On Fake News Detection: Machine Learning Approaches

Santhosh Kumar G, PhD
Cochin University of Science and Technology
sancochin@acm.org

National Conference on Indian Language Computing - NCILC 2019
March 2019
Kochi, Kerala, India
Outline

• Fake News in Primitive and Modern Times
• Fake News and Cognition
• ML approaches to Fake News
  • Supervised and Unsupervised Approaches
• Datasets
• Way Forward
post-truth

Fake News

n. (ferk 'nju:z)
False, often sensational, information disseminated under the guise of news.

# collinsWOTY
Fake News from 1522 CE

Piero Aretino (1492 – 1556)
Some Similarities from the 1522 Story

1. Fake News to Influence Elections
2. Fake News Posted on the “wall”
On 6th Nov 2016
Sample Fake News from the Present Era

The Rs 2000 currency is designed keeping in mind to eradicate the black money issues using state of the art indigenous nano technology, every Rs. 2000 currency note is embedded with a NGC (Nano GPS Chip).

The unique feature of the NGC is it doesn't need any power source. It only acts as a signal reflector. When a Satellite sends a signal requesting location the NGC reflects back the signal from the location, giving precise location coordinates, and the serial number of the currency back to the satellite, this way every NGC embedded currency can be easily tracked & located even if it is kept 120 meters below ground level. The NGC cant be tampered with or removed without damaging the currency note.
Sample Fake News from the Present Era

[Collected via e-mail, November 2011]

This is a story that happened to the relative of a friend of mine. If you love your family and your friends, just tell them about this story.

There was a woman who suddenly died with blood coming from her eyes, ear, mouth, nose, skin and tongue. After the authorities investigated, they claimed she died of food poisoning.

Before she died, she ate a lot of shrimp, but so did the rest of her family and they didn’t die. It turns out that the problem was that she drank a liquid vitamin C supplement immediately after she ate the shrimp.

Basically, Shrimp contains an Arsenic Pentoxide (As2O5) and when she ate shrimp, she drank Vitamin C at the same time. This caused a chemical reaction in her stomach. The Arsenic
Impact of Fake News

Friends I am getting a lot of suggestions to take lakshmi tharu and mulatha.. This was popularised through social media...I took the risk of trying it on myself and many other popular alternate medicines suggested by friends and family.. It couldn't control my Tumor and rather took me to a very dangerous situation.. I will never suggest it as an alternative to the already proved medication.. Maybe after a formal medication all these can be used so that it doesn't return back.. I wish and pray there is further study and research on all these to create a proper medication for cancer..Please don't advice this to anybody as an alternative to chemotherapy or any formal medication and mislead people.. It is very dangerous... And never believe forwarded messages in social media blindly..

Jishnu Raghavan (1979-2016)
Definitions of Fake News

• Intentionally false and shared via the Internet [Klein et al, 2017]

• “Everything from malicious stories to political propaganda” [Cohen, 2017]

• Fake News is a continuum; some news have some bias, yet others misinterpret, and some others step into misrepresentation territory.

Confirmation Bias

What you Hear or See

What you believe

What you value

Can be manipulated easily in the days of social media

Experiment to Study the Persistence of Fake News

Story of Nathalie, the nurse

Rate Nathalie on Trustworthiness

The story about Nathalie stealing stuff was fake news

Rate Nathalie on Trustworthiness again

The first negative rating was seen to influence the second rating quite a lot.

Keersmacker and Roets, 2016
Election-time Fake News

October 2016

Pope Francis Shocks World, Endorses Donald Trump for President, Releases Statement

October 2016: shown to be false

November 2016: US Elections

One week before Karnataka Elections 2018

Janta Ki Baat survey predicts huge BJP gain in Karnataka assembly elections.

- BJP - 135 Seats
- JDS - 45 Seats.
- Congress - 35 Seats
- Others - 19 Seats

The latest poll survey, conducted by Janta Ki Baat with the sample size of a 10.20 lakh respondents, says BJP is crossing 135 seats in Karnataka, a jump of 95 seats compared to the previous assembly elections. This means, the BJP would end up as the single largest party with a clear Majority over the Congress. The survey says that the BJP could get anything in between 125-135 seats. PM Modi, Amit Shah, Yogi Adityanath and former Karnataka CM Yeddyurappa are campaigning relentlessly for BJP in the state.
Story so far…

• Fake News has been around
• More so in present times
• Fake News Spreads Fast
• Fake News Persists in the mind long after it has been shown to be fake
Data Types of relevance for Fake News

Source of (Mis)information

Content of (Mis)information
Could contain hashtag, URL, image

Context of (Mis)information
Date, Time, Location etc.

Propagation of (Mis)information
Retweet, Reply, Like etc.

Ref: ICDM 2017 Tutorial (Wu et al.)
Two Families of Approaches

• ML Approach
  • Underlying Premise: underlying patterns can help telling apart fake news from others

• DB Approach
  • Underlying Premise: If the news story is not well supported by a knowledge base of facts, it may be fake

Yields better to identifying emerging fake news topics

Needs a continuously updated KB to identify unseen kinds of fake news

Ref: AAAI 2018 Tutorial (Hassan and Lee)
Families of ML Approaches

• Supervised
  • Uses labelled data in order to build fake news identification approaches

• Unsupervised
  • Treats deviations from a common character as being anomalous and thus likely to be fake
Supervised Fake News Identification

Methods differ in:
What kind of features are used to learn the model?
What kind of models are learnt?
Features from the Source

Originality = original tweets vs. retweeted

Role = # followers vs. # following

Influence

Credibility

Ref: AAAI 2018 Tutorial (Hassan and Lee)
Features from the Content

• Lexical: What words are used?
• Syntactic: Past tense, Present Tense etc.
• Emotion: Emojis, emotional words
• Hashtags and URLs
• Content of the Photos
• Content of the Videos
Features from Propagation

Ref: AAAI 2018 Tutorial (Hassan and Lee)
Classification Methodologies

• Support Vector Machines
• Bayes Methods [EMNLP 2011]
• Conditional Random Fields [SocInfo 2017]
• Deep Neural Networks
Deep learning approaches to NLP

CNN for NLP
Recurrent Neural Networks (RNN)
Attention Mechanism
Recursive Neural Network
Reinforcement Learning
Deep Generative Models
Memory Augmented Network
CNN for NLP task

Unsupervised Fake News Identification

The large majority of works treat Fake News Identification as Anomaly Detection

1. Learn the character of the nature of documents as a statistical model.
2. Identify those that deviate from the general character, and label those as fake news.
Drawback of the Anomaly Approach

• Pre-supposes/Assumes that most news are not fake
  • This may be violated in many scenarios!
Unsupervised Fake News Identification

• Very fertile area of research, since it is very hard to get labelled data for supervised approaches to work

• Meta-Approach:
  • Understand the scenario, and identify on what ‘dimensions’ fake news varies significantly from legit news
  • Identify news pieces from either extremes on those dimensions, and label them as ‘fake’ and ‘legitimate’
  • Propagate the labels to the unlabelled news pieces based on how similar they are to those labelled in the previous step
Dimensions of Variation (1/4): Propagation Peaks

- Fake News tends to have multiple propagation peaks (e.g., reshares, re-tweets), whereas legit news tends to have just one peak

Friggeri et al., ICWSM 2014
Dimensions of Variation (2/4): Diffusion Network Pattern

• Kwon et al., ICDM 2013: Information on a topic diffuses from A to B if they are connected and B re-tweets A’s post on the topic; the diffusion network patterns are likely to be different for fake and legit news

Fake News Diffusion Pattern:
Large number of singleton nodes who got the information from outside the n/w

Legit News Diffusion Pattern:
Much more connected
Dimensions of Variation (3/4): Affective Character of Content

- Fake News is likely to be highly emotional;

(a) Fake News

(b) Real News

Figure 4: Emotional expressions for anger in fake news and real news. Compared to real news, fake news use more fierce and extreme words to express anger.

Dimensions of Variation (3/4): Affective Character of Content

• Fake News is likely to be highly emotional; so, if we can amplify the emotional content, fake news and legit may become better separable
Dimensions of Variation (4/4): Enquiry Responses

• When there is a fake news, there is a high likelihood that people would respond with a lot of enquiries

<table>
<thead>
<tr>
<th>Inquiry</th>
<th>Tweet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oh my god is this real?</td>
<td>RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured</td>
</tr>
<tr>
<td>Is this true? Or hacked account?</td>
<td>RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured</td>
</tr>
<tr>
<td>Is this real or hacked?</td>
<td>RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured</td>
</tr>
<tr>
<td>How does this happen?</td>
<td>#hackers RT @user: RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured</td>
</tr>
<tr>
<td>Is this legit?</td>
<td>RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured</td>
</tr>
</tbody>
</table>

Zhao et al., WWW 2015
What other dimensions make Fake News different from Legit ones?

- Topical Coherence?
- Left or Right Political Orientation?
- Referencing of contemporary as against historical events?
- Highly confident or skeptical narrative?
- Presence of Numbers or Statistics?
- Presence of Quotations?
Public Datasets and Challenges

Kaggle Challenge: https://www.kaggle.com/c/fake-news
SemEval 2019: https://competitions.codalab.org/competitions/20022
Fake News Challenge: http://www.fakenewschallenge.org/
Fact Checking Websites

Snopes.com : https://www.snopes.com/
Alt News : https://www.altnews.in/
FactCheck.org : https://www.factcheck.org/
Politifact.com : https://www.politifact.com/
theOnion.com : https://www.theonion.com/
Datasets

• LIAR
  • Textual dataset comprising true and false news collected from politifact.com

• Kaggle Fake News
  • News articles tagged as fake using the Bullshit Detector Chrome API

• BuzzFeedNews
  • US Politics Articles from various news agencies labelled as true or fake, from the context of the Presidential Election

• Wild Web Tampered Image Dataset
  • Realistic tampered images disseminated through social media

• …
Datasets

PHEME : social media rumor dataset
So, Can ML combat Fake News?

• Combating would include
  • Detection (including Early Detection)
  • Throttling (preventing their spread through the network)
  • Repair (undo the damage that they have already done)

Long way to go!

We are starting to see what techniques for effective Fake news detection would look like.
Thanks!
acknowledgement

sancochin@acm.org
Acknowledgement

Dr. Deepak Padmanabhan,
Professor, Queen's University Belfast, UK for sharing the slides used by him for ICDSE 2018

https://pure.qub.ac.uk/portal/en/persons/deepak-padmanabhan(d64bccad-bac3-4336-9336-d27af8995e94)/activities.html