



Article Title: **Instagram Fake and Automated Account Detection: A Review**

Instagram Fake and Automated Account Detection: A Review

Abilash SB¹, Sujitha R²

¹Student, Department of MCA, Narayanaguru College of Engineering, Anna University Chennai.

²Assistant Professor, Department of MCA, Narayanaguru College of Engineering, Anna University.

ABSTRACT

Fake engagement is one of the significant problems in Online Social Networks (OSNs) which is used to increase the popularity of an account in an inorganic manner. The detection of fake engagement is crucial because it leads to loss of money for businesses, wrong audience targeting in advertising, wrong product predictions systems, and unhealthy social network environment. This study is related with the detection of fake and automated accounts which leads to fake engagement on Instagram. Prior to this work, there were no publicly available dataset for fake and automated accounts. For this purpose, two datasets have been published for the detection of fake and automated accounts. For the detection of these accounts, machine learning algorithms like Naive Bayes, Logistic Regression, Support Vector Machines and Neural Networks are applied. Additionally, for the detection of automated accounts, cost sensitive genetic algorithm is proposed to handle the unnatural bias in the dataset. To deal with the unevenness problem in the fake dataset, Smote-nc algorithm is implemented. In this paper investigating various methods used in the existing work for the Instagram fake account detection.

Keywords: Fake Engagement, Machine Learning, Online Social Networks, Instagram, Genetic Algorithm, Smote.

1 Introduction

Online Social Networks (OSNs) like Facebook and Instagram have becoming more and more popular and become the crucial part in Today's World. Beside the usage of OSNs as a medium of communication, they are also used to gain popularity and promote businesses. At the first glance, the popularity of an account is measured by some metrics like follower count or the properties of the shared contents like the number of likes, comments or views. Therefore, users of any might have a tendency to bolster its metrics in an artificial manner to get more benefits from OSNs.

There are some common ways to increase the reputation of an account in social media. These ways can be listed as usage of bots, buying social metrics such as like, comment and follower, and usage of some platforms or networks which enables users to trade metrics. A bot is a piece of software that completes automated tasks over the Internet. By a 2018 study done by Ghost Data, nearly 95 million Instagram accounts are automated. In 2016, bots generated more Internet traffic than humans. Additionally, by the creation of fake accounts, vendors sell likes



Article Title: Instagram Fake and Automated Account Detection: A Review

and followers very easily. For example, a company called IDigic sells 50k followers for only 250 dollars.

All of these actions listed above are inorganic and termed as fake engagement. In other words, fake engagement term covers all types of automated activities such as liking and commenting on posts, following accounts, uploading posts/stories. In addition, buying social media metrics can also be included in fake engagement terminology. The detection of users who inorganically grow its account is significant because inorganic growth makes businesses pay more to users than its worth for advertising, makes advertisers reach to wrong audiences, make recommendation systems work inefficiently, make access to quality services and product harder.

In the literature, there are some works and released datasets about the detection of fake engagement activity itself and the detection of users who engages inorganic activity in OSNs like Facebook and Twitter. The detection of Twitter fake accounts is studied in using support vector machines and logistic regression, in using graph-based methods, in using the joint usage of naive Bayes classifier with entropy minimization discretization.

In , fake account detection on Twitter is studied by applying the GAIN measure for weighting the all features used in the literature for this task and also proposes some security architectures and focuses also on Facebook. The main focus is on the detection of fake followers on Twitter with some machine learning algorithms. There are also some works in which Instagram are studied from the fake engagement perspective. Instagram has become one of the top social platforms. Instagram has reached about 1 billion monthly active user and 2 million monthly advertisers and users like 4.2 billion posts daily. Therefore, it is crucial to preserve the healthy environment in such an important social platform. In fake likes are tried to be determined on Instagram. In this study, the main concern is to estimate what is the probability that a user can like the post of another user based on the network closeness, interest overlap, liking frequency, influencer effect and link farming hashtag effect. Are about the detection of spammy posts and spammy comments on Instagram, respectively.

In Facebook employees studied the detection of malicious accounts from the requests sent on Facebook and Instagram but this method is not applicable with publicly available information because requests are reachable only from Facebook. From all of these studies, it is observed that there is not any work done regarding fake and automated account detection for Instagram with publicly available information; moreover, no publicly available dataset is present required for this analysis. In this work, we collect and annotate fake account and automated account datasets and present a detailed analysis on fake and automated account detection for Instagram using machine learning algorithms and explain the steps required for preprocessing. The dataset is available on <https://github.com/fcakyon/instafake-dataset>. For the rest of the paper, in Section II, fake account detection dataset and features are detailed. In Section III, automated account detection dataset and features and cost sensitive feature selection algorithm is given. In Section IV, implemented classification algorithms are detailed. Section V presents the results and section VI concludes the paper. This section is related with the detection of fake



Article Title: Instagram Fake and Automated Account Detection: A Review

accounts. Fake accounts are the accounts which are used to increase the popularity metrics of other users.

For this reason, they have a tendency to have a high following and low follower counts. Their liking behaviour may look randomly. The absence of profile picture and strange user names are the common characteristics of fake accounts. In this section, the dataset and the selected features for the dataset have been introduced for the detection of fake account. Then, the oversampling method is explained which is necessary for the unevenness in the number of real and fake account in the dataset. For the dataset, 1002 real account and 201 fake accounts has been gathered after extensive manual label, including accounts from different countries and fields. During this gather process, the points that is paid attention are follower and following counts, media counts, media posting dates or Distribution of classes in the fake account dataset is not even. This results in poor performance for the outnumbered class. SMOTE oversampling technique is utilized to increase number of samples for fake accounts. K is chosen as 5 for this work. In the implementation of SMOTE, SMOTE-NC is applied which considers not only the quantity classes but also the categorical classes. After applying oversampling, all classifiers are trained on equal number of training samples per class.

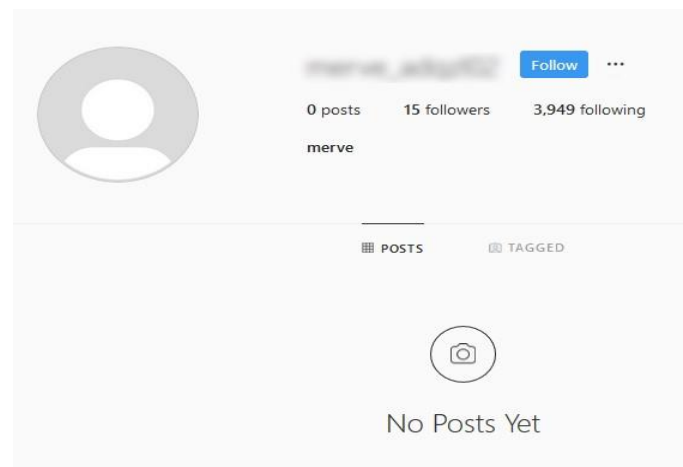


Figure 1: *Fake Account*

2 Related works

2.1 Worth Its Weight in Likes: Towards Detecting Fake Likes on Instagram

In 2018, I. Sen, A. Aggarwal, S. Mian, S. Singh, P. Kumaraguru, ve A. Datta, demonstrated that Instagram is a significant platform for users to share media; reflecting their interests. It is used by marketers and brands to reach their potential audience for advertisement. The number of likes on posts serves as a proxy for social reputation of the users, and in some cases, social media influencers with an extensive reach are compensated by marketers to promote products. This emerging market has led to users artificially bolstering the likes they get to project an inflated social worth. In this study, we enumerate the potential factors which contribute towards



Article Title: Instagram Fake and Automated Account Detection: A Review

a genuine like on Instagram. Based on our analysis of liking behaviour, we build an automated mechanism to detect fake likes on Instagram which achieves a high precision. Our work serves an important first step in reducing the effect of fake likes on Instagram influencer market.

2.2 Identifying Fake Accounts on Social Networks Based on Graph Analysis and Classification Algorithms

In 2018, M. Mohammadrezaei, M. E. Shiri, ve A. M. Rahmani, demonstrated that Social networks have become popular due to the ability to connect people around the world and share videos, photos, and communications. One of the security challenges in these networks, which have become a major concern for users, is creating fake accounts. In this paper, a new model which is based on similarity between the users' friends' networks was proposed in order to discover fake accounts in social networks. Similarity measures such as common friends, cosine, Jaccard, L1-measure, and weight similarity were calculated from the adjacency matrix of the corresponding graph of the social network. To evaluate the proposed model, all steps were implemented on the Twitter dataset. It was found that the Medium Gaussian SVM algorithm predicts fake accounts with high area under the curve=1 and low false positive rate=0.02.

2.3 Machine Learning Implementation for Identifying Fake Accounts In Social Network

In 2018, R. Raturi, demonstrated that, Social Networking is the main era of data transmission as well as data creation in a large scale and the reason why big data is created is very much known. Social networking is the main platform where tons of data is being created and by 2025 even Google data centers can't handle that kind of huge volume. Increase in the fake accounts are creating exponential growth of the volume and in this paper, we are proposing architecture for identifying the fake accounts in the social networking, especially in Facebook. In this research we are using Machine Learning to implement better prediction on identifying the fake account based on their posts and status on their social networking walls. We can consider Facebook and Twitter for this research and for the security purpose and availability of data, we are considering Twitter for this research. Twitter tweets and tagging, type of posts they are creating and sometimes some account people create mess in the society with their posts, identify those and block the fake and unwanted information circulating over the network for the peace and security. Twitter is used as main based with ML NLP for the processing of the text data and here we have to use sentiment analysis for identifying the goals we put on.

2.4 In a World That Counts: Clustering and Detecting Fake Social Engagement at Scale

In 2016, Y. Li, O. Martinez, X. Chen, Y. Li, ve J. E. Hopcroft, demonstrated that How can web services that depend on user generated content discern fake social engagement activities by spammers from legitimate ones? In this paper, we focus on the social site of YouTube and the problem of identifying bad actors posting inorganic contents and inflating the count of social engagement metrics. We propose an effective method, Leas (Local Expansion at Scale), and show how the fake engagement activities on YouTube can be tracked over time by analyzing the temporal graph based on the engagement behavior pattern between users and YouTube



Article Title: Instagram Fake and Automated Account Detection: A Review

videos. With the domain knowledge of spammer seeds, we formulate and tackle the problem in a semi-supervised manner --- with the objective of searching for individuals that have similar pattern of behavior as the known seeds --- based on a graph diffusion process via local spectral subspace. We offer a fast, scalable MapReduce deployment adapted from the localized spectral clustering algorithm. We demonstrate the effectiveness of our deployment at Google by achieving a manual review accuracy of 98% on YouTube Comments graph in practice. Comparing with the state-of-the-art algorithm Copy Catch, Leas achieves 10 times faster running time. Leas is actively in use at Google, searching for daily deceptive practices on YouTube's engagement graph spanning over a billion users.

2.5 Instagram Spam Detection

In 2017, W. Zhang vet H. Sun, demonstrated that in recent years, Instagram has become one of top 15 online social networks. However, popularity of Instagram also causes advertisement and spam posts flooding. Therefore, it is necessary to build a spam detection model to decrease number of spam posts in Instagram. We present a scheme applying feature-based method and supervised learning technique to detect spam posts from Instagram. We use K-fold cross validation to find best pair of supervised learning model and parameters of the model.

2.6 Detecting Spam Comments on Indonesia's Instagram Posts

In 2017, A. Akbar Septiandri vet O. Wibisono, demonstrated a several feature sets for detecting spam comments in social media contents authored by Indonesian public figures. We define spam comments as comments which have promotional purposes (e.g., referring other users to products and services) and thus not related to the content to which the comments are posted. Three sets of features are evaluated for detecting spams hand-engineered features such as comment length, number of capital letters, and number of emojis, keyword features such as whether the comment contains advertising words or product-related words, and text features, namely, bag-of-words, TF-IDF, and fast Text embeddings, each combined with latent semantic analysis. With 24,000 manually-annotated comments scraped from Instagram posts authored by more than 100 Indonesian public figures, we compared the performance of these feature sets and their combinations using 3 popular classification algorithms: Naive Bayes, SVM, and Boost. We find that using all three feature sets (with fast Text embedding for the text features) gave the best F_1 -score of 0.9601 on a holdout dataset. More interestingly, fast Text embedding combined with hand-engineered features (i.e., without keyword features) yield similar F_1 -score of 0.9523, and McNamar's test failed to reject the hypothesis that the two results are not significantly different. This result is important as keyword features are largely dependent on the dataset and may not be as generalisable as the other feature sets when applied to new data. For future work, we hope to collect bigger and more diverse dataset of Indonesian spam comments, improve our model's performance and generalisability, and publish a programming package for others to reliably detect spam comments.



Article Title: Instagram Fake and Automated Account Detection: A Review

3 Conclusion

In this paper, the different strategies for the detecting Instagram fake and automated accounts is studied. Detection of the fake and automated accounts which leads to fake engagement in Instagram is studied as a binary classification problem in this paper. To our knowledge, this is the first time for such an analysis over Instagram accounts. Our contributions with this work are: collection of datasets for fake and automated account detection, proposing derived features for fake and automated classification, proposing a cost sensitive feature reduction technique based on genetic algorithms for selecting best features for the classification of automated accounts, correcting the unevenness in the fake account dataset using the SMOTE-NC algorithm and evaluating several pattern recognition methods over the collected datasets. As a result, SVM and neural network-based methods achieved the most promising F1 score for the detection of automated accounts and neural network achieved. As a future work, recurrent neural networks can be utilized for the time series user data for a better detection of automated accounts. The biased features in the automated account dataset can be balanced by finding the suitable real users. Fake user detector explained in this paper can also be used for finding the suitable real users in the automated account dataset.

References

1. Phillip George Efthimion; Scott Payne; vet N Proffers, Year: 2018, "Supervised machine learning bot detection techniques to identify social twitter bots", SMU Data Science Review, Vol: 1, no: 2, pp. 5.
2. Mohammadreza Mohammadrezaei; Mohammad Ebrahim Shiri; vet A. M. Rahmani, Year: 2018, "Identifying fake accounts on social networks based on graph analysis and classification algorithms", Security and Communication Networks, Vol: 31, no: 2, pp.16 – 117.
3. Ahmed El Azab; A. M. Idrees; M. A. Mahmoud vet; H. Hefty, Year: 2016, "Fake account detection in twitter based on minimum weighted feature set", Int. Sch. Sci. Res. Innova, Vol: 10, no: 1, pp. 13 – 18.
4. Rohit Raturi; Year: 2018, "Machine learning implementation for identifying fake accounts in social network", International Journal of Pure and Applied Mathematics, Vol: 118, no: 20, pp. 4785 – 4797.
5. Stefano Cresci; Roberto Di Pietro; Marinella Petrocchi; A. Sogard vet; Maurizio Tesconi, Year:2015, "Fame for sale: Efficient detection of fake twitter followers", Decision Support Systems, Vol: 80, pp. 56 – 71.
6. A. Akbar Septenary vet; Okiriza Wibisono, Year: 2017 "Detecting spam comments on Indonesia's Instagram posts", Journal of Physics: Conference Series, Vol: 801, no: 01, pp. 012069.