Research proposal for Master's thesis

Surendra Singh Gangwar

ABV-Indian Institute of Information Technology and Management, Gwalior, India.

September 30, 2020

1 Tentative title

Suspicious Content and Profile Identification Based on Quantifying Data on Twitter

2 Keywords

- Social network security
- Spam Filtering
- User-Content features
- Natural Language Processing
- Machine Learning

3 Abstract

Twitter is one of the most famous social media platforms and because of its prevalence, spammers discover this stage to spam with clients. Twitter spamming is all the more compromising in light of the fact that its assortment of crowd, twitter clients length over all divisions of life for example it very well may be the educators or understudies, VIPs or politicians, marketers or clients or even the overall population.Because of URL shorteners, common and informal languages and abbreviations used on social networking sites filtering out the malicious content becomes a challenging problem. Industries and researchers have since used different techniques to eliminate spam content from social networking sites. Some of them are based only on user-based features, while others are based on the content-driven features of tweets. In our work, we will try to make a model that will combine both types of features and create some hybrid features and will also used relation based features to classify the content and users.The benefit of using the function of the tweet content is that we can discern the spam tweets regardless of whether the spammer creates another account that was impractical only with the content-based features of the customer and tweet. In this work we will perform the classification, evaluation and comparison of various spam separating strategies and sum up the general situation with respect to the exact pace of various existing methodologies.

4 Hypothesis and Objective

To filter out the malicious content becomes a challenging problem because of URLs shorteners, modern and informal languages, and abbreviations used on social networking sites. Spammers influence the users to click a particular URL or to read the content with specific phrases of words. The thesis would aim to complete the following objectives:

- We will try to develop a system where we can pass real time tweets to classify them as spam or non-spam using user and content based features.
- We will try to classify users for tweets as legit or spam users.
- We will analyze and compare the performances of different methods on our dataset preprocessed using Natural language processing.
- Extend this work to generate the more efficient model by bringing parameter tuning into consideration.

5 Proposed setup for implementation and experimentation

5.1 Programming

Python, Tweepy, Scikit-Learn, Keras,Pandas, Matplotlib, Numpy, NLTK, Word Embeddings.

5.2 Dataset

For our proposed model we will collect all the tweets in the range of 3 months. We will crawl the tweets using Twitter Developer API using Tweepy library. All the tweets will be classified into spam non-spam tweets, later we will try to classify the user as legit user and spam user.

Surendra Singh Gangwar
(2016IPG-107)

Dr. Santosh Singh Rathore
(MTP Supervisor)