Impersonation in Online Social Networks

Koosha Zarei

Under the supervision of Noel Crespi *and* Reza Farahbakhsh

Institut Polytechnique de Paris, Telecom SudParis

Abstract & Introduction

Impersonators are playing an important role in the production and propagation of the content on Online Social Networks, notably on Instagram. These entities are nefarious fake accounts that intend to disguise a legitimate account by making similar profiles and then striking social media by fake content. Our main **objectives** are: (1) Develop an exclusive crawler to collect necessary data from Instagram aligned with GDPR rules. (2) Focus on the **detection** process of impersonators by developing a **Deep Neural Network** architecture. (3) Leverage advanced NLP techniques to analyse the **behaviour** of these identified fake identities.

Keywords: *Bot; Impersonators; Fake Profile; Fake Content; Fake Engagement; Fake Identities; Instagram; Social Media.*

Research Questions

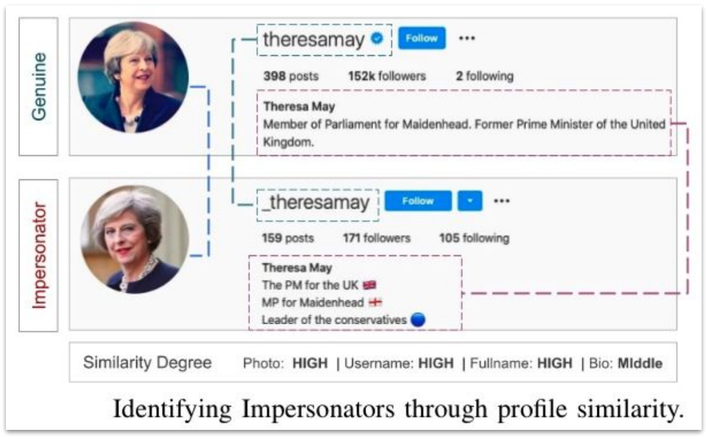
- Who are the impersonators and what are their behaviours? How we can identify impersonator?
- Among them, how many distinct hidden groups exist? What are their characteristics?
- Can we identify impersonator-generated content?

Motivation. We aim to study the behaviour of one important kind of the **fake identities** on OSNs which is called Impersonators. These identities exist in different platforms with various objectives. We characterize them, study their behaviours, analyse their published content and engagements, and eventually we develop a Deep Neural Network to detect them.

Impersonation. Impersonation is where (sometimes malicious) users create social media accounts mimicking a legitimate account. For example, impersonators or imposters maybe accounts that pretend to be someone popular or a representative of a known brand, company.

Lawsuits. Several lawsuits have taken place in the United State (along with other countries), where criminal impersonation is a crime. It involves assuming a false identity with the intent to defraud another or pretending to be a representative of another person or organisation.

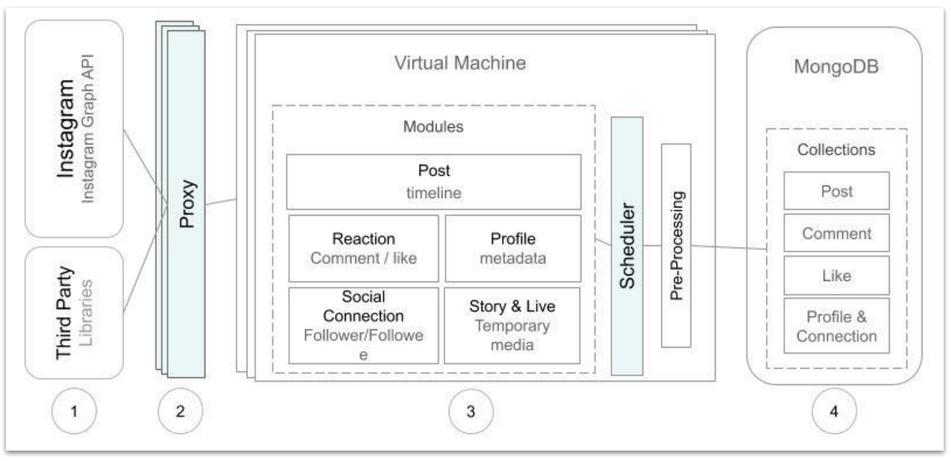
Crawler & Dataset [1, 2]



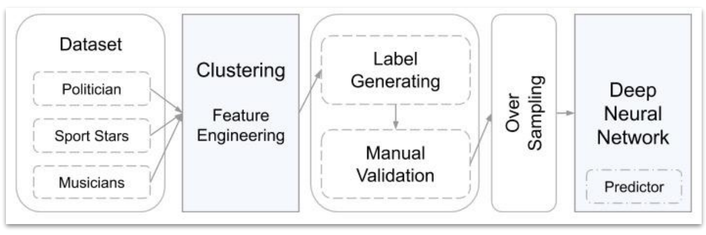
Identifying Impersonators through profile similarity.

Crawler. In order to collect the Instagram public content, we need to design and develop a specific crawler which is able to handle a lot of tasks simultaneously. This crawler connects to Instagram, download various public data content concurrently, perform some NLP based pre-processing steps, and finally store them in a NoSQL format database. In line with Instagram policies and ethical consideration on user privacy defined by the community.

Dataset. In order to perform our research, we continuously crawl and collect data from instagram. We hugely consider posts, comments, likes, profiles, and the relations between accounts. We get this information from different public communities. We start collecting in December 1, 2018, until August 30, 2020. During this period we have collected **3.5M** comments and **13M** likes from **4.2M** public posts from **1.2M** publishers.

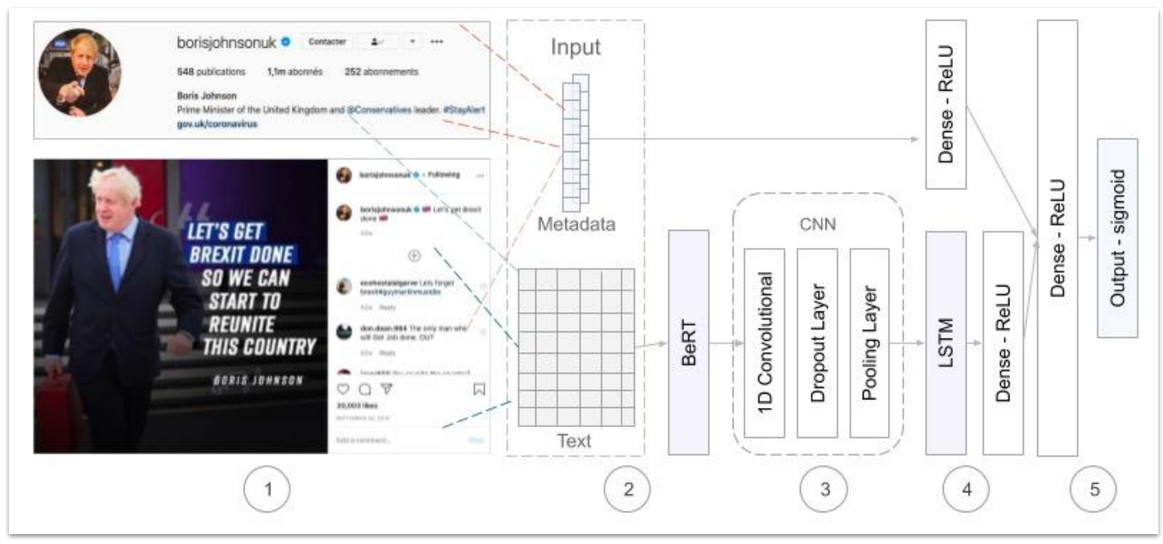


Identification. In order to do that, on Instagram, we compare the account characteristics of public users in various communities (compare to a genuine user) to measure the similarity level. To measure the user similarity, we use profile characteristics Instagram **Username**, **Display name**, and **Bio** are all collections of strings, so we can apply similarity measure algorithms such as Cosine Similarity technique. For matching **Profile Photos**, we leveraged a known library ``Face Recognition" built with deep learning.



Clustering & Engagement. We investigate the published-content and public reactions of impersonators in three major communities such as Politician, News agencies, and Sports stars. We eventually detected **4K** impersonator with different account similarity levels. Then, based on user behaviours and profile characteristics, we cluster them into three groups including **Fans**, **Ordinary users**, and **Bots**. We investigate this by applying clustering algorithms such as **K-Means** on impersonators.

A Deep Neural Approach [3, 4]



Deep Learning. In order to separate the impersonator-generated post from genuine content, we propose a **Deep Neural Network** architecture that measures `profiles' and `posts' metrics to predict the content type of a post: `**bot-generated**', '**fan-generated**', or `**genuine'** content. Our proposed DNN architecture exploits CNN, LSTM, BERT and Dense Layers to process post content and profile metadata

Feature Set. We break the feature list into two principal categories:

- **Post Features.**
  all features that are obtained from the content of the post such as number of likes, number of comments, the caption, etc.

- **Publisher Features.**
  are extracted from the profile of the publisher profile.

Feature Set used in Deep Neural Network.

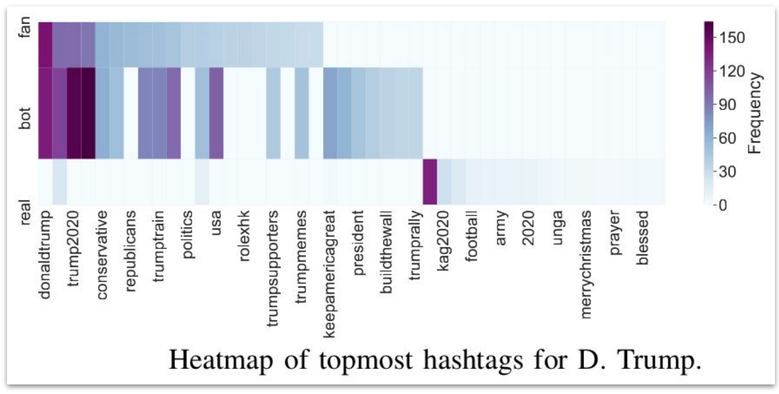
Post Features		Publisher Features	
Feature	**Type**	**Feature**	**Type**
caption text	*text*	similarity username	*numeric*
caption topics (LDA)	*text*	similarity fullname	*numeric*
post hashtag	*text*	similarity bio	*numeric*
tagged users in post	*text*	profile biography	*text*
like count	*numeric*	similarity photo	*numeric*
comment count	*numeric*	follower/followee/post	*numeric*
tagged users count	*numeric*	full name	*text*
mention users count	numeric	biography	*text*
hashtag count	numeric	username	*text*
overall sentiment of caption	numeric	following followers ratio [29]	*numeric*
overall sentiment of hashtag	numeric	followers posts ratio	*numeric*
media type (image or video)	*numeric*	bio emoji count	*numeric*
emoji count	*numeric*	bio hashtag count	*numeric*
url/website exist	*numeric*		*numeric*
date	*numeric*		

Result. First, we do classification using the proposed DNN architecture with only `post content' (CNN + LSTM), and we observe an increase in overall result by nearly 2% (Accuracy 78%). Then we re-run the classifier with both post content and profile metadata (CNN + LSTM). This helps to improve by almost 4.5% (Accuracy 83%). Finally, we add the BERT layer to our architecture (BERT + CNN + LSTM). This step additionally assists us to improve the overall efficiency by almost 4%, and we eventually achieve the accuracy of 86% in detecting post type.

Performance of the proposed architecture

Model	Accuracy	Precision	Recall	F1
Random Forest Classifier	0.76	0.78	0.77	0.76
Proposed DNN (post)	0.78	0.79	0.76	0.78
Proposed DNN (post + profile)	0.83	0.82	0.83	0.82
Proposed DNN (post + profile) + BERT	**0.86**	**0.85**	**0.86**	**0.85**

Impersonator Content. We observe impersonators target some specific events. For example D. Trump generally talks about internal issues such as `*jobs*', `*election*', `*maga*' in 86% of posts. Meanwhile, **Fans** publish relevant issues in 37% of posts. On the other hand, **Bots** talk principally about `*support trump*', `*best president*', `2020 election' in 68% of posts (positive sentiment).



Heatmap of topmost hashtags for D. Trump.

Future Direction & Refs

- An Hybrid Bot Detection in OSNs to detect and identify fake identities and fake content.
- Using BERT and Transfer Learning to increase the overall accuracy.
- Considering comments, stories, Reels and the connection between impersonators.

1. *Koosha. Zarei, R. F, and N. Crespi. Deep dive on politician impersonating accounts in social media. In 2019 ISCC, pages 1–6, 2019 [url].*
2. *Koosha. Zarei, R. F, and N. C. Typification of impersonated accounts on instagram. In 2019 IEEE 38th IPCCC, pages 1–6, 2019 [url].*
3. *Koosha. Zarei, R. F, and N. C. How impersonators exploit instagram to generate fake engagement? In ICC 2020, pages 1–6, 2020 [url].*
4. *Koosha. Zarei, R. F, N. C and G. T. Impersonation on Social Media: A Deep Neural Approach to Identify Ingenuine Content. In 2020 ACM/IEEE ASONAM, 2020 [url].*