Tracking Sources Of Online Disinformation Project Proposal

Cand.No. 183708
ALEXANDER HOLDERNESS
ah651@sussex.ac.uk
Supervisor: DAVID WEIR

October 15, 2020

With the advent of social media and wide-spread internet access, there has been a shift in the information ecosystem; In the past one might actively seek out their favorite broadsheet and possibly be able to place some trust in the journalistic process. Now, for most to have any presence online it means being exposed to information from possibly unreputable sources, pushing their own brand of news, be it suggested articles from a web browser or more commonly in the form of posts on social media. This makes it difficult to fit in an increasingly IT-centric society without at least some exposure to disinformation. Twitter is one such social media that has been plagued with false information; A 2018 study showed that inaccurate news reached more users than factually correct information, and that these falsehoods often even spread at a faster rate (Vosoughi, Roy, & Aral, 2018). As such Twitter will be a focus of this project, in its aim to track sources of disinformation online.

Projects Aim and Objectives

This project will use various fact-checking sites such as FactCheck, Snopes and GroundNews among others as a source for discovering potentially false claims. By accessing these reported instances of disinformation, queries can be built in Python which will then be run against an index of recent tweets in an attempt to match a false claim with a tweet. If a match is made then that tweet is recorded and checked to see if it contains a link. This link will then be examined to determine if it is the provenance of the purported claim, the result of which will be logged for the goal of monitoring these sources of disinformation.

Primary Objectives

- Investigate the level of disinformation versus accurate information for specific topics on Twitter.

- Research existing work on similar subjects for best practices and potentially novel methods.

- Identify the success cases and shortcomings in existing works.

- Implement a basic finished product early on to pinpoint necessary features, then refine according to the above objectives.

- Implement multiple models that target varying collections of fact-checking sites to see which combination is best suited for the task.

- Identify best way to represent and maintain collection of logged sources of disinformation.

- Investigate both the scope and scale of damage caused by disinformation on Twitter.

Extension Objectives

- Design a user-friendly interface for searching and viewing collections of false claims and their associated sources.

- Research ways of combating sources of online disinformation, making use of the tool or tools created during this project.

- Reach out to fact-checking sites to investigate the potential incorporation of the project's end-product in their fact-checking process.

Relevance

This project comes during a global pandemic and in the wake of a large increase in amount spent on online advertising campaigns for elections at high levels of government. This has meant a move toward an increased level of vigilance from independent watchdogs, governments and the technology firms who are in part responsible for hosting the "fake news"; the project sets out to assist this shift in attitude by coming up with solutions to the issue of tracking sources of disinformation as well as to research the spread of disinformation on social media and its wider impact.

As someone who believes skepticism to be a useful trait in this day and age, I am becoming increasingly encouraged by the gradual adoption of fact-checking tips (checking primary and secondary sources, being conscious of the reputation of the source and general skepticism towards outlandish stories) and the cautious realization of the potential scale of disinformation and by extension misinformation online. As such this project is of a personal interest to me too.

Required Resources

Twitter developer account(s) to be able to access their APIs, specifically Twitter's Search API to return collections of Tweets relevant to a specific disinformation related query.

Institutional account access for overleaf.com, an online LaTeX editor which will be used for the writing and editing of this project.
An account on Trello for handling project management.
A computer with a Python 3.X installation and various related natural language processing, web-scraping and Twitter packages installed, such as but not limited to:
Natural Language Processing: NLTK, spaCy, and cdQA.
Web-Scraping: Requests, Beautiful Soup 4, lxml, Selenium, Scrapy.
Twitter: Tweepy.
System Resources such as CPU, RAM etc. should be covered by most modern machines as this is not a computationally intensive project however if there is a need to scale up the Sussex University lab machines or a cloud-based solution such as Google Colab will most likely be sufficient.

Time Management

Weekly Timetable

This is an initial plan of the hours in a week that will be dedicated to this project - it marks 10 hours total currently and is subject to change depending on the needs of the project. Dedicated hours can be seen below marked in bold.

Time	Monday	Tuesday	Wednesday	Thursday	Friday
9:00		Seminar		Seminar	**Project**
10:00	**Project**		**Project**	Seminar	**Project**
11:00	**Project**	**Project**	**Project**	Lecture	
12:00					Lecture
13:00	Lecture	Lecture		Workshop	
14:00		Lecture	**Project**		
15:00	Lecture		**Project**		
16:00				**Project**	

Bibliography

Chamberlain, P. (2010). Twitter as a vector for disinformation. *Journal of Information Warfare*, *9*(1), 11–17.

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of "fake news" and online disinformation in europe. *Reuters institute factsheet*.

Galitsky, B. A. (2015). Detecting rumor and disinformation by web mining. In *Aaai spring symposia*.

Kumar, K., & Gopalan, G. (2014, 09). Detecting misinformation in online social networks using cognitive psychology. *Human-centric Computing and Information Sciences*, *4*, 14. doi: 10.1186/s13673-014-0014-x

Pierri, F., Artoni, A., & Ceri, S. (2020). Investigating italian disinformation spreading on twitter in the context of 2019 european elections. *PloS one*, *15*(1), e0227821.

Pierri, F., Piccardi, C., & Ceri, S. (2020). *A multi-layer approach to disinformation detection on twitter.*

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. Retrieved from `https://science.sciencemag.org/content/359/6380/1146` doi: 10.1126/science.aap9559