

Modeling of User's Tweet Behavior to Enhance Profile's Influence

Esraa Almajhad, Abdullatif M. AlAbdullatif, Esam Alwagait,
and Basit Shahzad^(✉)

College of Computer and Information Science,
King Saud University, Riyadh, Saudi Arabia
bent.majhad@hotmail.com,
{amalabdullatif, alwagait}@ksu.edu.sa,
basit.shahzad@gmail.com

Abstract. The communication among the individuals having commonality in interests has been empowered by the emergence of the social media. Twitter is one of the social media platform used by individuals, activist, politicians, academicians and celebrities and is used for diversified purposes. Despite being a brilliant medium in facilitating the process of communication, Twitter still has a gap in targeting specific people, attracting attention, and increasing the opportunity to get more interaction with the user's followers. This is due to the nature of the timeline that presents the recent tweets every time the user logins his account. In this situation the many significant tweeters may go un-noticed. In this paper we have addressed this problem by proposing a system that uses a novel method dedicated to discover the appropriate times to tweet based on the analysis of the active followers' usage behavior. Then, a tool that utilizes this method launches tweets during the time these followers are expected to be active. The tool helps the tweeters to receive the highest percentage of engagement and the proliferation of tweets among their targeted followers by finding the best times to tweet. The results of the experiments show the effectiveness of this system in raising the level of activity and interaction with the user's tweets.

Keywords: Tweeter's behavior · Active followers · Predictive behavior · Online following behavior

1 Introduction

Online social networks have become a social phenomenon that is growing day by day. Along with many other social networks available, Twitter is one of the media that helps in the rapid spread of short messages (so-called Tweets) if properly used; this means reaching thousands of people in a short period of time and at a lower price. The difficulty the tweeter sometimes faces is the dynamics of the followers' timelines [1–3]. With the ability to follow a large number of tweeters, the timelines of the followers are becoming much longer, which will decrease the possibility of interaction and paying attention to all tweeters' tweets. This may occur because the followers focus their attention on the tweets that appear in the visible region and display the last events in

their timelines without paying attention to reading other tweets in the middle. This interview asked people about the tweets they are interested to read and interact with when they login to their accounts on Twitter. The result was that 86 % of the answers are limited to the latest news that appear in the visible region. What we need is a study to pinpoint the “best” time window to tweet. This time window will have most of the active followers connected to a tweeter. This time window will be classified by the followers' activities.

The focus of this paper is to develop a system that finds out the best times to launch tweets at a particular user level, instead of guessing times. This is achieved through design and implementation of an algorithm that will analyze and mine the user's tweets. We fetch the user's tweets from Twitter API to extract the useful metadata that will help us specify the active followers. Then, we analyze and mine the user's active followers' timelines to identify time windows of their presence. This will help calculate the best times for a particular user with respect to the times when most active followers are online. Finally, when the user types certain tweets, the tool built in our system stores and launches tweets in the best times in order to obtain the maximum rate of interaction from the followers. We demonstrate the efficiency of our system by choosing random samples taking into account they contain activities to identify the peak times. Then, we calculate the rate of communication and activities before and after using the proposed system. Several measures are used to demonstrate the increase that occurs by identifying the peak times to post tweets for a particular user.

2 Related Work

Online social networks are useful for studying collaboration relationships, the structure of groups, learning opinions/or sentiments [4] and have an ability to produce a huge amount of data through review, text, discussion, blogs, news and reactions that will help to understand the people, communities and organizations. Social media therefore assist to establish relational data such as information about a user's friends, people who live in the same area and people who enjoy the same things. These data help to increase the chances of understanding the social world around the users [5, 6] has been identified that Six Degrees.com is the first social media network launched in 1997. Then a series of social media networks emerged such as MySpace, Facebook, YouTube, Skyblog and more [7]. Through these media, the consumer is able to learn from the experiences of others and find more information about a specific product before he/she makes a purchase decision [8]. This research is based on Twitter, a brief introduction and composition of twitter is presented here [8–10].

In 2006, Dorsey launched an online social networking service called Twitter. Since its launch, Twitter gained a great popularity to become the most visited site on the Internet. The number of users in Twitter service reached more than 500 million people in 2012 [8]. Twitter is a real-time and highly social micro-blogging service that allows twitter users to communicate with text messages consisting of maximum 140 characters called (tweets) which appear on a timeline. Despite this limitation of characters, users

can broadcast information including daily activities, give opinions on a particular topic, status, current events and more by tweets [8, 11]. The best time to launch tweets is from 1 pm to 3 pm, while the worst time is between 8 pm to 9 am. In addition, the organization identifies that Friday after 3 pm is the time that must be avoided to post tweets. The results of this study did not target users based on specific time zones similar to what was done in [12]. Alwagait and Shahzad [1] performed a statistical study that aimed to determine the peak times in the Twitter social network. This study targeted fixed time zone (Saudi Arabia). This study was conducted to identify the best time of the maximum number of online followers for a specific sample of Twitter users.

Holmes et al. [13] proposed a tool called HootSuite that works to manage and organize multiple social media networks. This tool was used to launch advertising campaigns and marketing on a number of social media like Facebook, LinkedIn and Twitter from one secure place. This tool allows monitoring more than 100 profiles during one dashboard. Furthermore, it allows adding more than one member for the management and assigns tasks to them. It provides reports and gives an overview of users' accounts showing the extent of the growth of followers and their engagement rate. In addition, it allows the scheduling of 100 messages and posts at specific times. For Twitter, in particular, it has facilitated the work for businessmen to monitor their brands and follow up statistics [14].

Su.pr [15] is a tool that provides URL shortening service to allow deployment on a number of social media networks due to limitations in the number of characters in messages such as those on Twitter. Tweet Reports takes a different approach from the other tools in the collection of information to provide statistics of the ideal time to post tweets. This approach relies on two methods: one method depends on the analysis of the optimal 25 influential user's followers last week. Then, it submits an hourly report including statistics for tweets and retweets [1, 2, 16–21]. The other method relies on the analysis of the tweets containing certain keywords to know the best times. Hence, the user can look for a specific word or name for a brand to discover the best times of circulation per day in the past week. As a result, it offers a report showing the times of most active users who post tweets or retweets related to a keyword. It is a very useful tool, but it can become better if it can also collect statistics about the most influential followers online and when users are active for a certain keyword [22]. Widrich [16] proposed a tool that shows the peak times for Twitter users. It depends on an algorithm to analyze 1000 past tweets and the response to them. This simple tool lacks other criteria, as it does not take into account the user's followers.

From the above discussion it is evident that there are some recent studies focused on identifying the best time to tweet and to identify the times when more followers of a given user are online. However, it is also evident that these studies have been undertaken in different contexts and parameters. This study will describe the best time to tweet considering the local data and hence addressing the social, cultural, and technological habits as well. The study will conclude by suggesting an algorithm, designing and implementing a tool and assuring that the suggestions made in this research are worthwhile for the individuals by validating the tool.

3 Proposed Algorithm

As we mentioned earlier, the Twitter's timeline consists of a set of tweets. This set is constantly increasing with a high number of tweeters and passage of time, which causes problems. Consequently, the probability that a follower interacts with Tweeter's tweets decreases. This is because the follower often focuses on the visible area, which is the first page of the timeline. Meanwhile, the follower does not give attention to the old tweets that do not appear in the visible timeline even when they are of great importance. This problem occurs because of the dynamic nature of the timeline that scrolls down tweets to other pages that become invisible to the follower when the number of tweets increases. To address this phenomenon, we propose an algorithm that works to launch tweets in the visible area of the followers in order to increase the rate of activity and response. This can be done by searching the best times when the user's followers are more active to post their tweets. Unfortunately, in the Twitter world, there may be users who have followers but do not communicate with them. This fact made our algorithm focus its attention only on the followers who interact with the user's tweets by replying or retweeting. These are called *active followers*. Therefore, the method starts with finding the active followers of the user. Then, it divides the timeline for each active follower into Hit/Miss Windows. Hit Windows means the time windows when the followers are active, while Miss Windows means the time windows when the followers are absent (i.e., not active). Consequently, our algorithm finds the common Hit Times of active followers to ensure the posting of tweets in the Hit Windows of the user's followers. Figure 1 displays the timeline of follower (n).

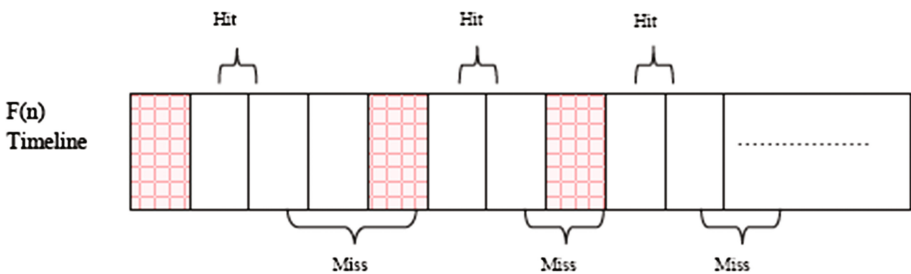


Fig. 1. Follower (n) timeline

Therefore, our algorithm consists of three stages:

- (a) Finding active followers.
- (b) Finding the Hit/Miss rate of time windows for each active follower.
- (c) Finding Global Hit Time of active followers.

3.1 Finding Active Followers

In order to find active followers, we analyze the user's tweets. Every tweet has some metadata associated with it. These metadata contain the creation time of tweets, the user

who posts tweets, the number of retweets, the number of replies, the identifier of the followers who retweet or reply and more. Thus, in order to determine active followers, we rely on two parameters, namely the number of replies and the number of retweets. Socially, each response to the user's tweet has a certain weight. Thus, the user's followers who reply to the user's tweet are more active than those who retweet. As a result, we assume (α) as a weight for each reply and (β) as a weight for each retweet, and then we find the best value for each variable after experiments. Then, we calculate the weight of each active follower through the following formula:

$$W(fn) = (\# \text{ of replies } (fn) * \alpha) + (\# \text{ of retweets } (fn) * \beta) \quad (1)$$

Where

$$\begin{aligned} W(fn) &: \text{The weight of follower } n \\ \# \text{ of replies } (fn) &: \text{Number of replies for follower } n. \\ \# \text{ of retweets } (fn) &: \text{Number of retweets for follower } n. \\ \alpha &: \text{Weight for a reply, } \beta : \text{Weight for a retweet.} \end{aligned}$$

Sometimes we face a problem when the weight of the active follower is too small. This occurs because some followers do not interact with the user significantly. In order to exclude such followers, we need to determine the *most active followers* with a high weight. Therefore, we take a certain percentage from the top weight of the active followers to calculate the smallest weight value that can be regarded as an active follower by the following formula:

$$\text{Minimum value of weight} = \max(w) * X \quad (2)$$

where

$$\begin{aligned} \max(w) &: \text{Maximum value of active followers weight.} \\ X &: \text{The percentage that will be taken from the highest weight of active followers} \end{aligned}$$

This percentage will be considered as a variable to reach the best ratio that can be found to determine the sample of the most active followers.

3.2 Finding the Hit/Miss Rate of Time Windows for Each Active Follower

From the first stage, we have a group of the most active followers. For each of them we identify the Hit/Miss rate of time periods for his/her timeline. Figure 2 represents the timeline for a single active follower. It can be observed that in some of the time slots, the follower is more active, while his/her participation in other slots is non-existent.

Therefore, we can determine that whether the follower is interactive or not based on the total practiced activities in each slot of time such as posting tweets, retweets or

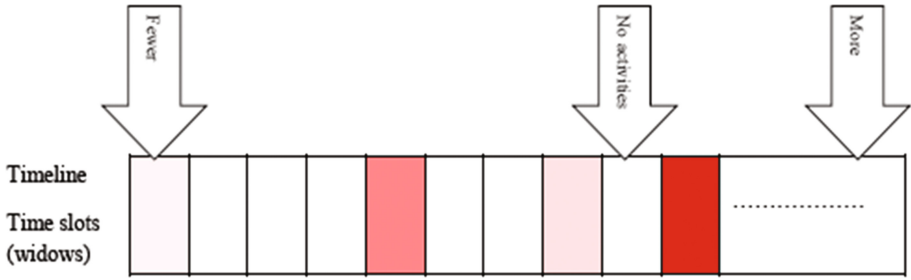


Fig. 2. A single active follower timeline

responding to them. All these activities are considered equal as far as the weight is concerned. Consequently, we can classify the time periods of active followers into: high interaction, called *Hit* period and low or non-existent interaction, called *Miss* period. We assume that dividing the timeline into equal periods takes (t) seconds. Then, we normalize the value of the total activities in each time period to a scale from zero to one. Figure 3 displays the scale to normalize total activities in each time slot.

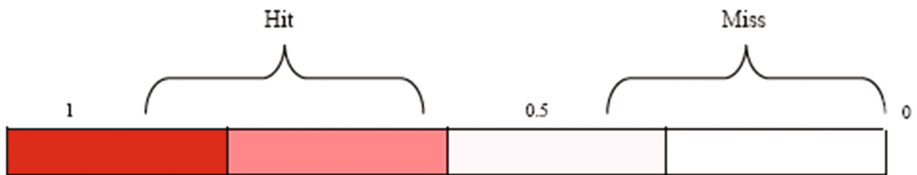


Fig. 3. Normalization scale for total activities in each time slot

All values from 0 to 0.5 are considered *Miss* time periods, while all values greater than 0.5 to 1 are considered *Hit* time periods, but this boundary may be altered after experiments as this is based on the initial assumption. The flow chart for *Hit*/*Miss* rate of time window for each active follower is shown in Fig. 4. This normalization applies to the total activities in each time slot according to the following equation:

$$Tx(Vfn) = \frac{Vfn - \min(fn)}{\max(fn) - \min(fn)} \tag{3}$$

Where

$Tx(Vfn)$: The value from 0 to 1 represents the total activities of follower(n) in time slot(x)

Vfn : Total activities of follower (n) in time slot(x)

$\min(fn)$: Minimum number of total activities of follower(n) over the timeline.

$\max(fn)$: Maximum number of total activities of follower(n) over the timeline.

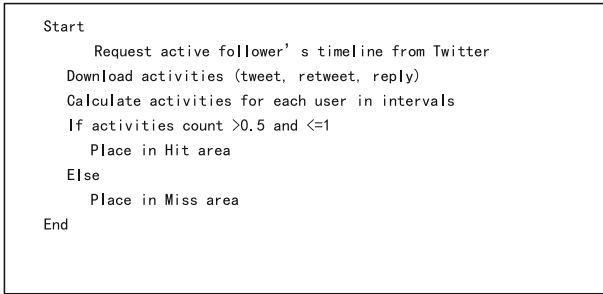


Fig. 4. Finding the Hit/Miss rate of time windows for each active follower

3.3 Finding Global Hit Time of Active Followers

Global Hit Time is a time when most active followers are Hit. From the previous stage, we have the timeline for each active follower segmented into Hit/Miss periods of time. At this stage, we should look for common Hit time periods for all active followers in order to post the tweets in them. However, finding a common Hit time period for all active followers is difficult, especially if the user has a large number of active followers. To solve this problem we rank the most active followers based on their weights. Followers who are most active take a higher value, while followers who are less active take a lower value. Then, the probability of the activities for each time slot $P(t)$ is the summation of the Hit/Miss rate value for a certain time slot multiplied by the rank of each follower. The highest values of $P(t)$ are the Global Hit Time (GHT) periods of active followers.

4 System Design and Implementation

The objective of this proposed system is to develop a tool that can be used in a Twitter platform. This system provides a new way to find the best times to tweet along with posting tweets in these times. The basic functional requirements consist of allowing any Twitter user to register in our system, retrieve the best times that will be calculated by a proposed algorithm. In addition, the proposed algorithm enables the system to send requests to Twitter API to download a particular user's and his/her followers' timelines using an information retrieval component, store the information in the system database, extract statistical data from the queried data, compute the best times and allow a weekly update of this data. Therefore, we need an application that handles the computation of the best times and the creation of tweets in a user-friendly way with good efficiency.

Apache is a most popular Web Server on Linux systems environments that was used in our project. It is used in a combination with the MySQL database engine and (PHP) scripting language. This configuration is called LAMP (Linux, Apache, MySQL and PHP), and forms a powerful and robust platform for the development of web applications [23]. In addition, Twitter API is a rich source that allows third party applications to retrieve data from Twitter. Therefore, in this application, we used a ready-made

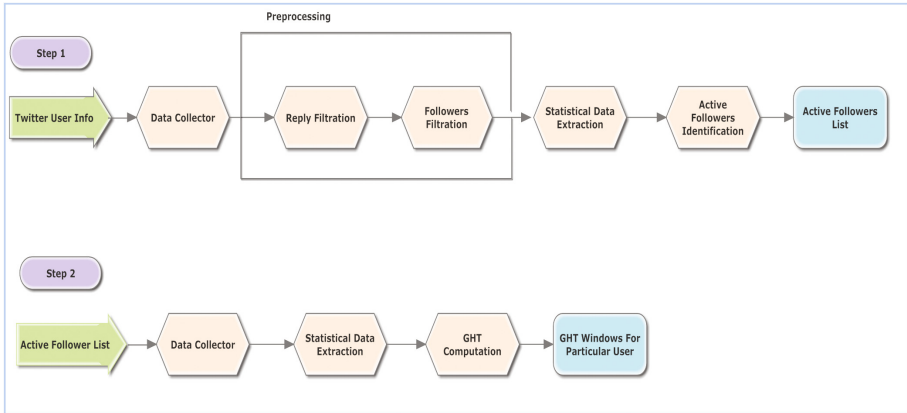


Fig. 5. System implementation flow chart

developed component to connect to Twitter API and retrieve useful information to compute GHT for a particular user. This environment is programmed in Java, and uses RSET API calls to get the required data. Figure 5 displays the sequence of the implementation of the system. This phase is focused to implement the basic objectives as mentioned in Sect. 3.1, 3.2 3.3.

- **Data Collector:** Twitter API allows third party applications to fetch data that aids developers in their applications. The data collector is responsible for requesting data that helps our application to compute GHT. It works to collect user information such as user tweets (user timeline), mention timeline, retweet to each tweet and followers ID. It can also obtain followers' timelines.
- **Reply Filtration:** Focusing on the timeline of a particular user we can determine that it contains tweets that are considered @reply or @mention. The main difference between them is their purpose and their delivery. Typically, @reply is directed to the particular user, but in @mention the username of the particular user may appear but it is not directed to him. Therefore, at this stage the system filters timeline based on the user replied to.
- **Followers Filtration:** It is common to observe that sometimes people do not follow back the twitter user. Yet, they reply and retweet his tweets. Thereby, by this stage the system filters persons who reply or retweet based on user followers ID. Such situations have not been addressed in this study and are considered out of scope.
- **Statistical Data Extraction:** In this stage, the system counts the reply tweets and retweets for each follower to determine active followers. In addition, it counts the activities for each active follower to maintain all possible statistical data that can be generated based on the available information.
- **Active Followers Identification & GHT Computation:** Based on proposed algorithm discussed in Sect. 3, the system calculates active followers for each Twitter user. The status of a follower is determined as proposed in Sect. 3. Also, the system computes the GHT windows for particular users depending on algorithm proposed in Sect. 3.

5 Validation

Our project aims to find the best time for the Twitter user to post his/her tweets with respect to the most active followers to identify the most interactive times. In the previous section, we proposed a solution to this problem in order to help the Twitter user release tweets in the visible region of the active followers to gain the highest level of communication with them. The proposed solution depends on selecting the Global Hit Times (GHT) for active followers of a particular user to launch tweets. Thus, the percentage of activity increases because of the increased response and the retweet rate. In this section, we investigate the effectiveness of the proposed algorithm developed as a tool in our system. This will ensure that we have achieved our objective of demonstrating an increased level of activity and interaction with active followers by targeting the time windows when active followers are online.

6 Conclusion

The results that emerged after using the proposed system gave a good indication of the success of the tool. By considering the more active followers the activity of the tweet viewership increased as the active follower's tweet got more visibility. For that, it cannot rely on the peak times or days for the majority of people because there are special times of the individual achieves high probability of activities, thus guaranteeing greater interaction and engagement with posted tweets. From the results of the experiments it can be deduced that there is a direct relationship between the number of followers and number of active followers. It was also noted that the for some accounts the number of followers increased while the experiments were completed. This may be due to an increase in the activity level of active followers by retweet, which contributed to the spread of tweets significantly and the gaining of new followers. This assumption provides an area of research which can be addressed in the future.

References

1. Alwagait, E., Shahzad, B.: Maximization of tweet's viewership with respect to time. In: 2014 World Symposium on Computer Applications and Research (WSCAR), pp. 1–5 (2014)
2. Alwagait, E., Shahzad, B.: When are tweets better valued? An empirical study. *J. Univers. Comput. Sci.* **20**, 1511–1521 (2014)
3. Shahzad, B., Alwagait, E.: Best and the worst times to tweet: an experimental study. In: 15th International Conference on Mathematics and Computers in Business and Economics (MCBE 2014), pp. 122–126 (2014)
4. Barbier, G., Liu, H.: Data mining in social media. In: Aggarwal, C.C. (ed.) *Social Network Data Analytics*, pp. 327–352. Springer, Heidelberg (2011)
5. Hansen, D., Shneiderman, B., Smith, M.A.: *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann, San Francisco (2010)

6. Adedoyin-Olowe, M., Gaber, M.M., Stahl, F.: A survey of data mining techniques for social media analysis (2013). arXiv preprint [arXiv:1312.4617](https://arxiv.org/abs/1312.4617)
7. Chen, G.M.: Tweet this: a uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Comput. Hum. Behav.* **27**, 755–762 (2011)
8. Gomathi, C., Gowtham, P.: Social media networking state of social media. In: Proceedings of National Conference on New Horizons in IT-NCNHIT, p. 205
9. DNA Kingston Training: Social Media Policy and Procedure. Kingston Training and Employment Pty Ltd, vol. 1 0113 SM SL, January 2013
10. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of social media. *Bus. Horiz.* **53**, 59–68 (2010)
11. Java, A., Song, X., Finin, T., Tseng, B.: Why we Twitter: understanding microblogging usage and communities. In: Proceedings of 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, pp. 56–65
12. Fisher, J.D.: Maximizing your Tweets—Twitter Infographic, May 2014. <http://www.fuseworkstudios.com/maximizing-your-tweets-infographic/>
13. HootSuite: About HootSuit, May 2014. <https://hootsuite.com/company>
14. Comm, J.: Twitter Power 2.0: How to Dominate Your Market One Tweet at a Time. Wiley, Hoboken (2010)
15. StumbleUpon: What features does Su.pr have? May 2014 <http://www.stumbleupon.com/help/business-tools/supr/>
16. Widrich, L.: Top 5 Tools to Better Time Your Tweets, May 2014. <http://mashable.com/2011/10/28/best-time-to-tweet/>
17. Al-Mudimigh, A.S., Ullah, Z., Shahzad, B.: A model for the efficient implementation of portals. In: 2010 2nd International Conference on Computer Engineering and Technology (IC CET), pp. V7-552–V7-554 (2010)
18. Al-Ohali, Y., Al-Oraij, A.A., Shahzad, B.: KSU news portal: a case study. In: International Conference on Internet Computing (ICOMP 2011) (2011)
19. Alwagait, E., Shahzad, B., Alim, S.: Impact of social media usage on students academic performance in Saudi Arabia. *J. Comput. Hum. Behav.* **51**, 1092–1097 (2014)
20. Shahzad, B., Alwagait, A.: Does a change in weekend days have an impact on social networking activity? *J. Univers. Comput. Sci.* **20**, 2068–2079 (2015)
21. Shahzad, B., Alwagait, E.: Utilizing technology in education environment: a case study. In: 2013 Tenth International Conference on Information Technology: New Generations (ITNG), pp. 299–302 (2013)
22. Aronica, J.: 4 Tools To Better Time Your Twitter Updates, May 2014. <http://socialfresh.com/how-to-target-your-twitter-audience-at-the-right-time/>
23. Pitt, L.F., Watson, R.T., Kavan, C.B.: Service quality: a measure of information systems effectiveness. *MIS Q.* **19**, 173–187 (1995)