Identifying Fake Account in Facebook Using Machine Learning

Ahmad Nazren Hakimi(✉), Suzaimah Ramli, Muslihah Wook, Norulzahrah Mohd Zainudin, Nor Asiakin Hasbullah, Norshahriah Abdul Wahab, and Noor Afiza Mat Razali

National Defence University of Malaysia, Kuala Lumpur, Malaysia
nazrennasir90@gmail.my,
{suzaimah, muslihah}@upnm.edu.my

Abstract. Nowadays people rely vigorously on online social networks (OSNs) that have attracted cyber criminals' interest in performing malicious acts. Furthermore, with the existence of illicit businesses that provide transactions of fake account services. This study focuses on identifying fake accounts in Facebook which is the most widely used in OSN. The methodology of this study is started with data collection, features identification and learning classifiers. The first process is to collect information on true and fake Facebook accounts. The second process is the use of Facebook user feed data to comprehend user profile activity and to identify a comprehensive collection of 5 characteristics that play a critical role in discriminating against fake users with true users on Facebook. Lastly, we use these characteristics and the identification of main classifiers based on machine learning that perform well in the assignment of identification out of a total of 3 classifiers namely K-nearest neighbour (KNN), support vector machine (SVM) and neural network (NN). The result shows that KNN generate 82% of the highest performing classifiers with classification precision. The findings have revealed that "likes" and "remarks" add well to the job of detection. However, although the precision value is not highly perfect, the findings of this study shows that most fake accounts are able to imitate actual users.

Keywords: Online social network · Fake account · Machine learning · Facebook

1 Introduction

Nowadays, online social networks (OSNs) has been a compulsory tool that every individual needs in their daily live. There are various OSNs platform that have been introduced such as Facebook, Twitter, Instagram and others that are widely used to share about peoples' daily activities pictures and videos (to name a few). With these kind of platforms, one can communicate with other people without boundaries. Based on studies from the Malaysian Communications and Multimedia Commission, until 2018, the rate of online social network usage in Malaysia has reached 87.4% [1]. This

H. Badioze Zaman et al. (Eds.): IVIC 2019, LNCS 11870, pp. 441–450, 2019.
https://doi.org/10.1007/978-3-030-34032-2_39

finding shows that Malaysian citizens are more likely to communicate and interact with each other through OSNs platform.

OSN is commonly associated with someone identity. According to [2], identity is a distinct object connected to a human being. "Name" is a typical instance of an individual. Every individual has a unique name to represent his or her identity. For instance, passport is also typically used to represent individual's identity. Passport normally contains name, date of birth, address, telephone number, nationality, fingerprints and photograph of the person. Although each person may have several identities, he or she must have a unique one in the sense that the identity is only belonging to him or her.

Recently, there are increasing issues regarding the utilisation of false identities in OSNs [3]. A typical situation for using such false identities is to impersonate someone with the goal to perform several criminal activities in the cyber space such as gathering further information for a spear, personal interest and also spread propaganda or campaign. Besides, the false identity is also used for distributing malware such as phishing attack, spamming and scamming [4]. Thus, the promotion of false identity will lead to the creation of fake account in OSNs, which are against the actual goal of the social network platform.

Instagram and Twitter have wealthy and fully functional Application Programming Interfaces (APIs) for acquiring appropriate, real-time and up-to-date user data. While Facebook APIs facilitates access to profile data such as user operations, friends, colleagues, and most fundamental user details (age, birthday, profile status, relationship status, likes, group details, etc.). Typically, the social network profile consists of two primary components of data: static and dynamic. Former is about the data that is statically set by user, while the dynamic one involves demographics and interests of users, and vibrant information refers to user activity and social network position [5]. These type of data can be detected using machine learning techniques such as K-nearest neighbour (KNN), support vector machine (SVM) and neural network (NN) [6].

As Facebook is one of the most popular OSNs[2], and often used by many people particularly in Malaysia [1], this study tends to identify the fake account of Facebook users using machine learning techniques towards data in the Southeast Asia countries. This is mainly due to the fact that most of Facebook users in these countries have identical demographic data such as time stamp, language and culture.

2 Literature Review

In order to use the service, most OSNs require a user to create a network profile containing their fundamental (sometimes private) data such as name, gender, place, e-mail address, etc. The openness of these social networking sites allow opponents to exploit the service by generating various types of fake profiles to perform illegal, adversarial, unlawful, false or malicious actions such as spamming, promotion and marketing, stalking, intimidation, defamation, etc. Specific reasons for setting false profiles, however, usually rely on the sort of social network being targeted. Adversary generates forged identities on networks such as Facebook and Twitter to access users' private data, endorse a specific brand or individual, or defame a user, etc. They strive to

monitor members' behaviour or gain the confidence of company experts for specialist locations like LinkedIn and Researchgate. Attackers often target dating websites to take advantage of individuals looking for perfect games and working colleagues by playing with their feelings or stealing private data to obtain cash from these customers. One of the most hazardous fake profiles on OSN dating is called Catfisher, a person who utilizes the websites of internet dating to tempt individuals into a scam romance.

According to [7] fake profiles can be splitted into five classifications, namely compromised profiles, cloned profiles, sybil accounts, sockpuppets, and fake bot profiles, as depicted in Fig. 1. Cloned profiles are divided into inter-site cloning and intra-site cloning. While fake bot profiles are splitted into social bots, spam bots, like bots influential bots and bots nets. On various online social networking sites, the five categories can be regarded as the distinct ways adversaries accomplish their ill goals.

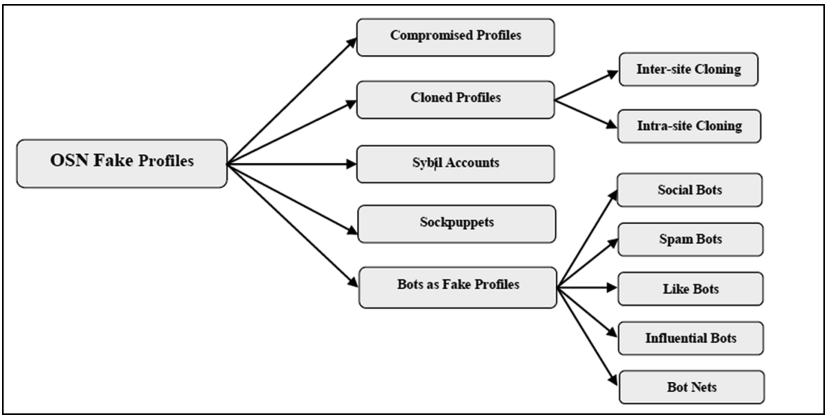

Fig. 1. Type of false online social network profile [7]

There are some attributes for the OSN detection technique that scholars should consider in order to analyse the characteristic that differentiates these profiles from the actual one. Preparation of precise and accurate function set must be prioritised in order to produce an efficient online social network detector. The characteristics can be either manually noted from social network sites or studied using literature study. It is also feasible, however, that some of the characteristics in the literature may not prove to be effective at the moment as opponents continue to change their behaviour to fool and bypass detection systems. Several scientists have recognized various characteristics of internet profiles from time to time in order to train their fake profile detection models [8] and, based on their nature, this study has classified them into 5 groups as follows:

a. **Network-based attributes**
   This attributes shows how fake account connect to their contact according to degree of relation such as first degree is their friend and second degree is for their friend of friends.

b. **Content-based attributes**
This attributes tells about how content based can lead us to detect anomalies activities around the Facebook fake profile.
c. **Temporal features**
This features study about time management of Facebook profile such as time-based activities.
d. **Profile-based features**
This features study about profile based activity like following number of other Facebook accounts, post activities and etc.
e. **Action-based features**
This features study on daily activities that has been performed. It includes how many tag has been posted, location sharing, friends tag and etc.

After studying all of these attributes and features, we manage to detect many of malicious activities that is not tally to real Facebook profile behaviour. This factor also can lead to decision-making rule that we can create during implementation of classification of Facebook fake account.

3 Methodology

3.1 Data Collecting Method

This study requires real-world Facebook datasets which are not openly accessible. There are some social graph datasets available which have profile-based feature data, however such datasets are in anonymised form and are unavailable to be used. Therefore, the study needs to obtain data from the Facebook API, although it is restricted to authorised user. These problems are often cited by authors who working on Facebook such as [6]. As Facebook is constantly updating the security policy on privacy, therefore it is hard to access the data without Facebook's permission [9]. Figure 2 shows the type of data collection techniques.

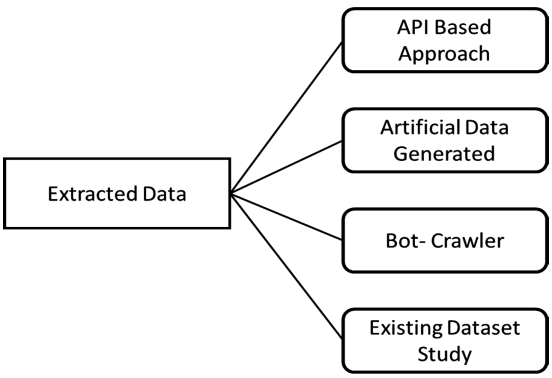

Fig. 2. Data collection technique [7]

According to [7] API-based and bot-crawler approaches are time consuming for data collection techniques and are extremely subject to user privacy and safety environments. To solve a fake account problem in a Facebook, this study utilised an artificial data generated which produce the synthetic data sample based on a network structure or the characteristics of existing datasets. Also, the synthetic data can be produced using different accessible instruments based on any current social network's recognized statistics or parameters. For example, a dummy data set can be generated for analysis purposes if the degree distribution, clustering coefficient, average centrality between the degree and other statistical parameters are known. The generation of artificial or synthetically data can be done by using various online data generators such as GEDIS Studio, Databene Benerator, Mockaroo etc. [10]. This study chooses Mockaroo online generator as the data created is more realistic and similar to the real data [11, 12]. There were 800 sample of data that are successful generated by the Mockaroo which comply with the feature of fake account dataset. Table 1 exhibits the details of the collected Facebook user data.

Table 1. Facebook user data collection

Serial	Description	Value
1	Total user	800
2	Real user	615
3	Fake user	185
4	Assumption real user	560
5	Assumption fake user	240

3.2 Features Identification

Following the collection of the different information characteristics, the next stage is t identify and define a set of characteristics extracted from these data characteristics that would assist as far as possible to distinguish true users and fake users. Finally, a set of 17 characteristics were select out of the different applicants as described in [13], but after a few revision had been done [7, 13–15] the study manage to trace out the most importance characteristics in order to detect fake account as shown in Table 2.

3.3 Learning Classifiers

This study uses monitored machine learning classification algorithms as a final phase in the methodology for detecting false accounts on Facebook. Supervised learners take annotated datasets as input and build predictive models that are used for tasks involving one value prediction using other values in the dataset. In this study situation, the two classes are true users and fake users.

The assumption of using teaching classifiers (as is the strategy followed by many other writers listed in associated work) is that the long-term values of characteristics are likely to differ for true customer accounts and false accounts engaged in multiple anomalous operations.

Table 2. Feature set table with description and intuitive justifications

Serial	Feature name	Feature description	Justification	Measuring methodology
1.	Average Post Likes Received	Average amount of likes a user receives in their own feeds (status, shared messages)	It is anticipated that fake accounts will post and share spam messages that are most likely to have a small count	It is possible to collect information from messages in a user feed like this
2.	Average Post Liked	Average number of posts that the user likes in a user feeds per day	It is anticipated that fake accounts will have a higher activity like ordinary users	This can be gathered from a user's feeds by reviewing the amount of articles where that user contains like information
3.	Average Post Comments Received	Average amount of remarks on the own posts of a user	It is expected that fake accounts will post and share spam messages, with very few comments most likely	Comment information in the feeds of a user can be gathered from the messages
4.	Average Post Comments	Average amount of remarks made by a user in their own feeds on the posts per day	It is anticipated that fake accounts will post and share spam messages, which can take the form of a big amount of comments	This can be gathered from a user's feeds by checking the amount of articles that include the user's comment information
5.	Average Friends	Average amount of friend that connected to user	It is expected that fake account may have huge number of friends rather than normal account	This can be gathered from viewing their friend attribute

K-Nearest Neighbor (KNN). KNN is a technique for classifying objects based on the nearest feature space training examples. One of the simplest of all machine learning algorithms is the k-nearest neighbor algorithm. Training method for this algorithm comprises only of storing the training data 'function vectors and labels. The unlabeled query point is simply allocated to the label of its closest k neighbors in the classification method [16].

Support Vector Machine (SVM). SVM is decision plane ideas that fines a decision's limit. The SVM's objective is to find a hyperplane in the amount of characteristics that clearly classify the data point. It is mainly a classier technique that performs functions in a multidimensional space by building hyperplane that differentiates instances of different class labels. Several constant and different categorical variables can be handled by SVM. SVM supports regression as well as classification [17].

Neural Network (NN). In NN, nodes are linked, sharing their resources to find the most precise outcome, updating the outcome of perception. It is also known as the

connecting computer network, which transmits inner values to each other. It has an input, output and hidden layer where input is where we insert the data, output is what is the outcome and hidden is where neural network learn itself about the dataset to generate output [18].

4 Evaluation

This section involves the assessment of the techniques that has been used to determine the efficiency of detecting fake accounts in Facebook. All of the above classifiers were introduced to a blended datasets consisting of a prior recognized real accounts belonging to the first and second stage of users in the social neighbourhood, as well as the Fake accounts. The dataset also includes user accounts that are friends of colleagues in the social neighbourhood, assumed to be real in active accounts and assumed to be Fake in inactive one. In creating projections for unknown user accounts, this study assessed the capacity of different machine learning classification models (KNN, SVM and NN) using Orange tools.

First step is cleaning dataset for learning classifier. In this clustering concept in Orange tool (Linear Projection and Circular Placement) for clustering is apply to the dataset. The combination of features mention in Table 2 is used to determine clustering process and categories the data into four clusters as follows (Fig. 3):

a. Fake account user (G1)
b. Assume Fake account User (G2)
c. Inactive User (G3)
d. Real User (G4).

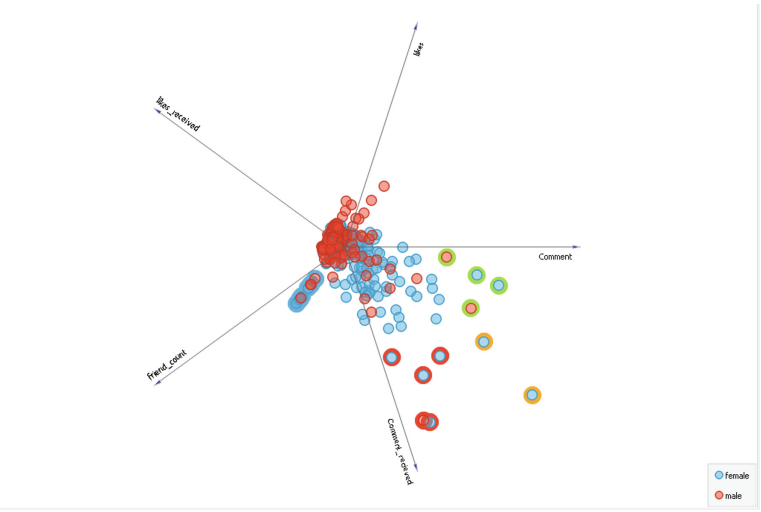

Fig. 3. Clustering process

Second, by using cluster that had been created, the study apply it on classifier learning and use three technique for learning classifier (KNN, SVM and NN) to get the best result of detection. Based on the previous research, the best classifier for detection fake account in Facebook are KNN, SVM and NN [5, 19]. By this it can alleviate the comparison between the three Learning Classifier, which are capable of delivering the best results. Table 3 exhibits the comparison that had been made between these the Classifier:

Table 3. Result of classifier

Serial	Model	Area under ROC curve	Classifier accuracy	Balance of F-score	Precision
1.	KNN	0.967	0.829	0.781	0.760
2.	SVM	0.794	0.729	0.685	0.665
3.	NN	0.958	0.800	0.777	0.772

Based on Table 3, Classifier Accuracy (CA) is the correct fraction of prediction model. If the value is closer to 1, the probability of the model prediction is high. According to the testing dataset, KNN model has the highest CA value (0.829) compared to other models. Figures 4, 5 and 6 depict the results of all the prediction models (KNN, SVM and NN).

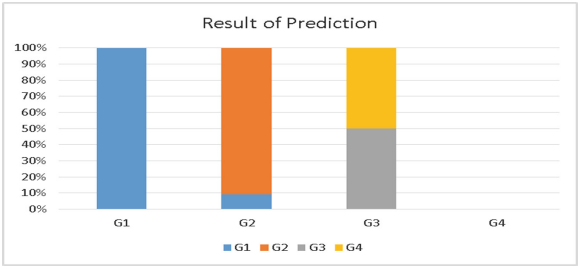

Fig. 4. Prediction result for KNN

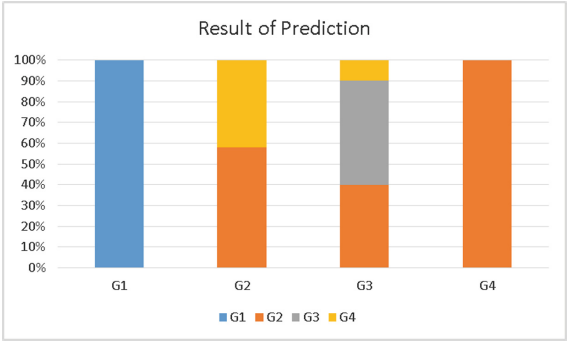

Fig. 5. Prediction result for SVM

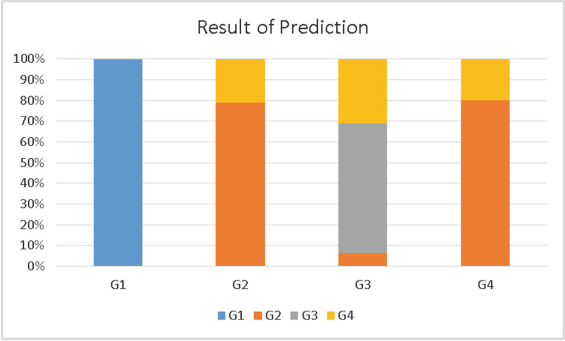

Fig. 6. Prediction result for NN

Figure 4 shows that the prediction of KNN algorithm, which is for G1 group almost all fake account can be detected due to strictly used number of neighbour that had been set. For G2, G3 and G4, up to 70% of detection can be achieved. Referring to Fig. 5, the prediction of SVM algorithm also up to 70% of detection for respective group. Figure 6 explains the result that had been obtain using NN. The result shows that up to 70% detection had been done. In a nutshell, all classifiers provided 70% precision of detection and 30% error rate.

5 Conclusion

Over the years, fake accounts have evolved constantly to avoid their detection. It is therefore essential to create methods to detect the false accounts. Based on the user profile operations and communication with other users on Facebook, this study reveals the fundamentals endeavour to detect the fake accounts in Facebook based on users from Southeast Asia countries. The study used artificial generated dataset for Facebook features as the fine-grained privacy settings on Facebook posed a major challenge to the collection of data. Then, the most frequently used machine learning classification methods are used to identify the highest classifiers. Future research is recommended to utilise hybrid approach on detecting fake account. Other characteristic parameters that can be used to detect fake account such as account ID, location data, devices that is used as a tool to browse social media also should be considered in future research.

References

1. MCMC: Statistic Internet usage survey (2018). https://www.mcmc.gov.my/resources/statistics/internet-users-survey. Accessed 23 July 2018
2. Romanov, A., Semenov, A., Mazhelis, O., Veijalainen, J.: Detection of fake profiles in social media - literature review. In: WEBIST, pp. 363–369 (2017)

3. Kumbhar, A., Wable, M., Nigade, S., Darekar, K., Student, B.E.: A survey on: malicious application and fake user detection in Facebook using data mining. Int. J. Eng. Sci. Comput. **7**(12), 15768 (2017)
4. Guess, A., Nagler, J., Tucker, J.: Less than you think: prevalence and predictors of fake news dissemination on Facebook. Asian-Australas. J. Anim. Sci. **32**(2), 1–9 (2019)
5. Rao, P.S., Gyani, J., Narsimha, G.: Fake profiles identification in online social networks using machine learning and NLP. Int. J. Appl. Eng. Res. **13**(6), 973–4562 (2018)
6. Albayati, M.B., Altamimi, A.M.: An empirical study for detecting fake Facebook profiles using supervised mining techniques. Informatica **43**(1), 77–86 (2019)
7. Fire, M., et al.: A sneak into the Devil's Colony - fake profiles in online social networks. J. Supercomput. **5**(1), 26–39 (2018)
8. Ali, A.M., Alvari, H., Hajibagheri, A., Lakkaraju, K., Sukthankar, G.: Synthetic generators for cloning social network data. In: BioMedCom, pp. 1–9 (2014)
9. Facebook Data Policy: (2018). https://www.digitaltrends.com/social-media/terms-conditions-facebooks-data-use-policy-explained/accessed. Accessed 16 Aug 2019
10. Software Testing Help : Top 10 Best data Generatools in 2019. https://www.softwaretestinghelp.com/test-data-generation-tools. Accessed 14 Aug 2019
11. Generated Data: Generated Data about. https://www.generatedata.com/#t2. Accessed 15 Aug 2019
12. No Title9. Mockaroo Realistic Data Generator. https://mockaroo.com/. Accessed 15 Aug 2019
13. Gupta, A., Kaushal, R.: Towards detecting fake user accounts in Facebook. In: ISEA Asia Security Privacy Conference 2017, ISEASP 2017, vol. 1, pp. 1–6 (2017)
14. Feizy, R.: An evaluation of identity in online social networking: distinguishing fact from fiction (2010)
15. Gheewala, S., Patel, R.: Machine learning based Twitter Spam account detection: a review. In: Proceedings of the 2nd International Conference on Computing Methodologies and Communication, ICCMC 2018, pp. 79–84 (2018)
16. Likhon, A.M., Rahman, A.S.M.M., Choudhury, M.H.: Detection of fake identities on twitter using supervised machine learning. Brac University (2019)
17. Kim, J., Kim, B.-S., Savarese, S.: Comparing image classification methods: K-Nearest-Neighbor and support-vector-machines. Appl. Math. Electr. Comput. Eng. 133–138 (2012)
18. Kudugunta, S., Ferrara, E.: Deep neural networks for bot detection. Inf. Sci. (Ny) **467**, 312–322 (2018)
19. Raturi, R.: Machine learning implementation for identifying fake accounts in social network **118**(20), 4785–4797 (2018)