SAKARYA UNIVERSITY JOURNAL OF COMPUTER AND INFORMATION SCIENCES
VOL. 2, NO. 3, DECEMBER 2019
DOI: 10.35377/saucis.02.03.649708

An Analysis of the Characteristics of Verified Twitter Users

Abdullah Talha Kabakus[1*], Mehmet Simsek[2]
[1]Corresponding Author; Duzce University, Faculty of Engineering, Department of Computer Engineering; talhakabakus@duzce.edu.tr; https://orcid.org/0000-0003-2181-4292; +903805421036
[2]Duzce University, Faculty of Engineering, Department of Computer Engineering; mehmetsimsek@duzce.edu.tr; https://orcid.org/0000-0002-9797-5028; +903805421036

Received 21 November 2019; Revised 15 December 2019 Accepted 30 December 2019; Published online 31 December 2019

Abstract

Twitter, the most popular microblog, contains a large variety of users as a result of its huge popularity. Twitter manually verifies the accounts which are deemed worthy of public interest. As a natural consequence of being verified, users trust these verified accounts since they represent legitimate users, and are managed by authorized users. To the best of our knowledge, Twitter has never revealed the requirements of being verified. In this study, in order to shed light on the characteristics of verified Twitter users, a software, which is based on Python programming language that utilizes a recent dataset, which consists of 297,798 verified Twitter users, was implemented within the scope of this study. The characteristics of verified Twitter users such as being public, and having a customized profile were revealed as a result of the analysis of the utilized dataset.

Keywords: Twitter, microblog, social media analysis, verified users, verification

Doğrulanmış Twitter Hesaplarının Karakteristiklerinin Analizi

Öz

En popüler mikroblog olan Twitter, sahip olduğu devasa popüleritenin sonucu olarak çok çeşitli kullanıcı kitlesine sahiptir. Twitter kamu yararına olacağı inanılan hesapları elle doğrulamaktadır. Doğrulanmanın doğal bir sonucu olarak kullanıcılar, bu hesapların meşru kullanıcıları temsil etmesinden ve yetkili kullanıcılar tarafından yönetilmesinden dolayı bu hesaplara güven duymaktadır. Elde ettiğimiz en iyi verilere göre, Twitter doğrulanmanın gereksinimlerini hiçbir zaman açıklamamıştır. Bu çalışmada, doğrulanmış kullanıcıların karakteristiklerine ışık tutmak amacıyla bu çalışma kapsamında Python programlam dili tabanlı 297.798 doğrulanmış Twitter kullanıcısı içeren güncel bir verisetini kullanan bir yazılım geliştirilmiştir. Bu veriseti üzerinde yapılan analizler sonucunda doğrulanmış kullanıcıların kamuya açık olma, kişiselleştirilmiş bir profile sahip olma gibi ortak karakteristikleri açığa çıkartılmıştır.

Anahtar Kelimeler: Twitter, mikroblog, sosyal medya analizi, doğrulanmış kullanıcılar, doğrulama

1. Introduction

Twitter is the most popular microblog that provides users to post status messages (which are called tweets) up to 280 characters regarding anything they want to share from personal feelings to critical official announcements. Twitter recently revealed that, as of the first quarter of 2019, Twitter has 330 million monthly active users [1]. Another recent official report indicates that users watch 2 billion videos on Twitter per day [2] despite watching videos is the third reason for people to use Twitter after news and photos. According to another report, 500 million tweets are sent each day which means 5.787 tweets are sent per second [3]. As a natural consequence of this popularity, Twitter's audience includes but not limited to regular users, celebrities, company representatives, newsagents, politicians, government agencies, and even country presidents [4]–[6]. Twitter verifies accounts such as newsagents, organizations, and public figures [7] which (1) attract the attention of public interest [8], and (2) whose identities are manually authenticated [9]. Twitter has never explained the requirements of being verified. In order to shed light on the characteristics of verified Twitter users, we have implemented software based on Python programming language that utilizes a recent dataset which consists of 297,798 verified

Twitter users. According to the analysis of this dataset, the characteristics of verified Twitter users were revealed.

2. Material and Method

Each Twitter account is identified by a unique username which is basically textual information that is used to mention other users while posting tweets. In addition to that, Twitter automatically assigns a unique numeric identifier (*a.k.a.* id field) called "*id*". Apart from that, all other information regarding users is optional. Twitter verifies an account when this account represents an official or public property, which could be a person or an organization as well. It is reported that there are a large number of fake/spam accounts [10], [11] despite the serious actions that have already been proposed by Twitter [12]. These fake/spam accounts try to look like official accounts thanks to them misleading their profiles by using the same profile and background picture, same description, and even the same name with the regarding official accounts. When this fact is considered, verifying accounts is a useful practice that lets Twitter users (1) trust that a legitimate source is authoring their tweets [13], and (2) securely contact these verified accounts for whatever reason which could be an information retrieval or conveying their messages, etc. Verified users are identified by a blue "tick" badge alongside their usernames which are available on both the profile page and every user interface that the verified account's tweets are presented as an example of verified Twitter users, Twitter's own profile, is presented in Figure 1. To the best of our knowledge, there is no definition of the requirements of being verified by Twitter. To this end, in order to have a clear idea about the characteristics of verified users, we have utilized an up-to-date Twitter dataset [14] which consists of 297,798 verified Twitter users.

Figure 1 An example of verified Twitter users, Twitter's own profile

Each information that is publicly available on Twitter clients (i.e. web user interface, mobile app, etc.) can be retrieved through the API provided by Twitter [15]. The dataset, which was used within this study, was also constructed thanks to this API, and it contains the features listed in Table 1 for each user. According to the latest documentation of Twitter API [15], the features *contributors_enabled*, *follow_request_sent*, *geo_enabled*, *has_extended_profile*, *is_translation_enabled*, *is_translator*, *notifications*, and *profile_use_background_image* are deprecated, hence they will be returned *null* when queried.

Table 1 The features available in the utilized dataset regarding each user

Feature	Description
contributors_enabled	A flag to indicate that the user has an account with “contributor mode” enabled that allows for tweets issued by the user to be coauthored by another account [16]. A deprecated feature that will always be *null*.
default_profile	A flag that indicates whether or not the user has altered the theme or background of his/her user profile.
default_profile_image	A flag that indicates whether or not the user has uploaded his/her profile image.
description	The description (*a.k.a.* biography) information of the user which is issued by himself/herself.
favourites_count	The number of tweets that were favorited by the user.
follow_request_sent	Indicates a follow request has sent to the user when the friendships are queried. A deprecated feature that will always be *null*.
followers_count	The number of users that follows the user.
following	Queries the user follows the other one. A deprecated feature that will always be *null*.
friends_count	The number of users that the user is following (a.k.a. *followings*)
geo_enabled	Indicates whether or not the user enabled geographic data attachment while posting tweets. A deprecated feature that will always be *null*.
has_extended_profile	A deprecated feature that will always be *null*.
is_translation_enabled	A deprecated feature that will always be *null*.
is_translator	A flag that indicates the user is a participant in Twitter’s translator community [16]. A deprecated feature that will always be *null*.
listed_count	The number of public lists that the user is a member of.
notifications	A deprecated feature that will always be *null*.
profile_use_background_image	A flag that indicates whether or not the user uses a background image in his/her profile instead of the default blank one. A deprecated feature that will always be *null*.
protected	A flag that indicates whether or not the user has chosen to protect his/her tweets.
statuses_count	The number of tweets sent by the user.
url	The website URL of the user.

Each of the features listed in Table 1 was investigated in order to have a clear idea about the characteristics of verified Twitter users. To this end, a software was implemented using Python programming language, which is responsible for (1) loading the data from the dataset, which is stored as an *ndjson* (Newline Delimited JSON) file, (2) creating a dataframe object from the loaded data thanks to the *Pandas* library [17], (3) calculating the insight regarding each feature as a result of analysis of the available data, and (4) draw charts using *matplotlib* library [18] in order to visually represent the experimental result.

The type of analysis of the features depends on the data type of the feature. If the feature is numerical, the minimum, maximum, and mean values were calculated. If the feature is categorical, the choices (which were binary in our case) of users were revealed. The textual features namely *url*, and *description* were converted to categorical through their existences for each user. Hence, the feature *has_url* indicates whether or not the website of the user is defined on his/her profile. In a similar fashion, the feature *has_desc* indicates whether or not the description of the user is defined in his/her profile. The analyzed features and their data types are listed in Table 2.

Table 2 The analyzed features and their data types

Feature	Data Type
default_profile	Categorical
default_profile_image	Categorical
has_desc	Categorical
favourites_count	Numerical
followers_count	Numerical
friends_count	Numerical
listed_count	Numerical
protected	Categorical
statuses_count	Numerical
has_url	Categorical

The dataset which was utilized within the proposed study contains both of tweets and the users who posted these tweets. A sample user data from this dataset is listed in Table 3.

Table 3 A sample user data from the utilized dataset

Property	Value
default_profile	false
default_profile_image	false
favourites_count	89,665
followers_count	18,544
friends_count	2,496
listed_count	112
protected	false
statuses_count	462,764
geo_enabled	true
url	null
verified	false
profile_use_background_image	true
lang	null
notifications	false
contributors_enabled	false

3. Results and Discussion

When the numerical features were analyzed, the characteristics listed in Table 4 were obtained. Verified accounts tend to have a high $followers\_count \div friends\_count$ ratio which is also used to calculate reputations [19] and legitimation of Twitter users [20].

Table 4 The characteristics of verified Twitter users in terms of numerical features

Feature	Minimum	Maximum	Average
favourites_count	0	1,922,026	4,764
followers_count	1	109,581,520	122,921
friends_count	0	4,553,626	2,302
listed_count	0	3,227,621	541
statuses_count	0	23,034,977	15,538

When the categorical features were analyzed, the following conclusions were drawn as the experimental result is visually presented in Figure 2:

- *Verified accounts tend to have website information defined on their profiles* which is reasonable as they represent public properties such as presidents, governments, companies, etc.
- *Verified accounts tend to be public* which is reasonable as (1) they represent public properties, and (2) they aim to reach as many people as possible.
- *Verified accounts tend to use their own background image in their profiles* as a replacement to the blank default one.
- *Verified accounts tend to customize their profile (i.e. profile picture, theme, etc.)* as it is necessary for them to describe themselves well.
- *Verified accounts tend to have a description* as it is necessary for them to describe themselves well (i.e. what is the aim of this profile, etc.).
- According to the experimental result regarding the numerical features, verified accounts tend to have much higher number of followers compared to the number of friends. Similarly, they tend to be listed in the lists that are created by others users in order to ease keep tracking.

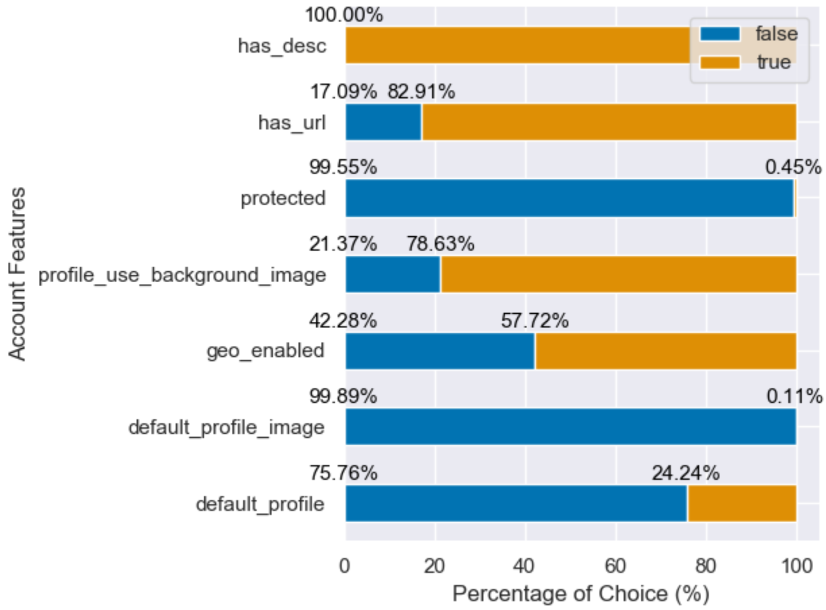

Figure 2 The characteristics of verified Twitter users in terms of categorical features

4. Conclusion

Twitter has a wide range of users, which vary from regular users to even country presidents. Users have the freedom of choosing their username (if it has not been used by someone else before as it should be unique), name, profile and background pictures, as well as defining website information and a brief description. As a natural consequence of this design, parody accounts whose aim is to look like the real (legitimate) accounts for their malicious targets. To this end, Twitter provided a “verification” mechanism to remark the accounts that address the public such as celebrities, organizations, and politicians. But Twitter has never revealed the requirements of this verification mechanism. In order to shed light on the characteristics of these verified accounts, a recent dataset, which consists of 297,798 verified Twitter users, was analyzed thanks to the implemented Python software. According to this analysis, the insights of verified users were revealed in this study. Our analysis showed that verified users have many common features. Verification is an indication of trustworthiness. People adopt the views of people they trust more easily. From this point of view, verified Twitter users are also likely to be influential people.

Referanslar

[1] "Q1 2019 Earnings Report," *Twitter*, 2019. [Online]. Available: https://s22.q4cdn.com/826641620/files/doc_financials/2019/q1/Q1-2019-Slide-Presentation.pdf. [Accessed: 15-Nov-2019].

[2] "Agency Playbook," *Twitter*, 2019. [Online]. Available: https://cdn.cms-twdigitalassets.com/content/dam/business-twitter/resources/Twitter_Agency_Playbook_2019.pdf. [Accessed: 15-Nov-2019].

[3] "2018 Twitter Report," *Mention*, 2018. [Online]. Available: https://info.mention.com/hubfs/Twitter Engagement Report 2018 %7C Mention.pdf. [Accessed: 15-Nov-2019].

[4] A. Pak and P. Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining," in *Proceedings of LREC'10, the Seventh International Conference on Language Resources and Evaluation*, 2010, pp. 1320–1326.

[5] S. Waisbord and A. Amado, "Populist communication by digital means: presidential Twitter in Latin America," *Inf. Commun. Soc.*, vol. 20, no. 9, pp. 1330–1346, 2017.

[6] H. Kwak, C. Lee, H. Park, and S. Moon, "What is Twitter, a social network or a news media?," in *Proceedings of the 19th International Conference on World Wide Web (WWW '10)*, 2010, pp. 591–600.

[7] R. Li, K. H. Lei, R. Khadiwala, and K. C. C. Chang, "TEDAS: A twitter-based event detection and analysis system," in *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, 2012, pp. 1273–1276.

[8] I. Paul, A. Khattar, P. Kumaraguru, M. Gupta, and S. Chopra, "Elites Tweet? Characterizing the Twitter Verified User Network," in *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, 2019.

[9] M. Hentschel, O. Alonso, S. Counts, and V. Kandylas, "Finding Users We Trust: Scaling Up Verified Twitter Users Using Their Communication Patterns," in *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media (ICWSM 2014)*, 2014, pp. 591–594.

[10] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online Human-Bot Interactions: Detection, Estimation, and Characterization," in *Proceedings of the 11th International Conference on Web and Social Media (ICWSM 2017)*, 2017.

[11] C. De Micheli and A. Stroppa, "Twitter and the underground market," 2013. [Online]. Available: https://nexa.polito.it/nexacenterfiles/lunch-11-de_micheli-stroppa.pdf. [Accessed: 16-Nov-2019].

[12] Y. Roth and D. Harvey, "How Twitter is fighting spam and malicious automation," *Twitter*, 2018. [Online]. Available: https://blog.twitter.com/en_us/topics/company/2018/how-twitter-is-fighting-spam-and-malicious-automation.html. [Accessed: 16-Nov-2019].

[13] H. L. LaMarre and Y. Suzuki-Lambrecht, "Tweeting democracy? Examining Twitter as an online public relations strategy for congressional campaigns'," *Public Relat. Rev.*, vol. 39, no. 4, pp. 360–368, 2013.

[14] Jason Baumgartner, "All Verified Twitter Users (100% complete) in ndjson format : datasets," *pushshift.io*, 2018. [Online]. Available: https://www.reddit.com/r/datasets/comments/8s6nqz/all_verified_twitter_users_100_complete_in_ndjson/. [Accessed: 15-Nov-2019].

[15] "API reference index — Twitter Developers," *Twitter*, 2019. [Online]. Available: https://developer.twitter.com/en/docs/api-reference-index. [Accessed: 15-Nov-2019].

[16] B. Erşahin, Ö. Aktaş, D. Kilmç, and C. Akyol, "Twitter fake account detection," in *2nd International Conference on Computer Science and Engineering (UBMK 2017)*, 2017, pp. 388–392.

[17] "pandas: Python Data Analysis Library," 2019. [Online]. Available: https://pandas.pydata.org. [Accessed: 16-Nov-2019].

[18] "Matplotlib: Python plotting — Matplotlib 3.1.1 documentation," 2019. [Online]. Available: https://matplotlib.org. [Accessed: 16-Nov-2019].

[19] T. Velayutham and P. K. Tiwari, "Bot Identification: Helping Analysts for Right Data in Twitter," in *Proceedings - 2017 3rd International Conference on Advances in Computing, Communication and Automation (Fall) (ICACCA 2017)*, 2018, pp. 1–5.

[20] S. Yardi, D. Romero, G. Schoenebeck, and D. Boyd, "Detecting spam in a Twitter network," *First Monday*, vol. 15, no. 1, 2010.