

11-830 Computational Ethics for NLP

Lecture 13: Fake News and Influencing Elections



Carnegie Mellon University

Language Technologies Institute

Fake News and Elections

- Ads, recommendations
- Fake news
- Election influence



Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam.

Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam
- Buy Me! – sent to only those who might buy me
 - Hard to target that population (and you want more people to buy)



Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam
- Buy Me! – sent to only those who might buy me
 - Hard to target that population (and you want more people to buy)
- Buy Me! – I'll help you with your latest endeavor
 - Try to target the interest of new people to buy me

Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam
- Buy Me! – sent to only those who might buy me
 - Hard to target that population (and you want more people to buy)
- Buy Me! – I'll help you with your latest endeavor
 - Try to target the interest of new people to buy me
- Buy Me! – I'll help you with <your latest endeavor>
 - Actually personalize the message to include personalized phrases

Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam
- Buy Me! – sent to only those who might buy me
 - Hard to target that population (and you want more people to buy)
- Buy Me! – I'll help you with your latest endeavor
 - Try to target the interest of new people to buy me
- Buy Me! – I'll help you with <your latest endeavor>
 - Actually personalize the message to include personalized phrases
- Buy Me! – I'll help you with <your latest endeavor>
 - “It helped my granddaughter with her latest endeavor” – John from Pittsburgh

Let's Advertise ...

- Buy Me!
 - People don't always respond to general spam
- Buy Me! – sent to only those who might buy me
 - Hard to target that population (and you want more people to buy)
- Buy Me! – I'll help you with your latest endeavor
 - Try to target the interest of new people to buy me
- Buy Me! – I'll help you with <your latest endeavor>
 - Actually personalize the message to include personalized phrases
- Buy Me! – I'll help you with <your latest endeavor>
 - “It helped my granddaughter with her latest endeavor” – John from Pittsburgh
- “Everybody bought me and you wont believe what happened next ...”
 - Your whole sphere seems to have bought me.

Fake Reviews

- Try to be a verified purchaser
- Be specific about the project
 - Not just ... “Great product, arrived on time”
- Add some self disclosure for realism
 - “My 6 year old granddaughter loves it, “Granny, I love my Tesla K80 24GB GPU” she says.
- Generate multiple different reviews
 - Different classes of user
 - “Works great on Linux”
 - “Works on my Mac”
 - “Once Update has finished running, I know it’ll work great”
- But reviews are still best written by humans
 - They can be adapted automatically, and posted automatically
- Automatically posted when some one mentions the product

Review vs News

- “News” is perceived to be more authoritative
 - But user-written “reviews” are more genuine
- Many “news” articles also advertise the product
- Many ads are press releases designed to be quoted as news
- You can make your reviews be like news.
- You have to release them via a recognized News site
 - ... or not
- Different headlines but same story
 - Looks like there is more news about X
- Generate references to the articles
 - Pay for links
 - Tweet/retweet about them

News Flash

- Panel: Neural Networks and Deep AI
- Panelists: Geoff Hinton, Yoshua Bengio, Elon Musk and Emma Watson
- Thursday 21st March 10:30-noon,
- Rashid Auditorium

- More details: <https://seminars.scs.cmu.edu/>

News Flash

- Panel: Neural Networks and Deep AI
- Panelists: Geoff Hinton, Yoshua Bengio, Elon Musk and Emma Watson
- Thursday 21st March 10:30-noon,
- Rashid Auditorium

- More details: <https://seminars.scs.cmu.edu/>

- You wont believe what happened next ...

Clickbait

- Making people click on links
- Things they like
 - Kim Kardashian something something
- Things they want to know
 - Next Avengers movie will be released ...
- Things left unsaid
 - Something, something, you wont believe what happened next
- All using reinforcement learning to find the best headline
 - Kardashian Avengers bitcoin deep learning, you wont believe what happened next

So what happened to Truth?

- It maybe never was there ...
 - News reports about things I know about are always wrong in the details, I'm just pleased that all the other news is correct
- We could fact check everything
 - “water runs downhill” 17.5K documents
 - “water runs uphill” 116K documents
 - “flat earth” 11m vs “spherical earth” 300K
- Identify “good” sources of facts
 - But we actually want opinion too
 - Who decides truth?

Trustworthiness

- Jeff Pasternack and Dan Roth at UIUC/UPenn
- Identify sources for fact checking
- Present multiple views when searching
 - “Is milk good for you?”
 - Gave side-by-side search results for and against
 - This was preferred by most subjects (sometimes)
- But probably wont work when people are already charged in one direction

Confirmation Bias

- Humans see things to confirm their biases
 - “Well that’s probably only one example” vs
 - “I bet there are many more examples like this”
- Arguments are rarely actually rational debates
 - Besides you’re just clearly wrong anyway ...



Exploiting Human Behavior for Gain

- You probably can't change peoples views
- But you can amplify them

- I'm a democrat but my vote doesn't really count
 - Healthcare will still be too expensive under either party
 - News: "Democrats will cut healthcare costs"
 - Okay maybe I will vote

Getting People to Vote

- Rayid Ghani, Chief Scientist of Obama campaign 2012
 - Masters from MLD, now at U of Chicago leading “Data Science for Social Good”
- Amplifying Activism
 - Find marginal constituencies
 - Find registered democrats in the area
 - Identify their key interests (education, healthcare etc)
 - Send them messages about their key interests
 - Ask for donations
 - Measure success in sending messages
 - Do it again

Getting People to Vote

- Attenuating Apathy
 - Find marginal constituencies
 - Find registered democrats in the area
 - Identify their key interests (education, healthcare etc)
 - Send them messages about their key interests
 - Get them worked up about the election
 - Get them to vote

- It doesn't take much to change an election result

Getting People to Not to Vote



Getting People to Not to Vote

**Save Time
Avoid The Line**

Vote from home.

**Text "Hillary" to 59925
and we'll make history together**

This November 8th.

MUST BE 18 OR OLDER TO VOTE. ONE VOTE PER PERSON. MUST BE A LEGAL CITIZEN OF THE UNITED STATES.
VOTE BY TEXT NOT AVAILABLE IN GUAM, PUERTO RICO, ALASKA OR HAWAII. PAID FOR BY HILLARY FOR PRESIDENT 2016

Getting People to Not to Vote

- Deflect voters
 - Its not worth voting
 - Poll estimates show X is overwhelmingly winning
- Mislead voters
 - Vote by text to
 - Vote early on March 9th (but its actually March 6th)
 - You need government ID to vote

Misleading Voters Through News

- Show relevant News stories
 - Stories of interest to the particular voter
 - No longer a general editor/newspaper
 - Only see things in your news feed
 - Overwhelmed with obviously fake stories so ignore everything
 - Add fake facts to real stories
 - Question objectivity itself
 - Call “Fake News” for anything you don’t like

Targeting Influence

- Companies already do this
- Cambridge Analytica (from Wikipedia)
 - Part of SCL Group: a global election management company
 - Financially backed by Robert Mercer (early pioneer of Statistical MT)
 - Das Magazin: CA's methods based on Kosinski 2008 using profiling based on facebook "likes" and smartphone data.
 - Behavioral microtargeting

Can this be stopped

- Companies and Countries already do that
 - “Russia did it all”, “It was North Korea’s fault”
 - Could be a excuse, true, or just misinformation
- Where to draw the line
 - What is the difference between Riyad Ghani and CA?
- Can you ever define legality
 - You must allow people to campaign
 - You have to avoid creating unfair laws about campaigning
 - You want to stop unfair vote manipulation
- Does it actually work
 - Depends who you ask (the answer is itself biased)

Science of Manipulation

- Marketing and Advertising
 - We want to influence people
- Public Service Announcements
 - Influencing the populace to do “good” things
- Psychology
 - Studying human behavior
- Psychohistory (Asimov’s fictional “Foundation”)
 - Modeling group behavior
- Manipulation for good/bad
 - Make better decisions
 - Evolve better political systems

Unseen Consequences

- Its not just about deliberate/opportunistic manipulation
- Access to diverse information flow
 - Allows personalization of choice of interests
 - Moves your information flow to areas of interest
- But with personalization comes limitations
 - You only see the areas you want to see
 - Your own information bubble
 - But everyone I talk to online likes My Little Pony
 - You never see people liking other things so your “normal” changes

Rise of the Independent Star

- No longer manufactured from central organization
- Justin Bieber and Logan Paul
- Youtube allows for self-created stars
 - Those who manage themselves well succeed
 - May not be the most intellectual content, but its popular
- Unconventional organizations end up being in control
 - Google/Facebook/Amazon become unexpected gateways

Cambridge Analytica and Microtargeting

- Please read The Guardian 17th March 2018:
<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>
- Watch (if you can) “Brexit: The Uncivil War” Channel 4 Movie (2019)