I J C T A, 10(8), 2017, pp. 581-584
ISSN: 0974-5572

Modeling Power users in Twitter

G.S. Mahalakshmi*, **K. Koquilamballeand**** **and S. Sendhilkumar*****

ABSTRACT

An enormous number of tweets are generated every day. This provideshuge amounts of data to analyze, recognize patterns, construct models andpredict user behavior. Tweet Analysis can help understand user behaviorand help service providers improve their user experience. In this paper,we propose methodologies to identify whether the baseuser is a power user based on their tweets, favorites, re-tweets, hash tagsand mentions.

Keywords: Power User, Twitter, Tweets, Influential User, User Interest, User Behavior

1. Introduction

Twitter is an online social networking service that enables users to send andread short 140-character messages called "tweets". Registered users can readand post tweets, but those who are unregistered can only read them. Usersaccess Twitter through the website interface, SMS or mobile device app.Twitterwas created in March 2006 by Jack Dorsey, Evan Williams, Biz Stone, andNoah Glass and launched in July 2006. The service rapidly gained worldwidepopularity, with more than 100 million users posting 340 million tweets a dayin 2012. The service also handled 1.6 billion search queries per day.Users cantweet via the Twitter website, compatible external applications (such as forsmartphones). Users may subscribe to other users tweets this is known as"following" and subscribers are known as "followers" or "tweeps". Individualtweets can be forwarded by other users to their own feed, a process known asa "retweet". Users can also "like" (favorite) individual tweets.Twitter allowsusers to update their profile via their mobile phone by apps released for certainsmartphones and tablets.

Detection of the influential user have helped users with providing interestedtweets to them. Our aim is to extend the idea of influential users, a user is saidto be a power user if he/she gets influenced by users and he/she influences users.A power user hence, influences and gets influenced. Tweets of a user along withhis favorites,re-tweets,mentions and hash tags are collected and given for furtherprocessing. After processing is done, a user who influences the base user is found.

2. RELATED WORK

In micro-blogs, user's posts are influencing other users and are therefore get attracted. This attraction is towards the posts and not for the user of the post. But if the attraction persists continuously, eventually the user gets socially attracted to the motivational user. There may be many such motivations existing in the network for any particular user. Therefore, gradually usage of blog services become viral. In other words, the motivational user indirectly compels the users to utilize the services of the blogging network via the social posts [1].

These motivational users can be sometimes influential too. Influential users (IUs) are those users who motivate the other users to perform several actions on the posts published by him / her [2]. IUs are used by

* Department of Computer Science & Engineering, Anna University, Chennai, Tamilnadu, INDIA, *Email: gsmaha@annauniv.edu*

** Department of Computer Science & Engineering, Eswari Engineering College, Chennai, Tamilnadu, INDIA, *Email: koquilamballe465@gmail.com*

*** Department of Information Science & Technology, Anna University, Chennai, Tamilnadu, INDIA, *Email: ssk_pdy@yahoo.co.in*

marketing agencies for viral marketing. Topology based and hyperlink based IU viral marketing is quite popular. Diffusion histories are also analysed to determine IU users on the go.

Discovering top-*k*influential users plays a central role in many social network applications. This is generally performed for a given item in commercial marketing. This approach uses diffusion traces and on-line relationships for identifying the top-*k*-influential users. Topic communities are evolved and ranked using activeness, follower counts, and follower participation rates [3].

On-line support forums also use such approaches for marketing. Here, sentiment analysis is performed over user posts to obtain the user expertise on various topics and then the users are ranked to find the most influential user in every support topic [4]. Language also plays a role in determining social influence [5]. Machine learning approaches and statistical language models are being used to detect influential users.

3. PROPOSED WORK: POWER USER BASED TWITTER ANALYSIS

When users login on Twitter, they see a stream of tweets sent by friends whichcomposes their timeline. Many of these tweets are conversational tweets and/orare not of personal interest to the user. The goal of our model is to detect thecycle of influence for a particular user so that they can interact more with thatinfluencer.

A user is said to be a power user if he/she has an influential user and alsoinfluences other users. In this paper, we use the Power User Detection methodto identify power users in a given dataset. Power users are common links betweentwo sets of users in a given dataset. Power Users can be used to identify superinfluential users in a given dataset. In this, the influential userswith respect to the base user are identified.

3.1. Dataset Collection

We used the Twitter API to gather information about a user's social links andtweets. We launched our crawler for all user IDs ranging from 0 to 80 million.This API has a restriction of 15 requests per 15 minutes. We did not look beyond 80 million, because no single user in the collected data had a link toa user whose ID was greater than that value. Out of 80 million possible IDs,we found 54,981,152 in-use accounts, which were connected to each other by1,963,263,821 social links. We gathered information about a user's follow linksand all tweets ever posted by each user since the early days of the service. Intotal, there were 1,755,925,520 tweets. Nearly 8% of all user accounts were setprivate, so that only their friends could view their tweets. We ignore these usersin our analysis. The social link information is based on the final snapshot of thenetwork topology at the time of crawling and we do not know when the linkswere formed.

The network of Twitter users comprises a single disproportionately large connected component (containing 94.8% of users), singletons (5%),and smaller components (0.2%). The largest component contains 99% of alllinks and tweets. Our goal is to explore influence of users, hence we focus onthe largest component of the network, which is conceptually a single interactiondomain for users. Because it is hard to determine

Table 1
Details of the tweet files

Name of the file	*Parameters*
screen name tweets.csv	id, account created date, tweet,entities, retweet count, favorites count, in reply to screen name, language
screen name retweets.csv	id, account created date,tweet, entities, retweet count, favorites count, in reply to screen name, language
screen name mentions count.csv	screen name whichthe user has mentioned and its count
screen name hashtag count.csv	@screen name whichthe user has used and its count

influence of users who havefew tweets, we borrowed the concept of active users from the traditional mediaresearch and focused on those users with some minimum level of activity. Weignored users who had posted fewer than 10 tweets during their entire lifetime.We also ignored users for whom we did not have a valid screen name, becausethis information is crucial in identifying the number of times a user was mentioned or retweeted by others.

After filtering, there were 1,048,636 users, whom we focus on in the remainder of this paper. We have also collected the dataset based on some unique characteristics as 4 csv files separately for each user based on his/her screen name in twitter (Table 1).

Majority of the dataset was collected using the Tweepy Python Module. Thisis a wrapper API for the Twitter API. Python was used to collect the dataset.Python was the primary programming language used to collect the dataset.Around 10 GB of dataset was collected to test the Power User Detection Method.

3.2. Processing the Dataset

Once the dataset has beencollected, it has to be processed in order to make any inference. Dataset processing plays a major role in phase one as it segregates the dataset in to vitalparts which can be used during the score calculation stage. Processing is donebased on the tweet information contained in the dataset. Dataset processing involves four major sub stages. These stages help model user behavior andprovide information on user interests.

Table 2
Details of the tweet dataset

Name of the User	*Domain*
Akshay Kumar	Bollywood Actor
Atlee	Kollywood Director
Bill Gates	Microsoft, USA
Charlie Sheen	Hollywood Actor
Dalai Lama	Buddhist Monk
DeepikaPadukone	Bollywood Actress
Katy Perry	American Singer and Lyricist
Real Hugh Jackman	Australian Actor, Singer and Producer
Roger Federer	Swiss Tennis Player
Sachin Tendulkar	Indian Cricket Player
Samuel Jackson	American Actor, Producer

3.2.1. Calculating User Mentions

Every tweet by the base user may contain mentions of other users. The numberof user mentions for every user is calculated and stored in a separate file. Theuser mentions are obtained from tweets, retweets, favorites and hashtags. Ifa hashtag forms a substring of a user, the user mentions count of that useris incremented by one. User mentions is one of the important factors for scorecalculation as the base user directly mentions the target user in the tweets. Usermentions from retweets are also added.

3.2.2. Retweet History

A Retweet is a tweet shared by the base user but created by another user.Retweets help in understanding what topics the user wants to share with others. In this paper, retweets is majorly used in topic modeling so obtain the topicswhich interest the user. For every retweet, the mentions count of the owner ofthe retweet is incremented by one.

3.2.3. Hashtag Analysis

Hashtag refers to a word that begins with the symbol "#". Hashtags generallyrefers to collection of words used by a user to describe the context of the tweet.Hashtags are used in topic modeling and user mentions count as mentionedabove.

3.2.4. Favorite Tweet Analysis

Favorites refer to the tweets liked by a user. Favorites majorly define theinterests of the base user. For every favorite, the mentions count of the owner ofthe favorite tweet is incremented by one. Favorite tweets can be used to modelthe favorite topics of the base user.

3.2.5. Score Calculation

The scores are provided to each user in the files mentioned above based on constant multiplier value. The score is assigned for each user with respect to the base user.

$$\begin{aligned}\text{Final User Score} &= 1 * \text{Tweet Mentions Count} + 0.5 * \text{Hashtags Mentions Count}\\ &+ 0.5 * \text{Retweets Mentions Count} + 1 * \text{Favorites Mentions Count}\end{aligned} \quad (1)$$

The scores file is sorted ina non-increasing order based on the scores of each user. The top ten users areobtained from the new sorted list. The top ten users are stored in a separate file.The file serves as the input to phase two. The top 10 users are the influentialusers with respect to the base user and the user with the highest score beingthe most influential among them.

4. CONCLUSION

This paper proposed methodologies for modeling user interest and behavior via analysis of tweet parameters, which facilitated the identification of most influential user for any given user. The future work focuses on identification of users who are most influenced by the given user. By statistical and semantic inferences between both the sets new insights related to power user detection shall be obtained.

REFERENCES

[1] Yadav, Mahendra Kumar, and Manoj Kumar. "Determining influentialusers in blogosphere-A survey." Green Computing, Communication andConservation of Energy (ICGCE), 2013 International Conference on. IEEE, 2013.

[2] Singh, Sushil, Nitesh Mishra, and Shantanu Sharma. "Survey of various techniques for determining influential users in social networks." Emerging Trends in Computing, Communication and Nanotechnology (ICE-CCN),2013 International Conference on. IEEE, 2013.

[3] Guo, Jing, et al. "Item-based top-k influential user discovery in social networks." Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on. IEEE, 2013.

[4] Munger, Tyler, and Jiabin Zhao. "Identifying Influential Users in On-lineSupport Forums using Topical Expertise and Social Network Analysis." Proceedings of the 2015 IEEE/ACM International Conference on Advancesin Social Networks Analysis and Mining 2015. ACM, 2015.

[5] Shalaby, May, and Ahmed Rafea. "Identifying the Topic-Specific Influential Users Using SLM." 2015 First International Conference on Arabic Computational Linguistics (ACLing). IEEE, 2015.