Team 39 - Fake News Detection

Group Members:
Rabiya Noori
Jerri Zhang
Iris Yoon
Renee G. Reynolds
Hannah Mei

Problem Statement

Disinformation and manipulation in the media have always existed, but with the advent of social media, there has been an increase in the amount of misinformation spread, the ability of anyone to contribute false information, and the speed at which this information can be spread to larger audiences. Fake news has even drawn more attention as it is widely used to spread conspiracies theories and misconceptions about politics. For this project, we are interested in minimizing the spread of misinformation. We will explore machine learning methods of natural language processing (NLP) and recurrent neural networks (RNN) as means to identify and address the issue of fake news in different categories.

Specifically we will focus on answering the question:

For a certain piece of news (article, tweet etc.) is it true or false?

Data

Our main data source consists of 17,903 fake news articles and 20,826 real news articles, which are part of a publicly available Kaggle dataset. Each news item has been classified as ‘True’ or ‘False’ by the non-profit fact-checking website PolitiFact. The fake articles mainly come from satirical news websites such as addictinginfo.org and the majority of the real news articles are taken from Reuters.

Here is a summary of our data sources:

1. https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset
    - Main datasource
    - 17,903 fake news; 20,826 real news
    - Each news item has been classified as ‘True’ or ‘False’ by the non-profit fact-checking website PolitiFact.
    - The fake articles mainly come from satirical news websites such as addictinginfo.org
    - Majority of the real news articles are taken from Reuters.

2. Fake news from 4 years ago (around time of 2016 election)
    - https://www.kaggle.com/mrisdal/fake-news
    - Could be useful for addressing the question “Can we track the evolution of fake news over time”

Methods

Exploratory Data Analysis + Data Cleaning

- preprocessing techniques
    - tokenization, stemming, generalization or weighting words
    - to convert tokenized texts into features:
        - TF-IDF (term frequency-inverse document frequency)
        - LIWC (linguistic inquiry and word count)
    - pre-learned word embedding vectors for word sequences:
        - word2vec
        - GloVe
    - when using entire articles, an additional preprocessing step is to identify the central claims from raw texts
        - rank the sentences using TF-IDF and DrQA system
            - closely related to word embeddings, named entity recognition, disambiguation or coreference resolution

Baseline Model

RNN: Recurrent neural networks recognize the data’s sequential characteristics and use patterns to address classification and regression tasks. Each recurrent unit in the network contains value composed by input at time t and memory from previous input at time t-1. For each input sentence or article of length T, the network unfolds for a sequence of recurrent state of length T and updates its memory with each word. During training, the network applies backpropagation through the recurrent state and improves the network by optimizing the loss function.

Word Embedding: Word embedding uses efficient, dense representations of words, which transform similar words into similar encodings. Specifically, an embedding is a dense vector of floating point values learned by the model during training.

Transfer Learning

BERT: Oftentimes with complex NLP problems, using a pre-trained model is common to avoid starting from scratch. For our work we will try to find pre-trained models (e.g. BERT) relevant to our problem and then train the final layers to fit to our data.

Analyzing the results

LIME: Oftentimes when we work with NLP models, we want to see what words or phrases the model is looking at when making classification. LIME provides the importance of individual words and phrases in the corpus and helps explain the model's output.

Some further questions we might explore:

- What are the topics/areas that are most susceptible to fake news?
    - What is the relationship between location and fake news i.e. are there certain areas which have higher incidences of a certain type of fake news?
- Can we identify conspiracy theories or themes of fake news relevant to COVID-19 and the upcoming elections?

Visualizations

- word cloud
- frequency n-gram

Final Product

We would like to package our analysis in the form of a web tool or chat bot. In this tool, the user will have the ability to input a news article and the tool will flag the article as fake or real. If the article is flagged as fake, the fake text will be highlighted as explanation for the result.

An additional action item is determining common themes of fake news across various of the States, so that campaigns can be run to dispel myths or conspiracy theories; this is relevant for fake news surrounding elections.

Timeline and Milestones

Week 1	Sep 25	Fri		Complete Project Outline; Confirm Data Source
	Sep 26	Sat		
	Sep 27	Sun	Project Outline Due	
Week 2	Sep 28	Mon		EDA; Reading on BERT/Transformer
	Sep 29	Tues		

	Sep 30	Wed	1st Mentor Meeting	
	Oct 1	Thur		Model building, Transfer learning; writing up report draft
	Oct 2	Fri		
	Oct 3	Sat		
	Oct 4	Sun	Report Draft Due	
	Oct 5	Mon	2nd Mentor Meeting	
	Oct 6	Tues		Improve model performance; LIME; Final report; Packaging/API; Poster
	Oct 7	Wed		
	Oct 8	Thur		
	Oct 9	Fri		
	Oct 10	Sat		
Week 3	Oct 11	Sun	Final Report Due	
	Oct 12	Mon	3rd Mentor Meeting	
	Oct 13	Tues		
	Oct 14	Wed	Final Presentations due	
	Oct 15	Thur		
Week 4	Oct 16	Fri	Final Presentation day	

Concerns

The primary concerns with our projects are:

- Acquiring latest dataset from categories other than political news
- Time constraints
- Definition of fake news can be ambiguous. The easy kind of fake news are those that contain blatantly incorrect facts. However, there are numerous types of misinformation whose truth value depends on the person, such as the following:
    - Selective editing -- based on factual statements that imply false information.
    - News with no objective truth. These can be particularly vulnerable to the reader's political views e.g. sexual assault cases
    - This means that the model will reflect the bias of whoever labeled the data. To fully understand the limits of a fake news detector, it's important to understand the criteria through which news have been labeled as fake.