

Report

Twitter Analysis

Submitted By:

IIT2016106 - Aswin VB

IIT2016105 - Aswanth K

IIT2016133 - Utpal Aman


BIM2016501 - Druval CR

IRM2016006 - Sarath Nandimandalam



Table Of Contents

1. Introduction
2. Related Work
3. Proposed Methodology
4. Results and Discussion
5. Conclusion and Future Work
6. References



Abstract: Over the years, use of Online Social Networks (OSNs) has exploded and thus, causing a need of studying and understanding users' behavior online. The excessive use of online social networking causes a great increase in anomalies. Anomalies in OSNs can signify irregular and often illegal behavior. Detection of such anomalies has been used to identify malicious individuals, including spammers, sexual predators, and online fraudsters. For detecting the anomalies dataset of the Twitter network is used and analyzed for user behavior via analyzing their tweets to find whether it is an anomalous or not.

1. Introduction

With a significant increase in the number of active social media users, the chances of an account being fake or anomalous have certainly gone higher. Thus, it is very much needed to identify whether a user is anomalous or not. And if the user is found to be anomalous, one should restrict the user from using the social media any further because that user might be a threat to other genuine users. On the other hand, the user can also be a robot which can be used to manipulate the environment of the social media. Here the social media is twitter and the dataset constitutes of the tweets of different users, their timings of the tweet and the content of the tweet. We have certain aspects that we can use to differentiate between a genuine user and an anomalous user. Also, we have given certain weightage to each of the aspects which eventually help us to calculate the FAL rank of the user based on which we categorize user into two categories, anomalous and genuine.

2. Related Work


Technological advances we have achieved today has equipped us with different techniques and may be a software for large scale computing whether as batch jobs or as real time stream computing jobs. These technologies can be used for anomaly detection in Social Networking websites.

Many research work have been done in the past to detect whether a user is anomalous or not. Principal Component Analysis was used to assess user behavior. Based on this researchers used to decide if a user is anomalous or not. Another challenge was to detect honey-profiles or fake profiles. This technique is applied to detect the ad clicks on Facebook and using it they detected 13% fake users. Another group of researchers used fake profiles to follow the suspected or anomalous users and for this 300 fake profiles were created. In the data collected in a span of over 11 months a lot of spam tweets and spam profiles were found. They studied the behaviour of spam bots as to how they react with the users and other profiles and categorised them into four different categories, Displayer Bragger, Whisperer and Poster. Also, some other researches have been done the link to which have been provided in the references.

After reading all the research papers, we arrived at the conclusion that spammers and other malicious users can be dangerous for any Online Social Networking. They are a threat to users private data(personal information) also. These users must be eliminated for a fair conduct of the OSN which will also help the environment in the OSN to be fair and not manipulated. In order to eliminate such users we need to identify them first. Certain parameters had also been mentioned in the research works previously done from which we picked some and used them by combining all the parameters for a successful anomalous behavior detection Social Network(Twitter).

3. Proposed Methodology

- **URL Ranking**



In this, we fetch the URL that the user has shared in his tweet. We then send this URL on the ALEXA website and fetch the rank of the URL. We count the number of URLs which has their rank less than 2,00,000. Ie, we increment the counter if rank is less than 2,00,000. Now we calculate a ratio based on this, as

Ratio=(Counter/Total number of URLs posted)*10

- **Adult content**

in this, we fetch the URL that the user has shared in his tweet. We made the dataset of all the URLs that are possible to have adult content. If the URL is found to contain adult content then the user is directly declared as an anomaly.

Ratio=10 (If adult content is present)

Ratio=0 (If no adult content is present)

- **Similarity of Tweets**

In this, we check full tweet and not only URL. We get all the tweets and check each tweet with its 3 previous and 3 next tweets. This forms the cluster of 7 tweets. In this cluster, if there exist any two tweets that are 75% similar then we increase the counter. Thus we calculate if there exist any two similar tweets around the particular tweet and calculate the rank.

Ratio=(Counter/Total number of clusters)*10

- **Time Difference**

In this, we check full tweet and not only URL. We get all the tweets and check each tweet with its 3 previous and 3 next tweets. This forms the cluster of 7 tweets. In this cluster, if there exist any 5 tweets that have a time difference of less than 2 minutes i.e. any 5 tweets if have been posted within 2 minutes then we increase the counter. Respective ranks are calculated.

Ratio= (Counter/Total number of clusters)*10

- **WOT Ranking**

In this, we fetch the URL that the user has shared in his tweet. We use the Web Of Trust (WOT) API to check the reputation of the URL i.e. whether it is a good URL or a URL which contains some malware or is not trustworthy. If more than 5% URLs

are found to contain malicious content then the user in this particular category is declared as Anomalous.

Ratio=10 (If 5% of tweets are malicious)

Ratio=0 (If less than 5% of tweets are malicious)

Parameters Weight Assignment and Final Anomaly Level (FAL) Assessment

We assign each of the above parameters a different weight in our formula according to the intensity with which each of them is spamming. Weights are assigned as follows :

1. Time Difference: 0.15
2. Similarity of Tweets: 0.25
3. URL Ranking: 0.30
4. WOT Ranking: 0.30
5. Adult Content: 1

Suppose the ratios obtained from the above methods are :

1. Time Difference: a
2. Similarity of Tweets: b
3. URL Ranking: c
4. WOT Ranking: d
5. Adult Content: e

FAL value calculation :

If e is 10, then $FAL = 10$

Else, $FAL = a*0.15 + b*0.25 + c*0.30 + d*0.30$

As discussed above, we can find out the FAL with the values of a,b,c,d,e parameter. This algorithm is run on a dataset containing twitter handles which will, in turn, produce a dataset of the a,b,c,d,e and FAL values upon which classification algorithms can be applied.

Classification algorithms used are :

- K-nearest neighbors (**KNN**)
- **Naive Bayes** classifier
- Random Forest
- Support Vector Machine (**SVM**)
- Decision Tree

The result of these classification algorithms will be a confusion matrix which is often used to **describe the performance of a classification model** (or "classifier") on a set of test data for which the true values are known.

		predicted class		
		class 1	class 2	class 3
actual class	class 1	True positives		
	class 2		True positives	
	class 3			True positives

```

[[4 0 0]
 [2 0 0]
 [0 0 1]]
Accuracy : 71.42857142857143

```

(A sample 3x3 confusion matrix(Left) and Output confusion matrix(Right))

What can we learn from this matrix?

- There are three possible predicted classes: "Anomalous", "Suspected" and "Non-Anomalous".
- The confusion matrix in Right image consist a total of 7 predictions (e.g., 7 twitter handles were being tested for the presence of Anomaly).
- Out of those 8 handles, the classifier predicted "Anomalous" 6 (4+2) times, and "Non-Anomalous" 1 time.
- In reality, 4 handles in the sample were Anomalous and 1 was Non-Anomalous and 2 were suspected category.

Let's now define the most basic terms, which are whole numbers (not rates):

- **true positives (TP):** These are cases in which we predicted true
- **true negatives (TN):** We predicted False, and they are True.
- **false positives (FP):** We predicted True, but they are False
- **false negatives (FN):** We predicted False but they actually do are False

With these values, we can calculate the accuracy of the classifier,

Accuracy : $(TP+TN)/total = 5/7 = 0.71$

4. Result and Discussion


Dataset Description :

We are using a dataset which contains 600 twitter handles. 100 tweets of each user are fetched by the tweepy library along with its date and time when it was tweeted. The algorithm will create a dataset of the a,b,c,d,e, FAL parameters which are applied to classification algorithms.

Experimental Evaluation :

We have to enter the username which we have to check, this will fetch the last 100 tweets of the user and is given to the algorithm to find FAL value and the result will be displayed

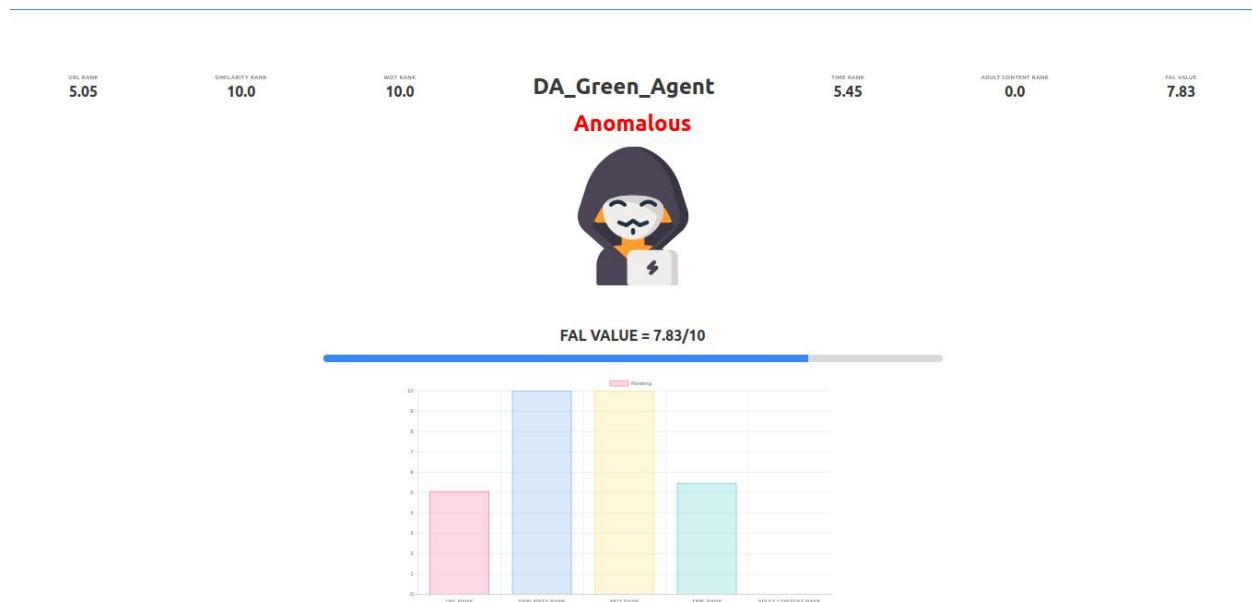
TWEEZY


TWEEZY

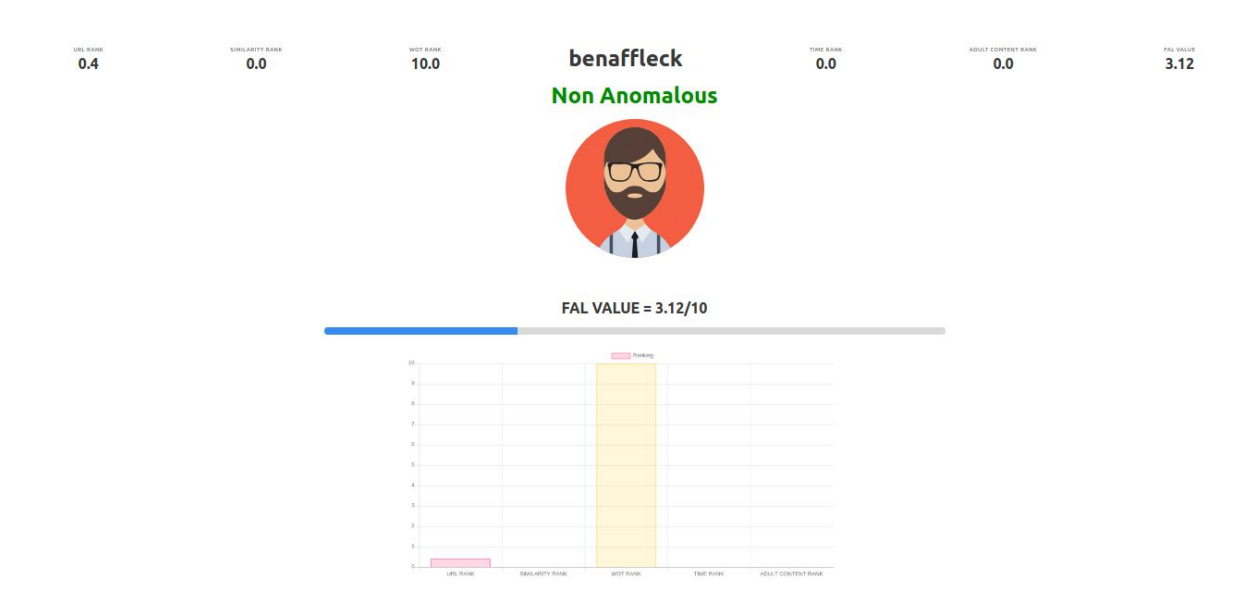
Input twitter username

Submit

(Twitter handle is entered)



(Example for Anomalous User)



(Example for Non-Anomalous User)

For the classification part, the same algorithm is run on a dataset of Twitter handles which will produce a dataset of a,b,c,d,e, FAL values. Which is passed into the various classification algorithm and the confusion matrix along with its accuracy is obtained

```
KNN Classification
=====
[[4 0 0]
 [2 0 0]
 [1 0 0]]
Accuracy : 57.14285714285714

Naive Bayes Classification
=====
[[4 0 0]
 [2 0 0]
 [0 0 1]]
Accuracy : 71.42857142857143

Decistion Tree Classification
=====
[[4 0 0]
 [2 0 0]
 [0 0 1]]
Accuracy : 71.42857142857143

Random Forest Classification
=====
[[4 0 0]
 [2 0 0]
 [0 0 1]]
Accuracy : 71.42857142857143

SVM Classification
=====
[[4 0 0]
 [2 0 0]
 [0 0 1]]
Accuracy : 71.42857142857143
```

(Output of Classifier)

5. Conclusion and Future Work

The approach discussed in this work covers all methods for checking whether a user is anomalous or not. Using the classification algorithm we can predict whether a given Twitter user is anomalous or not. Future works will include adding more methods depending upon the spamming methods which are being used in the future.

6. References

- <https://ieeexplore.ieee.org/document/8204141/references>
- V. Chauhan et al., "Anomalous behavior detection in social networking," *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2017, pp. 1-5.
- G. Stringhini, C. Kruegel, G. Vigna, *Detecting spammers on social networks*, pp. 1-9, Jun. 2010.
- Rose Yu, Huida Qiu, Zhen Wen, Ching Yung Lin, Yan Liu, "A Survey on Social Media Anomaly Detection", *ACM SIGKDD Explorations Newsletter*, vol. 18, no. 1, pp. 1-14, 2016.
- Cao Xiao, David Mandell Freeman, Theodore Hwa, "Detecting clusters of fake accounts in online social networks" in *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, ACM, pp. 91-101, 2015.