studying misinformation on Twitter with data viz + natural language processing

or, how I built this dashboard without writing a single line of JavaScript

20191022

@mathemakitten

hi hello i’m helen

- mostly machine learning
- interests in language models and misinformation
- Python 

whalefakes
@whalefakes

yeah do you realize that at least one whale is on drugs

10:21 PM · Jul 9, 2019 · Twitter Web Client

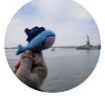
4 Retweets **28** Likes

whalefakes
@whalefakes

there are no laws that say a humpback whale cannot drive

7:18 PM · Aug 20, 2019 · TweetDeck

2 Retweets **6** Likes



whalefakes
@whalefakes

from a scientific standpoint, there is no such thing as the fuckin prince of whales

9:15 PM · Jul 14, 2019 · TweetDeck

2 Retweets **14** Likes

other things i am

- terrible at JavaScript
- honestly i’m amazed that the internet even works

what I wanted to do at RC

- study misinformation on Twitter
- with data from the Twitter Elections Integrity Hub
- study information operations directed at Hong Kong

a problem

- can we predict when an account transforms into a propaganda-related account?
  - **spoiler**: it’s really hard

the goals

- make this data accessible for everyone
- raise awareness about Twitter misinformation campaigns
- simplify my life while conducting exploratory data analysis for natural language tasks

Classic Dashboard

Creatively crafted Dashboard for your needs

$ 72548	72548	2548	1548
TOTAL PROFIT	ACTIVE USERS	NEW ORDERS	NEW COMMENTS
Monthly progress	Monthly progress	Monthly progress	Monthly progress

$ 72548	72548	2548	1548
TOTAL PROFIT	ACTIVE USERS	NEW ORDERS	NEW COMMENTS
Monthly progress	Monthly progress	Monthly progress	Monthly progress

Production February 2017

Export Print

Production February 2017

Export Print

helen
@mathemakitten

hello, i am a machine learning dev who just spent four hours struggling with (what seems to be??) the deep end of data viz/d3/JavaScript and i have a WHOLE NEW APPRECIATION for front-end complexity

i will now proceed to navigate the web with a newfound sense of awe✨

7:36 PM · Sep 28, 2019 · Twitter Web App

View Tweet activity

19 Likes

demo

some tools used

- wrapper on JavaScript + D3.js

- wrapper on React

- natural language processing library for entity detection and lots of other fun stuff

- machine learning library
  - including prebuilt language models

things i learned a non-comprehensive account

- it’s expensive to deploy a webapp which requires 7GB of startup memory with Heroku or Google App Engine
- **what I did instead**: host the dashboard myself on Google Cloud Platform with Flask (a web framework), gunicorn (a web server) & nginx (a reverse proxy)

more things i learned

- lots of NLP has been built on top of English corpuses
  - which is fine, until you need to do things in not-English
- NLP on spam is hard

stuff which didn't work

- encoding tweets from lots of languages with the Tensorflow Multilingual Universal Sentence Encoder and then clustering them for topic modeling
  - **spoiler**: it just learns to separate tweets by their original language

more stuff which didn't work

- entity detection on non-English tweets
- trying to translate 10M tweets from various languages with the Google Translate API

(languages are hard)

things I still don’t know

- how to write any JavaScript
- how the internet works

stuff still in progress

- predicting the timestep at which an account starts being propaganda-related
  - sub-problem: figuring out what to do with the large amount of non-English tweets in an account history

thank you

@mathemakitten

@whalefakes