Towards automating the detection of misinformation

Data Incubator Capstone Project Proposal

Haritha Gangavarapu
MS Data Analytics
George Mason University

Motivation

- Misinformation is everywhere:
  - In 2009 Belkin used Amazon's MTurk to get fake reviews for its products by paying incentives to people.
  - In India, there were mob killings because of the rumours spread on whatsapp about child abductions in many states.
  - A gunman barged in to a restaurant having been exposed to fake news that the owners were involved in child abuse.
  - Rumors about transfer of club soccer players shoot up their prices to millions above the set valuation.
- Reputation management firms manage the image of a business.

Current Systems

- Human annotated fact checking sites
  - Politifact
  - Gossipcop

- Rule based web browser extensions
  - BS Detector: List of web pages that are blacklisted
  - Media Bias Fact Check

Open Source Datasets

- Two major Datasets
  - LIAR
    - 12,836 human-labeled short statements from Politifact
    - Metadata
      - Speaker name, party affiliation etc.
  - Fake News Net
    - 854 News articles from Politifact and Gossipcop
    - Temporal and Social Media based metadata
- Others
  - Buzzfeed
  - Fake News Challenge for Stance Detection

Performance of current Machine Learning based systems

- LIAR (Wang, 2017)
  - Combined Text and metadata analysis
    - Highest test accuracy of 27.4% with Convolution Neural Network

- Fake News Net (Shu & Mahudeshwaran, 2018)
  - Textual analysis
    - Highest accuracy of 62.9% with Convolution Neural Network
  - Social context Analysis

GOAL

To develop a holistic and intelligent system that can identify and effectively classify misinformation.

The system must consider all the components of information media:

- Text to analyse linguistic Structure
- Images
- URLs
- Extent of dissemination on social media

Proposed system - Work done

- Data Exploration
- Experimentation with word vector representation
- Replication of current state of the system
- Baseline models and accuracy scores
- LIAR
    - Regularized Logistic Regression: 26.7%
    - SVM: 25.3%
    - Naive Bayes: 26.6%
- FakeNewsNet
    - Regularized Logistic Regression: 83.0%
    - SVM: 59.1%
    - Naive Bayes: 63.1%

Proposed system - In progress

- Deep Learning based methods for Text classification
  - Convolutional Neural network
  - Recurrent Neural network
    - Traditional RNN, LSTM and GRU
    - Bidirectional Recurrent Neural Network
- Neural Language Models for learning representations

Till now:

- Cleaning the data and tested various tokenizers, stemming techniques
- Handled sparseness in data due to word frequency representation
- Language models are computationally expensive to run

Proposed system - Next Steps

- Textual Entailment using updated information from Google search API
  - Attention based methods
  - Needs Human in the loop to verify authenticity of the source
- Apply image analysis methods to verify the authenticity
  - Needs to be a cold start(cannot use available Imagenet)
- Methods inspired from NER and Semantic Role Labeling tasks
  - syntactic/semantic parsing is an expensive process
  - Measured performance on large texts is not known yet

Thank you

References

- Liar dataset https://arxiv.org/pdf/1705.00648.pdf
- FakeNewsNet https://arxiv.org/pdf/1809.01286.pdf
- Politifact https://www.politifact.com/
- Gossipcop https://www.gossipcop.com/
- BS detector https://gitlab.com/bs-detector/bs-detector
- Whatsapp rumors - India https://en.wikipedia.org/wiki/Indian_WhatsApp_lynchings
- Detecting Fake reviews https://www.microsoft.com/en-us/research/video/detecting-fake-reviews/