Detection of Fake and Clone accounts in Twitter

Swathi K S
4th sem,
Dept of MCA,
JNNCE,Shivamogga,
Karnataka,India.
swathisharath88@gmail.com

Sandhya R
Associate Professor,
Dept of MCA,
JNNCE,Shivamogga,
Karnataka,India
sandhya_r@jnnce.ac.in

Abstract: Online social networks (OSNs) are networking hubs that allow people to interact with real-world relationships. The popularity of OSN is increasing day by day, and the security and privacy issues associated with it are also increasing. Fake and duplicated profiles on all social media pose a dangerous security issue for social network users. User profile duplication is a serious threat of stealing existing user data, creating fake profiles, and exploiting them to destroy the identity of the original profile owner. Even threats such as phishing, stalking and spam. A fake profile is the creation of a profile on behalf of a person whose ID already exists or a company that does not actually exist on social media in order to perform malicious activity. This article suggested a detection method for detecting fake and clone profiles on Twitter. Fake profiles are detected using rules that can effectively classify fake and real profiles. Two methods are used to detect profile cloning. One uses the similarity measure and the other uses the C4.5 decision tree algorithm. Similarity considers two types of similarity: attribute similarity and network relationship similarity. C4.5 checks clones by building a decision tree with information in mind. A comparison is made to see how these two methods work in clone profile detection.

KEYWORDS: *Security,privacy,similarity measures,network relationship*

1. INTRODUCTION

In recent years, the Internet has evolved and smart terminals have become more and more popular. With this in mind, online social networking (OSN) is becoming an important channel for people to get information, disseminate information, make new friends, and have a lot of fun. The complexity of the structure of online social networks, the large, fast and difficult traceability of information generation, the impact of user acceptance, the creation of content for users, group interactions, and the dissemination of information on online social networks. Is social stability, organizational management models, and people's daily lives and lives. Taking Twitter as an example, detecting Twitter spam simplifies the process of analyzing, directing, and monitoring social network events and regulates network management. If the victim was a child, the risk is more dangerous. Profile duplication attacks steal existing user profile information and create fake profiles that are used. It was abused to tamper with the identity of the original owner of the profile. There are two types of profile cloning: profile cloning techniques for the same site and cross-site. When user credentials are obtained from the network to clone on the same network, it is called cloning the same site profile. In cross-site profile replication, an attacker retrieves user information from one network and creates a replication profile on another network where the user does not have an account. As the registration process on social networks has attracted more and more users, the creation of fake profiles has also increased significantly. Attackers create fake profiles to connect with victims and cause malicious activity.

2 . LITERATURE SURVEY

Currently, existing works make it difficult to investigate Twitter spam, especially element selection and location calculations.

1) When determining emphasis, genealogy often chooses an indistinguishable type of quality.[1]Content-based and customer profile-based attributes for detection. Many types of attributes in the informal community of rare customers are unique compared to the attributes of common customers, so it is not enough to accurately convey the state of the information.

2) In [8] calculation decisions, analysts primarily use AI calculations to manage the location of spam in interpersonal organizations. Considering the possibility of characterization, scientists have provided mathematical structural features to distinguish spam clients.

3) The actual dataset of the informal organization shows the long tail effect. H. This is a heterogeneous dataset that contains a wide variety of non-spam that far outweighs spam.[10] Performance suffers when these managed AI calculations are recognized by non-uniform datasets. Similarly, you need to take advantage of multi-tiered attributes and perform calculations prepared to show end-to-end reachability even if the dataset is non-uniform.

3. PROPOSED METHODOLOGY

To make the user's social life safer, we proposed a detection method that can detect both fake and duplicated profiles.

1. Detection of fake profiles

2. Detection of clone profile using similarity measure

3. Clone profile detection using C4.5 algorithm Clone detection using the similarity measure was superior to C4.5 and was able to detect most of the clones supplied to the system

3.1 SYSTEM DESIGN

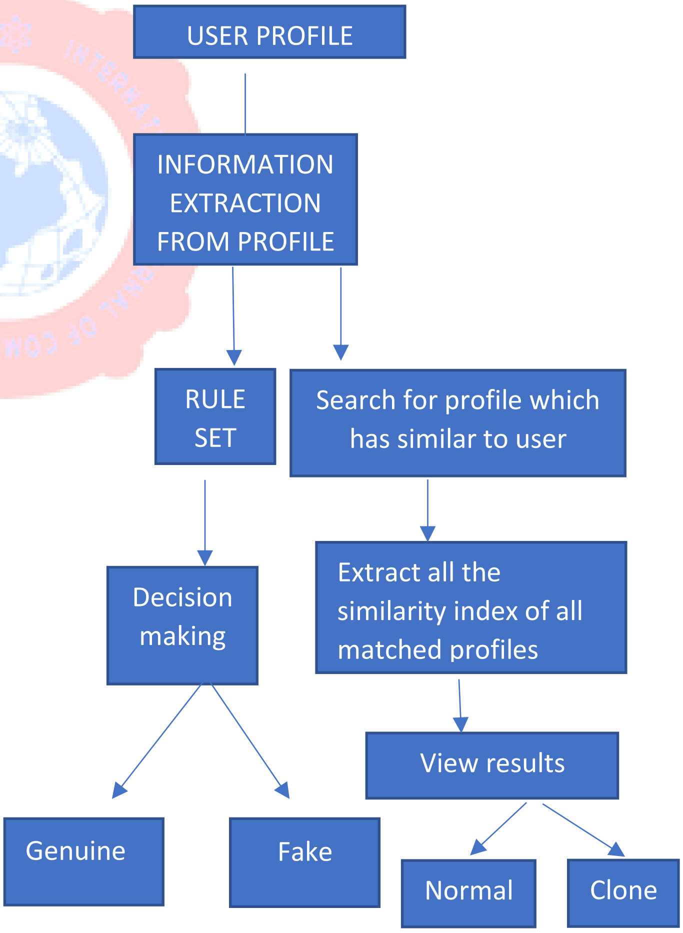
USER PROFILE
INFORMATION EXTRACTION FROM PROFILE
RULE SET
Search for profile which has similar to user
Decision making
Extract all the similarity index of all matched profiles
View results
Genuine
Fake
Normal
Clone

Fig 1 System flow diagram

4. EXPERIMENTAL RESULTS

Here fake and clone account will be detected using the similarity measure algorithm. Modules used here are

Fake Profile Detection

This module is used to detect fake Twitter profiles. Here, fake profiles are detected using rules that effectively distinguish between fake and real profiles. Some of the rules used to detect fake profiles are: Fake profiles usually do not have a profile name or image. Account description is not included. Geo-enabled fields are incorrect because you don't want to reveal your location in tweets. Usually, you may create a large number of tweets, or your profile may not create tweets, etc. The rule is applied to the profile. For each matching rule, the counter is incremented if the counter value is greater than the predefined threshold. The profile is labeled as fake.

Clone Profile Detection using Similarity Measures

This module detects clones based on attributes and network similarities. The user profile is used as input. User identification information is extracted from the profile. Profile that is Searches for attributes that match the attributes of the user profile. If the similarity index is calculated and the similarity index is greater than the threshold, the profile is labeled as follows: Clone, otherwise normal.

Clone Profile Detection using C4.5 algorithm

This module uses the C4.5 algorithm to detect if the specified profile is a clone. C4.5 is the decision tree algorithm used for classification. Build a decision tree based on the given data. At each node of the tree, the attributes that most effectively separate the sample set into subsets are selected. The division factors used in C4.5 are information gain and entropy. Select and determine the attribute with the highest information gain and repeat. Partitioned subtree. The C4.5 algorithm finds similarities between attributes by creating a tree-like structure. The specified profile is compared to a profile that already exists in the database. If the specified profile matches one of the profiles in the database, then that profile is said to be a clone. Otherwise it is normal.

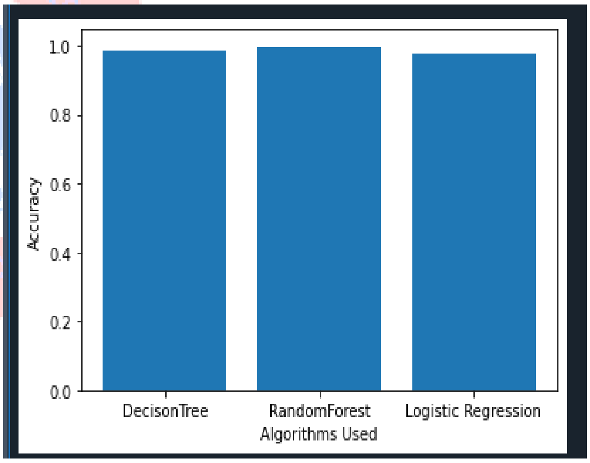

Fig 2 Accuracy graph

5.CONCLUSION

Fake and duplicated profiles have become a very serious problem on online social networks in this project. We hear any threat from these profiles in our daily lives. Therefore, a detection method that can detect both fake Twitter profiles and duplicated Twitter profiles has been proposed. A set of rules was used to identify the fake. This allows you to classify fake and real profiles when in use. Clone detection was performed using the similarity measure and the C4.5 algorithm, and comparisons were performed to confirm performance. Clone detection using the similarity measure was superior to C4.5 and was able to detect most of the clones supplied to the system. In this work, we only considered the forgery and clone detection profile attributes. In the future, we can extend this work to consider tweets by applying some NLP techniques.

6.REFERENCES

[1]Tengetal proposed a self-adaptive, coordinated intrusion detection model built by applying environmental model classes, agents, roles, groups, and objects (E-CARGO). Wu et al.

[2] found that most of the latest spam 36detection methods are based on feature selection and machine learning classifications (DT, RF, NB, etc.). Liuetal.

[3] reviewed the schemes and systems proposed to address more and more cybersecurity threats. In this task, you can extract all the information from the data source and apply analysis / algorithms (such as machine learning) to make decisions in an efficient way. Sunetal.

[4] provided an overview of emerging fields and research prospects. H. Prediction of cyber security incidents. We have also extracted and summarized research methods for predicting cybersecurity incidents at critical stages. In a study by Coulter et al.

[5] A new research method for data-driven cybersecurity (DDCS) has been demonstrated and considered for application to social and internet traffic analysis. DDCS has shown a strong link between data, models, and methodologies during recent work on Twitter spam detection and IP traffic classification key validation. Tang dynasty.