Automatic Identification of Fake Profiles in Online Social Networks

Kumari Pushpam [1], Kumari Sonam[1]
Tapan Kumar Dey [2]

[1] M. Tech, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur
[2] Asst. Professor, CSE Dept., R. V. S College of Engg. & Tech., Jamshedpur

Abstract

In the modern era, the social life of individuals has shifted towards becoming connected with online informal communities. These communities have revolutionized the way we pursue our social life, making it easier to connect with them and keep up with their updates. However, with the rapid growth of these communities, there has been a rise in issues such as false profiles and online bullying. There is currently no practical solution to address these issues. To address this issue, we have developed a system with which the programmed identification of false profiles is feasible and efficient. This system uses order methods such as support vector machine, nave bayes and decision trees to classify the profiles into false or true classes. As this is a software-based recognition method, it can be easily connected to online informal organizations that have a large number of profiles whose profiles cannot be physically inspected. In today's world, users are bombarded with unnecessary information during surfing, which are posted by false users. Research has shown that 20% to 40% of profiles in social networks like Facebook are fake profiles. This detection of fake profiles leads to solution using frameworks.

Keywords: Online Social Network, Fake Profiles, Classification, Natural Language Processing.

1. Introduction

Social media has become the most popular recreational activity on the internet today, with hundreds of millions of people using it and spending billions of minutes doing so. Online social media is a platform where individuals have the opportunity to stay connected with their loved ones, share their news, and connect with others who share similar interests. Social Networks use front end technologies to enable users to create permanent accounts in order to get to know each other. Platforms such as Facebook, Twitter, and other platforms are being developed in tandem with humans to enable users to consult with each other. The online accounts accept people with similar interests, which makes it easier for users to stay connected with existing friends.

The information contained in online profiles will be either static or dynamic. The information that can be provided by the individual at the time of creating the profile is referred to as static knowledge, while the information that can be recounted with the assistance of the network is referred to as dynamic knowledge. Static knowledge involves demographic information about a person and their interests, while dynamic knowledge includes information about a person's runtime habits and location within the network. Most of the current research relies on static and dynamic data. However, this is not applicable to many social networks, where some of the static profiles are easily visible and some of the dynamic profiles are not immediately apparent to the network. A single researcher proposed several approaches to address the issue of fake identities and malicious content on online social networks, each with its own merits and drawbacks.

2. Literature Survey

Halim et al. proposed a stratification of social networks to identify the circle of customers affected by malicious events, using latent semantic analysis. The results of the spatio temporal coincidence were then compared to the original organization/ties associated with the network, which could be encouraging as the organization generated by the spatio time coincidence and the actual one is very close together. Once the worth of threshold was set to the correct level, the number of nodes, or actors, was developed to enable higher photo resolution. Overall, the scan suggests that latent semantic indexing is very effective for detecting malicious content, provided the feature set is chosen competently.

Chai et al. on this paper serves as a testament to the inspiration gained. Although the prototype approach has employed the most efficient normal systems in the context of normal language processing and human-PC interaction, the results achieved from the user trial are significant. By comparing this simple prototype approach with a fully deployed menu procedure, they have found that users, particularly beginner users, prefer the common language dialog-based approach.

Yang et al. proposed a detection method called Invitation which looks at the frequency, rate of outbound and inbound requests, and the clustering coefficient. It was used as a SVM classifier and was 99% accurate. The ground truth showed 1000 legit and 1000 fake accounts provided by Renren.

Additionally, Yang et al. proposed Graph-based features (local clustering coefficient, betweenness centerity, and bi-directional link ratio), Neighbor-based features (e.g. average neighbors' followers), Automation-based features (API ratio, API URL ratio, API Tweet similarity) and Timing-based features to construct various classifiers. The accuracy of the classifier was 86%. The ground truth demonstrated that spam Twitter accounts were defined as those containing malicious URLs 2060 spam accounts.

The Random Forest was designed by Stringhini et Al. and was based on a variety of criteria, including the accepted friend request rate, the URL ratio, the similarity of the messages, the frequency of the selection of friends, the number of messages sent, etc. The accuracy of the algorithm was estimated to be 2% for Facebook and 1% for Twitter. According to Honeypots, there were 173 spam accounts registered for Facebook and 361 for Twitter.

The research conducted by De Cristofaro and Viswanath identifies Facebook like farms through the implementation of honeypot pages. Additionally, the analysis of anomalies in the like behavior of black-market Facebook accounts is used to identify them.

Adikari et al. demonstrate the effectiveness of identifying false profiles on LinkedIn. The paper demonstrates that false profiles can be identified with 84% confidence and a false negative rate of 2.44% when limited profile data is inputted. The methods employed include neural networks, Supervised Verification Method (SVM) and Principal Component Analysis (PCA). Additionally, features such as language proficiency, education, aptitude, recommendation, interest, and awards are employed. Ground truth characteristics of profiles, which are known to be fake, are posted on specialized websites.

3. Methodology

We used a machine learning and natural language processing system on this paper to detect fake profiles in online social networks. We're also adding the SVM Classifier and the Naive Bayes algorithm to improve the accuracy of the fake profiles. In order to identify fake profiles, we have gone through the following steps:

3.1 Collect Data and Pre-process of data:

We select all the attributes that the classification algorithm will use. Carefully select features that do not depend on other features, and those features that can improve the classification efficiency.

3.2 Generate fake accounts:

After selecting the attributes, we need the dataset of the previously identified fake profiles and real profiles for the training purposes of our classification algorithm. We have created a real profile dataset, while the fake profile dataset is provided by a private company that provides network appliance and cloud storage solutions.

3.3 Data Validation to find fake and real:

The attributes selected in Step 1 are necessary to extract the profiles (both fake and genuine) from the database. Social networking companies wishing to implement our system do not need to adhere to the scrapping procedure, as they can extract the features directly from the database. We requested to scrap the profiles, as there is no publicly available social network dataset for the purpose of identifying the fake profiles.

3.4 Create new features:

The dataset of fake profiles and real profiles is then prepared. Of the fake profiles, 80% are used for the training dataset, and 20% for the testing dataset. The training dataset contains 922 profiles, and the testing dataset contains 240 profiles. The classification algorithm is evaluated based on the efficiency of the training dataset.

3.5 Apply Neural Network/ Classification algorithm

Once the training and testing data are ready, the training data is fed into the classification algorithm. The classification algorithm learns from the training data and is expected to provide the right class levels for the test data.

3.6 Evaluate Results

Levels from the test dataset are eliminated and are left to be determined by the classifier trained. The classifier's efficiency is determined by dividing the number of incorrect predictions by the total number of predictions. Three classification algorithms have been employed and the classification efficiency of these algorithms has been compared.

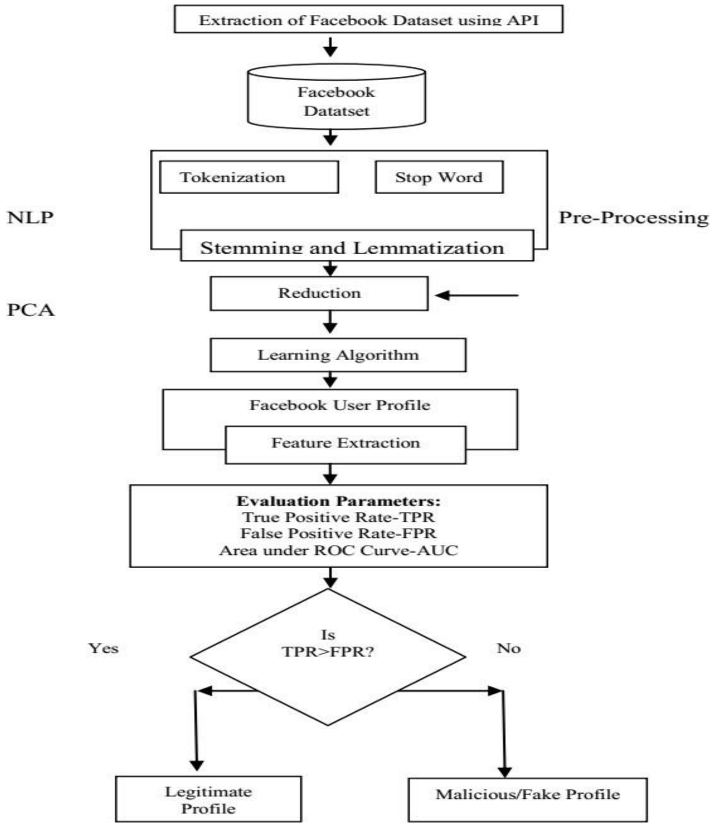

Fig.1: Framework of Proposed System

The proposed procedure consists of three main stages namely NLP Pre-Processing, Principal Component Analysis and Learning algorithms.

A. NLP Pre-Processing:

Pre-processing text is an important part of any Natural Language Processing (NLP) method, and the importance of it is: to reduce the size of the records (or knowledge) of the text content records, and to make the Natural Language Processing (IR) method more efficient and effective.

B. Principal Component Analysis:

The primary purpose of Principal Component Analysis is to acquire the basic knowledge from a table, to represent it as a collection of orthogonal new variables referred to as major accessories and to display a sample of the similarity of observations and variables as components in maps.

C. Learning Algorithms:

In our proposed system, we are using Support Vector Machine and Non-Naive Bayes Algorithms (SVM and NBA respectively). An SVM sorts information by finding an exceptional hyperplane, which separates all information aspects of type 1 from those of type 2. The best hyperplane for any SVM method is the one that has the largest line between the two classifications. An SVM sorts data by finding the exceptional hyperplanes that separate all knowledge aspects of one class from those of another class. Help vectors are the information aspects that are closest to keeping apart hyperplanes.

Naive Bayes is an algorithm that learns the probability that an object with specified characteristics belongs to a specific group/category. It is a probabilistic classification. Naive Bayes is named "naive" because it makes the assumption that the presence of a particular characteristic is independent of the presence of other characteristics. For example, if we are attempting to identify false profiles based on the time, date, or posts, language, and geography, all of these properties contribute to the likelihood that the false profile will be found.

4. Conclusion and Future Work:

4.1 Conclusion

The detection of fake profiles in social networks is achieved through the use of engineered features and the training of the account with machine learning models, such as neural networks and random forests. According to the predictions, the algorithm neural network generated 93% accuracy. It is hoped that in the future, new features that enable easy detection and identification, such as the implementation of skin detection, can be achieved with more accurate natural language processing techniques. Once Facebook introduces new features, it will be possible to identify fake accounts with ease.

4.2 Future Work:

The primary issue is that a person may have multiple Facebook accounts, which allows them to create false profiles and accounts in social media platforms. The concept is to attach an Aadhaar card number when signing up for an account, so that only one account can be created and there is no risk of creating false profiles.

5. References:

[1] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu. The socialbot network: when bots socialize for fame and money. In Proceedings of the 27th Annual Computer Security Applications Conference, pages 93-102. ACM, 2011.

[2] C. Wagner, S. Mitter, C. Korner, and M. Strohmaier. When social bots attack: Modeling susceptibility of users in online social networks. In Proceedings of the WWW, volume 12, 2012.

[3] Egele, M., Stringhini, G., Kruegel, C., Vigna, G., 2015. Towards Detecting Compromised Accounts on Social Networks.

[4]Chu, Z., Gianvecchio, S., Wang, H., Jajodia, S., 2010. Who is Tweeting on Twitter: Human, Bot, or Cyborg?, in: Proceedings of the 26th Annual Computer Security Applications Conference, ACSAC ’10. ACM, NewYork, NY, USA, pp. 21–30. doi:10.1145/ 1920261.1920265.

[5] A. Wang. Detecting spam bots in online social networking sites: a machine learning approach. Data and Applications Security and Privacy XXIV, pages 335-342, 2010.

[6] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In Proceedings of the 4th Workshop on Social Network Systems, SNS, volume 11, page 8, 2011.

[7] Liu Y, Gummadi K, Krishnamurthy B, Mislove A,” Analyzing Facebook privacy settings: User expectations vs. reality”, in: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference, ACM,pp.61–70.

[8] Mahmood S, Desmedt Y,” Poster: preliminary analysis of google?’s privacy. In: Proceedings of the 18th ACM conference on computer and communications security”, ACM 2011, pp.809–812.

[9] Saeed Abu-Nimeh, T. M. Chen, and O. Alzubi, “Malicious and Spam Posts in Online Social Networks,” Computer, vol.44, no.9, IEEE2011, pp.23– 28.

[10] Kazienko, P. and K. Musiał (2006). Social capital in online social networks. Knowledge-Based Intelligent Information and Engineering Systems, Springer.