CIS 520 Fall 2020, Project Proposal

Fake News Detection

Team Members:

- Haowei Liu; Email: hwliu@seas.upenn.edu
- Yuchen Zhang; Email: zycalice@seas.upenn.edu

---

1 Motivation

In the current era of information overflow, identifying information validity becomes a challenging task for news outlet, social media, and just the general public. This process becomes increasingly more relevant when platforms or individuals voice opinions incentivized by personal interests, and are not as much concerned about verifying these information. We see this phenomenon intensified in the midst of political turmoil. Therefore, we believe it's an important to build a scalable model to preemptively identify fake news from more legitimate ones, which would otherwise be too inefficient to go through human scrutiny. For social media and news collection agencies, this would help them to flag unreliable news. For finance/investment companies, this could help them to have a better sense of the sentiment and to predict the financial markets more accurately.

2 Dataset

With this goal in mind, we have located a dataset on Kaggle titled, **Fake and real news dataset**.[1] Formally, the dataset is called "ISOT Fake News Dataset".[2] It consists of a total of about 40,000 news articles, and manually annotated label of either real or fake news. This is a fairly balanced dataset, around 50% of the news article classified as fake news. This dataset was collected from real-word sources dated 2015 to 2017 - the real news had been obtained by crawling Reuters.com, whereas the fake news are collected from unreliable sources as flagged by third-party fact checking organizations[1].

News	Number of Articles
Real news	21,417
Fake news	23,481

Table 1: Data Summary

In addition to the labels, the provided features includes:

1. **title**: Title of the article; we will include this in the features.
2. **article**: The full article itself, including punctuation; we will include this in the features.
3. **subject**: For fake news, the subjects include "GovernmentNews", "Middle-east", "US News", "left-news", "politics", "News". For true news, the subjects include "World-News" and "PoliticsNews". Since the labels seems to be messy and unbalanced, we will not include these labels in training.

[1]Kaggle dataset source, https://www.kaggle.com/clmentbisaillon/fake-and-real-news-dataset.

[2]Kaggle dataset metadata - sources, https://www.uvic.ca/engineering/ece/isot/assets/docs/ISOT_Fake_News_Dataset_ReadMe.pdf

4. **date**: We could potentially use this feature, for example creating a feature indicating if the news was published on weekends or not. This will be explored at the end.

3 Related Work

One major challenge with our project is how to sensibly translate text information to numeric representation and use this information to predict the validity of a news article. This task falls under the widely discussed area of text classification. A variety of methodology and application have been proposed by previous research, and we found the following sources especially relevant to our project.

3.1 Blogs

In a medium blog post regarding text classificaiton, the author detailed a complete workflow of text classification task from preprocessing the raw text to running machine learning model.[2] The author implemented two context-free feature extraction method, one based on term frequency, TF-IDF and the other using deep learning, word2vec. With these extracted features, the author proceeded to perform logistic regression and naive bayes for classification. The models were evaluated using accuracy score, and TF-IDF combined with logistic regression produced the best performance. Our takeaway from this article is that we also intend to perform feature extraction using term-frequency and pre-trained deep learning models. Building upon the two classification models mentioned so far, we are also planning to experiment with other classification method, such as random forest and SVM.

To use deep learning techniques to create word embeddings, we will reply on huggingface documentations and blog posts like `http://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/`. This post provided a detailed tutorial and visual representation of the pre-trained DistilBERT model.

In sum, we intend to apply these aforementioned text embedding and deep learning methodologies to identify whether a news article is fake or real given its content.

3.2 Research and Academic Articles

1. **Automatic Detection of Fake News**[5]
   `https://web.eecs.umich.edu/~mihalcea/papers/perezrosas.coling18.pdf`
   In this paper, the authors used a linear SVM classifer, and used a 5-fold cross validation method. The paper used accuracy, precision, recall, and F-score as evaluation metrics. The paper grouped accuracy by feature types, finding that the best performing classifers are the ones "that rely on stylistic features", such as punctuation and readability. In addition, it is interesting that the paper tested domain-specific news with different types of features. For technology, features that represent readability are the most important predictors.

2. **The Language of Fake News: Opening the Black-Box of Deep Learning Based Detectors**[3]
   `https://cbmm.mit.edu/sites/default/files/publications/fake-news-paper-NIPS.pdf`
   This is an interesting paper using CNN to detect fake news. In addition, the model aimed to predict novel topics correctly. Although it is interesting, we will not do CNN or novel topics prediction centered training here due to resource and time constraints. We will use a fully-connected neural network model.

3. **Text-mining-based Fake News Detection Using Ensemble Methods**[4]
   `https://link.springer.com/article/10.1007/s11633-019-1216-5`
   Since this is not a free resource, we only examined the abstract. The paper found that using a combination of features and word-vector representations, an accuracy of 95.4% is achieved using **ensemble** method.

4 Problem Formulation and Methods

This is a supervised learning problem, with labels indicating if the article is a fake news or not. For this binary classification problem, the inputs are words in the news articles, titles and other potential features. Our project consists of two major steps.

1. **Preprocessing:** Because we have full articles and titles, we plane to use natural language processing techniques to preprocess the data and extract features. We will use two methods: 1) vectorize the words in news articles (using sklearn), and 2) context-based word embeddings (using pretrained bert models). Then we would attempt to use PCA to reduce the dimensions before feeding into models. For bert, we will consider including the punctuation. We will use distilBert.
2. **Models:** The second step is to use traditional classification models to identify fake news. We intend to use the following models:
   (a) Base models: Logistic Regression, SVM
   (b) Statistical Learning: Random Forest, Native Bayes
   (c) Deep Learning: Fully-connected Neural Network

5 Evaluation

5.1 Loss Function

Regarding loss function, we will be working with different models which inherently assumes different loss functions.

1. Logistic Regression: logistic loss
2. Support Vector Machine: Hinge Loss
3. Random Forest: Cross Entropy and Information Gain
4. Native Bayes: Negative Joint Log-likelihood
5. Deep Learning: Cross Entropy and Information Gain

5.2 Evaluation

We will use cross-validation to select parameters. We plan to use both accuracy and confusion matrix to evaluate our model.

1. Since we are dealing with a binary classification problem, and our data is fairly balanced between two target classes, thus we think the **accuracy** score is a fair evaluation metric.
2. Although, the evaluation criteria may change depending on the application of this model - whether we want to capture as many fake news as possible or that we want to be confident that the identified fake news are indeed incredible. Therefore, we are also planning to use F_1 score to evaluate our model performance.

6 Project Plan

Week 1: 11/4

- Explored datasets and agreed on one dataset

Week 2: 11/11

- Planned the project by reading blog posts and research on relevant work

Week 3: 11/18

- TO DO: Complete first round on data pre-processing

Week 4: 11/25

- TO DO: Start training and testing models

Week 5: 12/2

- To DO: Finalize the results, explore additional features/things to add on, and prepare deliveries

References

[1] Saad S Ahmed H, Traore I. Detecting opinion spams and fake news using text classification. *Journalof Security and Privacy*, 1(1), January/February 2018.

[2] Ishaan Arora. Document feature extraction and classification.

[3] N. OâBrien, Sophia Latessa, Georgios Evangelopoulos, and X. Boix. The language of fake news: Opening the black-box of deep learning based detectors. 2018.

[4] Raj N. Gala M. et al. Reddy, H. Text-mining-based fake news detection using ensemble methods. *Int. J. Autom. Comput.*, 2020.

[5] Alexandra Lefevre Rada Mihalcea Veronica Pérez-Rosas, Bennett Kleinberg. Automatic detection of fake news. *COLING*, 2018.