina, 2007), digital volunteers (Starbird & Palen, 2011), and emergency response organizations (Sarcevic, et al., 2012). Its appeal comes from its short message, broadcast, public nature: most posts can be seen by anyone which means that interactions are not "walled" awayto a restricted groups. Then we find that Twitter is a place where information converges from across the Internet, serving often as a way-finding resource to places where additional interactions are happening. As a result, it is a valuable research site because it helps organize the otherwise boundless space of the Internet. Specifically, Twitter is a socially-networked social media platform that allows users to broadcast 140-charactermessages (tweets) to their followers, and to receive broadcast messages from users they choose to follow.Twitter users (Twitterers) maintain an account profile with information they provide including account name,user name, account description and location. Users can be make Twitter lists of other Twitterers, grouped byconversation topic or some other user-defined classification, and publicize these lists to other users to "follow."Additionally, all tweets broadcast from public accounts and all profile information associated with theseaccounts, including follower and following counts and usernames as well as user-provided information, are available for public search through application programming interfaces (APIs) made available by Twitter, afeature that permits large-scale information sharing and diffusion during mass disruption events. Throughout Twitter's young life, users have introduced and adopted linguistic conventions to adapt the platform to support their needs. These include the retweet mechanism (RT @username) to permit message forwardingwith upstream author attribution (boyd et al., 2010) and the hashtag (#keyword) to support information searchand group formation (Messina, 2007)

That paper conclude individual behavior could serve as a identifying people tweeting from the ground during a mass disruption event and combine with singe twitterer The aim of the work reported in this paper is to test this hypothesis using models from machine learning techniques.

## II. RELATED WORK

### Content Analysis – Identifying On-the-ground Tweeters

To create training data for the machine learning classifier, we needed to determine for every Twitterer whetherthey were on the ground at the New York City protests at any time during our collection window and whether ornot they tweeted first-hand information about the event. Though Twitterers can designate a location for theiraccounts, research shows that this self-reported location does not always include valid geographic data (Hecht etal., 2011). During mass disruption events, self-reported location can also be inaccurate due to physical movement of the Twitterer (e.g. in cases of evacuation or convergence) or purposeful misinformation—e.g. many remote Twitter users changed their profile location to Iran during the Iran Election protests in 2009 (Reinikainen, 2009). Geolocation metadata may be valid and accurate, but only a small fraction of tweets have geolocation information. In our *Keyword-Search-and-Filter* dataset, only 124 of 53,296 Twitterers (0.23%) had geolocation metadata on any of their tweets. For these reasons, we could not rely on self-reported location information or tweet metadata to generate enough classification data for our study.

 

We used manual content analysis to make classifications for Twitterers, beginning with an investigation of all of their OWS keyword tweets—captured either by our Search API or Streaming API searches. If we could not make a determination from there, we then went to account owners' Twitter profile pages and read *all* of their tweets—those that contained protest keywords and those that did not. Because the events were broadcast live through the Global Revolution livestream video feed3, many Twitterers were tweeting real-time information from the ground without being physically present at the event. Conversely, there were a few Twitterers who we determined to be on the ground at the event but who were not tweeting information about the protest beyond assertions of being there, going there or having been there. Since our goal is to identify new information coming from on-the-ground sources, we classified Twitterers in the *Twitterer- Sample* into two groups: A) those who were on the ground and tweeting information from the ground, and B) those who were not on the ground or were not tweeting information about the protests from the ground. Of the 2,385 Twitterers in our sample, 106 were found to be on the ground and tweeting information from the ground (A), 2270 were classified as not on the ground or tweeting information from the ground (B). Determinations could not be made for 9 Twitterers and these were excluded from the remainder of the study.

| | **Location** | **Total# of Twitterers** | **% of Total for SVM Cassification** |
|---|---|---|---|
| | Total | 2385 | 100% |
| 1 | Ground &Tweeting Ground Info (Group A) | 106 | 4.46% |
| 2 | Not Ground or Not Tweeting Ground Info (Group B) | 2270 | 95.54% |
| 3 | Unknown – Excluded | 9 | NA |

**Current Analysis-**

**Data**

To assess retweeting as a practice, we draw on twice distinct but complementary data sets. The first one datasets provide quantitative retweeting; the second two describe data we use more directly in our analysis and discussion.

**Random sample of tweets**

The first dataset is a random sample of singe tweets captured at 50 minute intervals from the public timeline over the period 2/20/13-2/03/13 using the Twitter API. This sample includes tweets from unique user, but does not include tweets from those with protected accounts. This data set provides valuable insight into the prevalence of a variety of Twitter practices. Using this data, we found that:. 42% of tweets mention a user in the form '@user'; 68% of tweets with @user begin with @user and are presumably a directed @reply

- 50% of tweet contain a hashtag (#) with 81% of these also containing a URL
- 34% of tweets include a URL ('http:')
- 3% of tweets are likely to be retweets in that they contain 'RT', 'retweet' and/or 'via' (55% include 'RT', 1% include 'via' and 50% include 'retweet')

**Random sample of retweets**

Our second set of data is a random sample of 203 retweets captured from the Twitter public timeline using the search API over the period 2/26/09- 2/03/09. This sample is only from those who have public accounts and includes tweets from unique user.

This second set of data was captured independently of the first set through explicit queries for retweets of the form 'RT' and 'via'. While other syntax is often used to indicate retweeting and we certainly missed many retweets, these two variants still provide a diverse dataset of retweets. Analyzing these,

- 46% of retweets contain a hashtag
- 74% of retweets contain a URL
- 1% of retweets contain an encapsulated retweet (RT @user1 RT @user2 ...message..)

 

• 90% of retweets contain an @reply that refers to the
person retweeting the post Compared to the random sample of tweets, hashtag usage and linking are overrepresented in retweets. Of retweets containing RT, 5% were not followed by @user. In some cases, this was because the user didn't use the @ symbol (e.g., 'RT: username'). In other cases, there was a URL but no apparent attribution (e.g., 'RT http://url.com') or 'RT' followed by a quote and attributed to 'Anonymous.' We also found that 11% of retweets containing RT included text before the RT; these appear to mostly be commentary on the retweeted content. Over 9% of all retweets include a reference to the retweeter's handle. In other words, A retweets B when B's message refers to A. We call these 'ego retweets'.

User Data collection

In that retweets we can find the User data from following tools:

- Social Media Tools:
- Twitter Tracking and Analytics Tools
- Tracking Tools to analyze Social Media Stats

So we can get best practices that will enhance user ability to get whole content into the coveted top percentage position. This is important to fully understand as 93% of Twitter users retweet on some level and 21% retweet frequently. It is also relatively useful in exposing the most retweeted user and emerging retweet on single tweets in an easy to consume format. It updates very frequently and publishes stats on the most retweeted users of the previous hour and the associated tweet. We can also see the 525 retweeted users in the past 20 days period. This is a great tool and in fact it tends to drive even more retweets once gain visibility on the different lists that they publish. There are more advanced ways to track.

## III. EXPERINENT RESULT

The features they used are some statistics about follower growth, listed growth and times retweeted. And they use SVM to train a classifier. The results shows they correctly identified 67.9% of the on ground users. It's really interesting that only with those information can we correctly identify the local user.

In that previous result we can combine our data by using SVM tool we can get the ratio of single retweets to on the ground twitterers .The result shows may be 55% of the on the ground user.

**Referance**

[1] Qu, Y., Wu, P. & Wang, X. Online Community Response to Major Disaster: A Case Study of Tianya Forum in the 2008 China Earthquake. In Proc 42nd Hawaii Int'l Conf. on System Sciences, 2009, 1-11.

[2] Starbird, K., & Palen, L. 'Voluntweeters': Self-organizing by digital volunteers in times of crisis, Proc of CHI 2011.

[3] Starbird, K., & Palen, L. (How) Will the Revolution be Retweeted? Information Diffusion and the 2011 Egyptian Uprising, (to appear) Proc of CSCW 2012.

[4] Starbird, K. Palen, L., Liu, S., & others. Promoting Structured Data in Citizen Communications During Disaster Response: An Account of Strategies for Diffusion of the 'Tweak the Tweet' Syntax. In Christine Hagar (Ed.), Crisis Information Management: Communication and Technologies, Cambridge, UK: Woodhead Publishing Limited. (In press.)

[5] Starbird, K., & Stamberger, J. Tweak the Tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting

 