Why Did They #Unfollow Me? Early Detection of Follower Loss on Twitter

Suman Kalyan Maity
Dept. of CSE
IIT Kharagpur, India
sumankalyan.maity@cse.iitkgp.ernet.in

Ramanth Gajula
Dept. of CSE
IIT Kharagpur, India
ramanth139@gmail.com

Animesh Mukherjee
Dept. of CSE
IIT Kharagpur, India
animeshm@cse.iitkgp.ernet.in

GROUP '18, January 7–10, 2018, Sanibel Island, FL, USA
ACM 978-1-4503-5562-9/18/01.
https://doi.org/10.1145/3148330.3154514

Abstract

Having more followers has become a norm in recent social media and micro-blogging communities. This battle has been taking shape from the early days of Twitter. Despite this strong competition for followers, many Twitter users are continuously losing their followers. This work addresses the problem of identifying the reasons behind the drop of followers of users in Twitter. As a first step, we extract various features by analyzing the *content of the posts* made by the Twitter users who lose followers consistently. We then leverage these features to early detect follower loss. We propose various models and yield an overall accuracy of 73% with high precision and recall. Our model outperforms baseline model by 19.67% (w.r.t accuracy), 33.8% (w.r.t precision) and 14.3% (w.r.t recall).

Author Keywords

unfollow; social media; prediction

ACM Classification Keywords

H.4.m [Information Systems Applications]: Miscellaneous; J.4 [Computer Applications]: [Social and Behavioral Sciences]; K.4.2 [Computers And Society]: [Social Issues]

Introduction

Followership of users in social media is an important factor since it indicates social prestige and popularity for the users. Followers have a proportional impact on how far and wide one's message spreads and the rate at which one can get social recognition in form of reposts, shares, likes etc.[1] It helps in outreach, helps in forming new social relationships. Though people have studied followership gain, there are very few studies that looked into the other side of the spectrum of this online relationship - the "unfollowing" behavior. Like gain in followership, followership loss has also important social connotation and business implications. Twitter or other social media are extensively used by media houses, various industry outlets from technology to fashion, political personalities. Therefore, followership loss of such entities could mean decrease in face value and which could directly/indirectly impact business. For instance, boxer cum politician, Manny Pacquiao lost 2 million followers over his gay comments[2], Indian prime minister Narendra Modi reportedly lost 313,312 followers after announcing demonetization of 500 and 1000 notes[3]. In our dataset containing 9.3 million Twitter users, 26% of the users are found to have suffered a net loss in a two years span. For instance, a user from our dataset who had 114K followers, tweeted only about mundane details of his day to day activities and therefore lost 85 percent of his followers. Another user who had 176K followers, lost 55% of his followers most likely because the tweets mostly portray political propaganda and the tweet frequency is as high as ~ 200 tweets per day. Though both gain/loss in followership can be contextual to different relationships and situations, however, in this work we try to find holistically what factors - like *social behavior*, *textual content of posts*, *language usage and network structure* - lead to follower loss.

Dataset preparation We construct our dataset through web-based crawls of the profile information of 9.3 million users at two different time points – i) June 2014 and ii) September 2016. We then select those users who have at least 1000 followers in June 2014 and lost some followers by September 2016. We create two datasets based on followership gain/loss characteristics. *Dataset1* consists of users who lost at least 30% of their followers. *Dataset2* has users who lost at most 2% followers plus users who gained at most 2% followers. Our objective here is to identify features that discriminate this set (which corresponds to mostly accidental loss/gain of followers) from the set of users who incur a real loss of followers (i.e., *dataset1*). We further remove those users who did not tweet in English. We then randomly sample out 8000 users from *dataset1* and a similar number of users from *dataset2* for the subsequent study.

Related work: There have been few studies done by researchers to understand the dynamics of unfollowing in various OSNs. Kwak et al. [3] reported that 43% of active users unfollow at least once during 51 days. Twitter users have unfollowed those users who left many tweets within a short time, created tweets about uninteresting topics, or tweeted about the mundane details of their lives [2, 5]. Also Twitter users appreciate receiving more attention than giving when it comes to mentions, retweets etc., and this is pronounced in the act of unfollow [3]. Another popular mode of unfollowing in Twitter is burst unfollowing [6]. In this work, we propose models for early prediction of loss of followers on Twitter mainly focusing on the content and the language usage in the tweets posted by the users. In specific, we make use of the activities and the content of the tweets of the victim (i.e., the person losing the followers) only. Building such a model would enable the victim early in time to know the specific online behavior that could result in followership loss. Having such succinct clues can guide the victim as to how to contain his/her behavior to avoid loss of followers (which is usually very hard to accumulate).

Factors behind follower loss

The factors below attempt to extract the textual content and the language usage behavior of the victims.

Use of offensive/profane words in tweets: We use a list of offensive and profane words from https://www.cs.cmu.edu/~biglou/resources/bad-words.txt and manually label their offensive/badness score. We calculate badness influence per tweet as the sum of badness of the words used in the tweet normalized by the number of words in the tweet. The average badness influence of all the tweets of a user gives the *badness coefficient* of a user.

Repetitive content: word diversity: Repetitive content

[1] https://blog.bufferapp.com/definitive-guide-social-media-metrics-stats
[2] http://bit.ly/2yFnnHF
[3] http://bit.ly/2gxAzIs

is considered generally as boring in social media, unless the content is very trendy. Let T_u be the multi-set of words from all the filtered tweets of user *u* and *W* be the set of all unique words in T_u and $p(w|T_u)$ be the probability of word *w* belonging to T$_u$. We now define content diversity as $ContentDiv(u) = -\sum_{w \in W} p(w|T_u) \times log(p(w|T_u))$

Topic diversity: Topic diversity also captures the notion of repetitive content by finding topics in the tweets of user rather than directly using the words. We use LDA [1], for the discovery of latent subtopics and calculate the topical diversity for an user u as $TopDiv(u) = -\sum_{k=1}^{K} p(topic_k|T_u) \times log(p(topic_k|T_u))$ where T_u denotes the set of tweets as a document for user u.

Tweet rate: In Twitter, users would hardly want their feeds to be overflown by the tweets from a single other user. We capture this notion using the rate at which a given user is tweeting which simply is the time difference of the first and the last tweets of the user normalized by the total number of tweets so far (in the data) of the user.

Mentions per tweet: We calculate *MentionCoeff* as the average number of mentions per tweet. Users who *mention* infrequently are able to less engage other users and might get unfollowed.

Mention entropy: The *MentionCoeff* measure might implicitly (and incorrectly) indicate that a particular user *u*, who mentions only a small set of people very frequently, is very less probable of losing followers. However, these users might also be prone to losing followers. For example, users who follow 1000 people communicate with only about 70 people on average [2]. We capture this notion by using *MentionEntropy*. Let *M* be the list of distinct users mentioned by user *u* and let $p(m \mid u)$ denote the probability with which user *u* mentions user *m* in his/her tweets. So,

$$MentionEntropy(u) = -\sum_{m \in M} p(m|u) \times log(p(m|u))$$

Users who have a low mention entropy are more from *dataset1* indicating that users engaging only a particular set of other users in their tweets repeatedly are prone to lose more followers in future.

Usage of urls in tweets: Urls are popular in Twitter community for redirection. However, excessive usage of urls is usually not encouraged in the community because that is often interpreted as spamming. Users of *dataset1* have more average url count per tweet indicating that people who use excessive urls are prone to loss of followers.

Profile description and verification status of user: Profile description renders authenticity to a user profile. Interestingly, in our dataset, users who had profile description were less likely to lose followers. Verification status is also an important factor. Verified users usually have a net gain of followers. 90% of the total verified users gained followers in our dataset.

Network features We have constructed the following two networks - a) mention network b) content similarity network of the users in the dataset.
a) Mention Network: We consider the mention network of users in the full dataset where the nodes are the users and a directed edge is created from a to b if a mentions b at least once in his/her tweets. Only those users who have their (in-degree + out-degree) $>$ 0 are included in the network. $\sim$17% of users from both datasets combined are present in this mention network. We have used various centrality and clustering based features (appropriately scaled) *–in-degree centrality, out-degree centrality, betweenness centrality, closeness centrality, eigenvector centrality, clustering coefficient* from this network.

Tweet bursts: Bursts of tweets sometimes draw an unwanted attention in Twitter. An example of tweet burst includes a long story posted as a continuum of tweets. Twenty out of 22 respondents reported that they unfollowed 39 people because of burst tweets [2]. Consider the array T_u of tweets of a user *u* sorted according to the tweet arrival time. We define tweet burst as a maximal sub-array $T_u[i..j] \mid \forall k, i \leq k < j, t(k+1) - t(k) \leq 1000$ where $t(k)$ denotes the arrival time of the k^{th} tweet of the user *u*. Time period of a burst $T_u[i..j]$ is defined as $t(j) - t(i)$. In the equation 1000 is a chosen hyper parameter. We use the following set of features extracted from tweet bursts - *mean inter-burst arrival time, avg. time period of a burst, max. time period of a burst, minimum time period of a burst, no. of bursts*.
Tweet length: We observe that users with tweets having very short length and users with tweets occupying most of the allowed space, are more likely to lose followers.

Psycholinguistic aspects of tweets

We also perform psycholinguistic analysis of the tweets to observe if there exists any pattern leading to follower loss. The cognitive, linguistic and psychological dimensions are captured through different categories provided by the LIWC tool [7]. There are 64 different categories that LIWC extracts from the tweet texts. First, we collect the words related to each of these 64 categories. Next, we find for each category *c*, the number of words in the tweets of user u which belong to the category *c* and normalize this value by the total number of tweets of the user *u*. Some of the key points to note here are that users who lose more followers use more negation words, less inclusive words as well as less insightful words.

b) Content similarity network: We consider the tweets of the users as bag-of-words. We then compute user-user similarity using the Jaccard co-efficient between the tweets. We then construct a network with nodes as users and edges indicating similarity between word usage of users. Through inspection of the distribution of similarities, we prune those edges with similarity values less than 0.3. In the resulting graph, the similarity feature for a user is extracted as follows: for a node n, all the neighboring nodes whose corresponding users are in the training set are considered and the majority class of neighbors is used as a feature. Clustering coefficient of similarity network is also used as a feature.

Prediction Framework

In this section, we present our model to early predict follower loss. Apart from the content based features, we have also used the following features - no. of followers, no. of followees, followee/follower ratio.

Baseline Model: In previous studies, many link based features from the follower-followee network like homophily, link exchange, follower overlap, tie strength [3] have been used as factors for followership loss. We create a baseline model by only using those features which are from the perspective of user who get unfollowed and we shall compare our model with this baseline model.

Our Model: Doc2vec + features Apart from the features described above, we obtain vector representation of users using the state-of-art Doc2vec model[4]. The word vectors are trained from the dataset of tweets. We feed the user vectors along with the features to feed-forward multi-layer perceptron (MLP) and train using cross-entropy loss for the classification task. We perform a 10-fold cross validation to evaluate our model. We vary the values of K[4] (number of topics in LDA) and other hyper parameters to obtain the best results.

Table 1: Evaluation results.

Models	Accuracy	Precision	Recall	F1-score	ROC-area
Baseline Model	61%	0.65	0.70	0.67	0.62
Our Model	**73%**	**0.73**	**0.87**	**0.80**	**0.71**

Results: Table 1 summarizes the results. Our model significantly outperforms the baseline model by 19.67% (w.r.t accuracy), 33.8% (w.r.t precision) and 14.3% (w.r.t recall). To understand which features are discriminative we rank them by their χ^2 values. The top six discriminative features came out to be - *avg. tweet burst time period, max. tweet burst time period, tweet frequency, mention entropy, topic diversity, eigenvector centrality of mention graph*.

Conclusions and implications

In this paper, we identify various socio-linguistic factors behind followership loss and propose a feature-based model for followership loss prediction that achieves a good accuracy of 73% and significantly outperforms the baseline model. The most discriminative factors are related to the users' tweeting behavior - frequency of tweets, their burstiness, the engaging ability of the user and the topic diversity of the user's tweets.

Our research can be helpful for Twitter users in various ways - i) to early identify the followership loss in near future ii) enabling victims to quickly take corrective measures/actions to stop the trend of follower loss and iii) help the Twitter service as a whole to build a “tweet-properly”-like recommendation system for the subscribers to help them avoid unforeseen follower loss.

[4] best result for $K = 30$

REFERENCES

1. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022.

2. Haewoon Kwak, Hyunwoo Chun, and Sue Moon. 2011. Fragile Online Relationship: A First Look at Unfollow Dynamics in Twitter *(CHI '11)*. 1091–1100.

3. Haewoon Kwak, Sue Moon, and Wonjae Lee. 2012. More of a Receiver Than a Giver: Why Do People Unfollow in Twitter? *(ICWSM '12)*.

4. Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. *CoRR* abs/1405.4053 (2014).

5. Sue Moon. 2011. Analysis of Twitter Unfollow: How often Do People Unfollow in Twitter and Why? *(SocInfo '11)*.

6. Seth A. Myers and Jure Leskovec. 2014. The Bursty Dynamics of the Twitter Information Network. *CoRR* abs/1403.2732 (2014).

7. Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.