Classification of Followers of Influential Users on Twitter

Manoj Narayan Bisarahalli, Ramya Rajendra, Sahana Shreedhar Kulkarni

Problem Overview:

Twitter is a popular microblogging service in recent years. About 200 million tweets are generated per day. Each user has a provision to share information with their friends or public in the form of text messages up to 140 characters, which are popularly known as tweets. A substantial amount of data is being generated by spam or fake users.There is also retweet mechanism that allows users to share information with their followers which inturn spreads the information rapidly. Often a popular user is followed by a large number of other users. This creates a massive celebrity-centric social network ,which plays a major role in breaking news, controversial opinions on issues and latest events.Such accounts must be given importance to understand this spread of information and how the users are influenced.This helps in profiling their followers' behaviors. Therefore ,on analyzing when and what the followers comment on tweets of these influential users we can discover differential temporal patterns and word diversity in comments.This helps in classifying the followers as spammers and normal.

Data:

Data is collected actively from a list of influential users on Twitter. We make use of the Twitter API's 'statuses/user_timeline' endpoint to fetch the timeline statuses of each user in our list cautious of the rate limits imposed by twitter. Further, we fetch the replies to each of these statuses. Since we do not have a direct endpoint to do this, we follow a set of steps as described below:

- Collect @username and id of the status we need replies for.
- Use the search API with query parameters containing 'to:@username'
- Return tweets that match the field 'id_reply_to_status_id' to 'id'

This data is then processed by trimming out trivial words, symbols or emotion icons. Further we plan on processing even the status content for further analysis.

Method:

Our approach deals with classifying the followers into spammers and normal users. We consider a follower of influential users as a spammer, if the follower posts advertisement information or phishing messages when commenting on the tweets from influential users.

The method that we are using involves tf-idf technique which is a widely used statistical measure for analyzing the frequency and diversity of individual words in tweets. This process involves getting occurrence of every word in a tweet. We consider all the comments from a follower f on a given tweet, and use tf-idf to analyze the words in the comments from a single follower on the same tweet posted by an influential user and we further compute standard deviation (to check variation from normal comments) on every tf-idf values of the words to determine how frequently the word is used in a comment by a follower and how it can be categorized as a spam. The libraries that we plan to use are sklearn, nltk, numpy, twitterAPI.

To evaluate the performance of the proposed method, we manually examine the types of followers by visiting their profile pages and historical tweets, retweets, and comments on twitter, and determine whether each follower is a spammer or a normal user. Due to the limitation of the manual process, we randomly select one thousand accounts from millions of followers of these influential users. This set of randomly selected followers provides ground truth data for our experimental evaluation, and quantifies the accuracy of our proposed technique for classifying the followers of influential users.

Intermediate/Preliminary Experiments & Results:

From the data collected from twitter on a set of influential users, we analyse some of the numbers such as original statuses, original ratio and other counts. An output of the stats function is shown below for reference.

	screen_name	friends_count	followers_count	statuses_count	n_original_statuses	original_ratio
0	Jon4Lakers	3812	177917	56597	56530	99.881619
1	TechCrunch	1050	10114364	205164	205054	99.946384
2	DalaiLama	0	19299206	1511	1511	100.000000
3	Forbes	5334	15656258	228394	228355	99.982924
4	RaySiegel	1079	11301	16708	16433	98.354082
5	DaveKaval	1238	21489	13596	13433	98.801118
6	realDonaldTrump	47	66389012	45656	45656	100.000000
7	StationCDRKelly	415	5594510	3444	3264	94.773519
8	KylieJenner	1082	29087309	13449	12235	90.973307
9	deadmau5	619	3695882	42454	42288	99.608989

Related Work:

Paper 1: 'Detection of spam-posting accounts on Twitter' by Isa Inuwa-Dutse, Mark Liptrott, Ioannis Korkontzelos
-This paper talks about the approach for distinguishing spam and non-spam social media posts and gives some insight into the behavior of spam users on Twitter detection of spam by considering the pairwise engagement with each user.The idea sounds similar to ours, whereas we analyze when and what the followers comment on tweet.[1]

Paper 2: 'Detecting Social Spam Campaigns on Twitter' by Zi Chu, Indra Widjaja, and Haining Wang
-This paper has a collective perspective, and focuses on detecting spam campaigns that manipulate multiple accounts to spread spam on Twitter. It also designs an automatic classification system based on machine learning, and apply multiple features for classifying spam campaigns. Our approach uses tf-idf (term frequency inverse document frequency) to systematically study commenting and retweeting behaviors of the followers and to discover different temporal commenting patterns and word diversity in the comments.[2]

Paper 3:'Information assurance:detection of web spam attacks in social media' by Pang-Ning Tan, Feilong Chen, and Anil K Jain
-This paper shows a co-classification framework to simultaneously detect Web spam and spammers in social media Web sites based on their content and link-based features.We use two algorithms for classification: one is based on commenting behavior and pattern and another with small number of comments based on known spammers and known normal fans[3].

Paper 4: 'Identifying Twitter Spam by Utilizing Random Forests' by Humza S. Haider
-This paper user the random forest method to classify users.We use two algorithms for classification:one is based on commenting behavior and pattern and another with small number of comments based on known spammers and known normal fans. Our approach uses tf-idf (term frequency inverse document frequency) to systematically study commenting and retweeting behaviors of the followers and to discover different temporal commenting patterns and word diversity in the comments.[4]

Paper 5: 'Spam Detection on Twitter Using Traditional Classifiers' by M. McCord and M. Chuah
-This paper user the random forest method to classify spammers and non-spammers. We use tf-idf to systematically study commenting and retweeting behaviors of the followers and to discover different temporal commenting patterns and word diversity in the comments.[5]

Who Does What:

Manoj Narayan Bisarahalli - Working on collection of data and storage in the directory. Prepare a report on accuracy and other metrics.(collect and evaluate)

Ramya Rajendra- Working on network analysis of the project.Use the retrieves data and work on analysing it.(network and stats)

Sahana Shreedhar Kulkarni- Working on algorithms for training the classifier for the labeled data. Work on web application of the project.(train and web)

The above mentioned tasks are coordinated in the form of cycle of responsibilities rather than just static allocation of tasks.

Timeline:

So far: -Researched on problem statement.
-Gathered various research papers
-Researched on algorithms that can be used.
-Data collection through twitter API.
-Cleaned the data and made it ready for analysis.

11/2/19-Completion of evaluation of data.
11/4/19- Completion of network analysis for the project and labeling data.
11/9/19- Completion of preparation of a statistical report and Finish training a classifier on the collected data with some results.
11/12/19- Finish fine tuning our classifier to improve accuracy.
11/15/19-Completion of web app for project demo.
11/20/19-Completion of report and presentation.
The above mentioned activities are worked on parallely and a rough timeline for completion of each activity are been specified.

References:

[1] Detection of spam-posting accounts on Twitter, Isa Inuwa-Dutse, Mark Liptrott, Ioannis Korkontzelos
[2] Detecting Social Spam Campaigns on Twitter,Zi Chu, Indra Widjaja, and Haining Wang
[3] Information assurance:detection of web spam attacks in social media,Pang-Ning Tan, Feilong Chen, and Anil K Jain
[4] Identifying Twitter Spam by Utilizing Random Forests,Humza S. Haider
[5] Spam Detection on Twitter Using Traditional Classifiers,M. McCord and M. Chuah