Social Media Analytics & Facebook Inspector

Ronak Kumar
Computer Science and Engineering (CSE)
IIIT-Delhi
Okhla Industrial Estate, Phase III, Near Govind Puri Metro Station, Delhi-110020
Email: ronak15080@iiitd.ac.in

***Abstract*—This paper illustrates Social Media Analytics, its tools and techniques along with the use of Facebook Inspector, a browser plugin for real-time detection of malicious content on Facebook.**

***Keywords*—Facebook Inspector, Social Media, Analytics, Machine Learning, Natural Language Processing, API, RSS, Opinion Mining, Sentiment Analysis, Research Paper, LaTeX.**

I. Social Media Analytics

A. Social Media Terminologies

Social Media consists of web and mobile based Internet applications for ubiquitous exchange of data between users. It consists of Posts, RSS Feeds, Blogs, News etc. and plays an eminent role in research related to computational social science that investigates research questions using quantitative methods. Some of the terms used in social media text analysis are :-

Natural Language Processing: NLP is a field in computer science dealing with human computer interactions. It outputs meaningful information by text processing of natural language.

News Analytics: News Analytics is the process of collecting qualitative and quantitative attributes of given data along with determining human sentiment behind the given message.

Scraping: Scraping is the process of collecting raw social media data / unstructured text from various OSN's in either JSON or XML format involving the use of an API.

Text & Social Analytics: T&S Analytics involves the use of techniques like Information Retrieval, NLP, Lexical Analysis for extracting subjective information and predicting opinions in the post.

B. Research Challenges

Social Media analytics and scraping involves a lot of academic challenges for researchers and scientists. Some of them are :-

Data Protection & Cleansing: The unstructured data gathered from scraping poses a lot of challenges for it to being used in research work. Also, once we have big data for analysis, it becomes necessary for us to store it protected as scraping data is against terms and services.

Dashboard Analytics: Most of the social media platforms allow data to be accessed through their API's only. This however is reasonable only for computer scientists only as skills required for getting data are beyond most of the social media researchers.

Data Visualization: Visual representation of data is become increasingly important given the magnitude of available data. Graphical models and statistics should be proposed so as to understand the data properly.

C. Social Media Methodology & Critique

The demerit of using social media for academic research is access to vast amount of data and tools that allow indepth data analysis without being explicitly programmed. Most of the emerging social media resources are monetizing their data however, some companies like Twitter and GNIP have shown modest response for big data access. Research requirements for analyzing the content can be grouped into three different types:

Data: Big Data Research can be classified into four different types: (1) Social Network Media Data (2) News Data (3) Public Data (4) Programmable Interfaces.

Analytics: Social Media researches require explicit programming for their analytics. It is classified into three different types: (1) Analytic Dashboards (2) Holistic Data Analytics (3) Data Visualization.

Facilities: The volume of gathered social media data is huge and so storage of both principal data sources and individual projects needs to be met before. Remotely accessible computational facilities like: (1) Data Protection (2) Hosting Analytics (3) Data Visualization (4) Computational Resources are the essential required facilities.

D. Social Media Data

There is a rise in commericial services that are providing social networking raw data through various methods. In this section, I have described the data types and the formats in which the data is outputted.

1) ***Types of Data****:* The raw datasets can be subdivided into five categories as: (1) Historic Data Sets (2) Real Time Feeds (3) Raw Data (4) Cleaned Data (5) Value-Added Data

2) ***Data Formats****:* The most common formats of raw data available from the API's are: (1) HTML (2) XML (3) JSON (4) CSV

E. Social Media Providers

Social media resources are subdivided into three different types: (1) Freely Available Databases (2) Data Access Via Tools (3) Data Access Via API's. These are described as :-

Open Source Databases: Open Source Databases like Wikipedia provide a majority of all social media data as copies which are majorly used for database queries, social media analytics and solving database queries. The World Bank Data can also be a good example and provides information related to Health across the globe which later is used for computational analysis of posts.

Freely Accessible Sources: Google, being the biggest scraper of content in the world provides tools like Trends & InSights for scraping web content. It also has a range of packages, such as Google Analytics rather than only programmable HTTP API's for analysis. Google Trends, can compare up to five topics at a time for getting which topics have been searched in a particular geographic location.

Commercial Sources: There is an increase in the amount of commercial services which help social media data scraping and provide paid access through their simple analysis tools. Commercial Data Sellers like GNIP, DataSift came into picture after Twitter limiting their public available data. GNIP is regarded as the world's biggest social media data provider as was the first to partner with Twitter to make their data available globally. Also, it has partnered with Tumblr, FourSquare and various other OSN's and delivers social media data to over people in 40 Countries. The company provides detailed information for a particular user or a post using URL Expansion and language detection techniques. It also gives freedom to the user to extract any tweet from 2006 using its Search API and PowerTrack addon.

Data Feed Access via API's & RSS: For computer scientists, the most beneficial sources of getting social media data is via programmable methods uses the help of API's. Here, I have described the API's used in various OSN's along with the use of RSS Feeds for streaming social media data into databases.

1) ***Twitter****:* Twitter API allows researchers to access publicly available data from Twitter in JSON Format. The API is divided into two parts: (1) Search (REST) API (2) Streaming API. REST API queries Twitter Databases for recently posted tweets using specific keywords and hashtags by using an access token. On the other hand, Streaming API provides real time stream of Tweets using a unique User ID, Geographic Location etc. Twitter results are displayed in JSON Format in an array containing fields that are attributes of the given tweet. For Example- **created_at, from_user** are two fields in the JSON structure that depict the posting time of the tweet and the user details of the user who posted it. The JSON response from the API reveals the kind of results generated from the API. The API provides a limiting window of 15 minutes, thus making at most 3,200 calls in that window.

2) ***Facebook****:* Facebook's Graph API is a lot more restricted than Twitter hence it provides much less information about any public message. It stores all the data as objects and has three API's namely Graph, Keyword Insight and Public Feed API. These API's are queried by passing the unique ID / App Access Token to access the properties of an object. The data is returned in JSON Format with the fields representing attributes of a given message.

3) ***RSS Feeds & Blogs****:* RSS Feeds have become a major source of online social media data as majority of the sites provide content access using RSS. RSS (Really Simple Syndication) is a standard followed for publishing updates on a web based content site on an Internet Server making use of a XML File. RSS Feed Reader parses the RSS File, converts it into HTML and displays into layman language. News Feeds are delivered in formats like XML, JSON or CSV. Also, much of the so called **"geospatial social data"** is achieved from mobile devices that generate location based data. The data is divided into four types: (1) Location and Time Sensitive (2) Location Sensitive Only (3) Time Sensitive Only (3) Neither Location or Time Sensitive. GeoRSS is an emerging standard for encoding location in a particular feed using GML (Geography Marking Language) Format. With an increase in number of advanced mobile devices, the geographical identification of a tweet / post can be easily identified using the geotagged attribute in the JSON Structure.

F. Social Media Analytics Techniques

1) ***Sentiment Analysis****:* Sentiment Analysis is an attempt for taking vast amounts of user-generated data and convert into meaningful information. In this section, I have described Analytics using Sentiment Analysis and Text Mining. The general aim here is to evaluate the mood of a writer for a specific topic he is writing about. Automated Sentiment Analysis uses computational statistics along with Machine Learning to output results related to input text using training set and automate the sentiment determining process. Sentiment Analysis is broadly divided into specific subtasks as:

Sentiment Context: It is defined as the process of extracting opinions from the context of the text using various review portals and feeds.

Sentiment Level: It is the process of deciding level / strength of a opinion using the document or sentence.

Sentiment Polarity: It is a method of deciding whether the opinion for a given text is positive, negative or neutral. It is regarded as the most difficult analysis to conduct for a post. The most popular approach for assigning polarity is by using the scores (+1, 0, -1) for positive, neutral and negative opinion respectively. The total polarity score thus is calculated by taking the sum of all the scores from all the opinion words found.

2) ***Supervised Learning Models****:* Khan et al. classified the various popular Machine Learning Techniques for sentiment analysis. These are: (1) Naive Bayes (2) Maximum Entropy (3) Support Vector Machines (4) Logistic Regression (5) Latent Semantic Analysis.

G. Social Media Monitoring Tools

Social Media Monitoring Tools are tools for measuring what people think about a company or its product etc. For Social Media Monitoring, Google presents some free tools like Google Trends, Google Alerts, a detection tool for content change and for providing automatic notifications. It also has

a RSS Feeder namely FeedBurner for manual engagement of posts.

Text Analysis Tools: Natural Language Processing and Text Analysis Tools like OpenAmplifiy and Jodange are used to aggregate thoughts, statements and feelings from social media data. A lot of free tools like Stanford NLP Group Tools and LingPipe are used for linguistic analysis of human written language. Python provides a good collection of libraries for Natural Language Processing, **NLTK** that performs automatic sentiment analysis on given documents.

Data Visualization Tools: The Data Visualization Tools provide capabilities for gaining insights from big data. It helps to identify trends and relationships from the input data. Quick ad hoc visualization on the data can be used to identify patterns in the text for large scale data sets frameworks like Apache Hadoop.

SAS Sentiment Analysis: Sentiment Analysis and Social Media Analytics (SAS) is regarded as the best analytics software for Business Intelligence, Data Management and for predictive analysis. It has automated the process of data scraping from various websites including social media. It identifies trends and emotional changes and combines statistical modeling with linguistics to output accurate sentiment analysis results. SAS has a user-friendly interface for developing models where users can upload sentiment analysis models directly to the server for minimizing manual model deployment. It is used in various OSN's like Facebook or Twitter and other review posting sites like TripAdvisor etc. and has a major role.

II. Facebook Inspector

A. Preface

Facebook Inspector, a REST API based browser plug-in tool for real-time detection of spam content uses class probabilities obtained from two independent supervised learning models to identify malicious content posted on Facebook. These models are based on an extensive feature set, which have an accuracy of over 80% if only used single. It has collected a dataset of about 4.4 Million posts generated from 17 news-making events and has analyzed about 0.97 Million Posts by having an average response time of about 2.6 seconds per post. It has been downloaded 2,500 on both Chrome and Firefox Browsers.

B. Introduction

Social Network Activity rises rapidly during news-making events, natural catastrophes, etc. Ex- The 2014 FIFA World Cup Final saw over 3 Billion Posts on Facebook within a span of 32 days all around the globe. This enormous amount of traffic causes sites like Facebook to become lucrative venues for posting malicious content and spam posts to compromise system reputation and seek monetary gains. Facebook, being the most preferred OSN, becomes the most vulnerable one. Researches depict that Facebook Spammers earn about $200 Million just by posting malicious links during major events. Some prior experiments reveal that spam detection techniques are able to detect less than half of the total number of malicious posts detected by our supervised model. Facebook Inspector does not rely on engagement level of a post i.e likes, comments etc. since these attributes are time dependent and may not be available at zero-hour. Some contributions towards building up this extension are (1) Characterization of spam content during 17 news-making events (2) Excluding engagement level of a post (3) Using a two-fold filtering methodology (4) Publicly available solution in the form of a browser extension.

C. Methodology

There exists a wide range of malicious content on various social networks which includes phishing, advertising, content originating from compromised profiles, click baiting etc. For efficient real time detection of malicious posts on Facebook, our model utilizes specific Facebook Post Features like page category and post type and excludes all the likes, comments and shares for a post for robust detection in of malicious content in real time. Our approach uses data collection using Facebook's Graph API from 17 news-making events and from different domains like political, sport, natural calamities, terror attacks and entertainment news. Facebook does not allow continuous random sample of public posts unlike Twitter and search needs to be repeated every 15 minutes to overcome any drawback caused due to API.

Dataset Creation: Multiple techniques have been used in the past to obtain ground truth data for fake profiles and spam content. These include third-party URL lookup, phishing blacklist, malware, malicious URL detected using CrowdFlower, Mechanical Turk and other human annotation techniques. These techniques are accepted widely, however they may not show correct results when used individually. For Example, URL Blacklists are likely to miss out on click baiting since the domain facebook.com never appears as a URL Blacklist. This content can easily be identified using human annotation making it impossible to detect large amount of posts. Automated techniques like Machine Learning causes further complications like over-fitting. For obtaining ground truth for Facebook Posts, we created two separate datasets as follows :-

1) ***Using URL Blacklists****:* For creating a labeled dataset of posts containing malicious URL, we started by filtering all the posts with more than one URL's. The set of URL's were extracted using the Python's Requests Package. In case of any error, we visited the URL using the LongURL API. Each URL is then subjected to six blacklist lookups like Google Safebrowsing, SURBL, Phishtank, WOT etc. For every domain, the data was passed through the VirusTotal API which categorized the URL as spam, malicious or phishing. The Web Of Trust Score also returns the reputation score for the given domain. A reputation estimate of below 60 indicated unsatisfactory. Also, the WOT Browser requires a confidence value of more than 10 before it presents any warning for the website.

2) ***Using Human Annotation****:* Human annotators were used to obtain ground truth dataset regarding whether a post is malicious or not. This methodology has been followed widely in the past to classify online social media data. We picked up a random sample of 500 posts per event and developed a web interface for the annotation task, thereby assigning unique login credentials for each annotator. All annotators were regular Facebook users and monetary reward of about $4

was given for annotating about 500 posts. They were provided with three options for each post: Post is a spam, Post is not a spam, Can't say. For our analysis, only those posts were selected for which all the annotators agreed upon the same label. After partial agreement posts were discarded, we were left with a final dataset of about 4,412 posts out of which about 571 were spam and 3841 not spam.

D. Analysis

We analyzed the Dataset in three aspects: (1) Textual Content & URL's (2) Entities who post malicious content (3) Metadata associated with the post.

Textual Content & URL's: I found that there are many campaigns in our dataset promoting a particular entity or event. Majority of these campaigns were event specific involving a lot of celebrities and famous personalities. This behaviour reflects attackers preferences of using content of an event for targeting these social media users. Attackers prefer to exploit users curiosity for some news making events along with posting spam content using specific keywords and hashtags. Further investigation on the dataset revealed that most common type of malicious posts about 52% contain URL's pointing to 18+ content which is marked unsafe for children. Second most common type of posts comprised of questionable and negative category URL's as reported by WOT. It accounted for about 45% posts and contains categories with malware, spam, phishing. About 38% posts were from untrustworthy sources on the internet and about 3% of them advertised a phishing URL.

Entities Posting Malicious Content: Facebook Content gets generated using two entities: (1) Users (2) Pages. Pages are public profiles gaining fans from users who like that page. Our dataset detected pages by checking the category field which is unique for pages. Our analysis revealed that pages are more vulnerable sources of posting malicious URL's and constitute about 21% of all the malicious entities. We also found 43 verified pages and only single user who posted malicious content. However, past researches excluded pages and only considered users while generating content reports. Hence, content posted by pages attracts a large audience, thus making them more lucrative sites for posting spam and malicious content.

Metadata: A lot of metadata gets involved for any post like application used for posting, time of the post, content of the post, location of the post etc. Facebook, a universal social network allows users to post content through a variety of platforms. Our analysis reveals that majority of legitimate content about 51% was posted through mobile apps, along with 15% of malicious content. This behaviour reveals that malicious entities primarily use web and third party applications to automate the process of malicious content spread. A significant difference was also observed in the content types. About 50% of legitimate posts with URL were either photos and videos uploaded directly on Facebook. This is regarded as the primary reason for the facebook domain to become the most legitimate domain in our dataset.

E. Challenges

A lot of challenges were faced during the course of our research. First of all collecting raw data from Facebook is itself a challenging task as the API has certain limitations. For FIFA World Cup 2014 Final, out of 3 Billion Posts, only 67,406 posts from the event were fetched from the API. Secondly, the amount of fields of information Facebook provides in very less as compared to other OSN's like Twitter. This amount of information is not sufficient for automating our analysis techniques to work efficiently. A majority of Facebook posts does not include text and also since likes and comments are not used for the analysis, a lot of rich features gets excluded for spam identification. Given, the rate of increase of information spread using OSN's, determining malicious content of OSN's especially for news making events has become of utmost importance. The WOT ratings used in creating the Dataset are obtained from crowdsourcing may be biased. The current architecture that Facebook Inspector uses only involved public Facebook Posts which may result in less detection of malicious posts.

III. Conclusion

This paper illustrates and surveys some of the social media software tools for scraping, data cleaning and sentiment analysis. It also discusses the increase in restricting public available data and depicts how researchers should use computational environments for experimentation. What is needed are public domain data facilities for easy access using cloud based facility. It also illustrates the use of Facebook Inspector, an automatic real time detection tool for malicious content posted on Facebook. In the era of increase use of social media for disseminating information faster than any method, Facebook Inspector can be a great tool for flagging spam and malicious content on Facebook, one of the fastest growing Online Social Network.

Acknowledgment

I would like to acknowledge Michal Galas who led the design and implementation of the UCL SocialSTORM platform with the assistance of Ilya Zheludev, Kacper Chwialkowski and Dan Brown. Dr. Christian Hesse of Deutsche Bank is also acknowledged for collaboration on News Analytics. I would also like to thank Manik Panwar and Bhavna Nagpal for development of Facebook Inspector and collecting survey data and Precog Resarch Group for consistent support.

References

[1] Dewan, P. & Kumaraguru, P. Soc. Netw. Anal. Min. (2017) 7: 15. https://doi.org/10.1007/s13278-017-0434-5

[2] Batrinca, B. & Treleaven, P.C. AI Soc (2015) 30: 89. https://doi.org/10.1007/s00146-014-0549-4