Fake News Detection with Deep Learning

Yuqi Liu / 301355758
Xinyue Ma / 301297142
Yufei Wang / 301338190

1 Motivation

1.1 NLP task

We will use deep learning to train a model that can estimate whether the stance of a body text from a news article relative to a headline. We will try to construct a model similar to XLnet, which is one of the most popular neural networks in the NLP field. And for dataset, we will use FNC-1 to train and test our model. And the labels for this task are unrelated, discuss, agree and disagree.

1. **Agrees**: The body text agrees with the headline.

2. **Disagrees**: The body text disagrees with the headline.

3. **Discusses**: The body text discuss the same topic as the headline, but does not take a position

4. **Unrelated**: The body text discusses a different topic than the headline

1.2 Aspect

So far, it is difficult for AI to detect fake news using any existing technologies since the dataset is not available and sometimes some fake news are even plausible from human perspectives. Due to the difficulties of directly detecting whether a story is real, we decide to work on stage I of the fake news challenge, which is to classify whether the content of the news is related to the headline.

1.3 Reasons

As the media becomes easier to spread the news around the world, only one piece of news can have a sensational response. However, there are many people making up stories with an intention to deceive for a secondary gain. And this is one of the most serious challenges facing the news industry today. Deep learning has the power to detect whether a story is real or not.

2 Approach

We were supposed to use XLnet, which is one of the most popular models for NLP, to train the model for the fake news challenge. However, XLnet is a large model, we are afraid that we can not have enough computing resources to use it. So, we plan to implement a baseline method for this task. We plan to implement the baseline model indicated in the http://www.fakenewschallenge.org/, which use hand-coded features and a GradientBoosting classifier. And after this, we will try to find other papers related to this challenge and try to improve the accuracy.

3 Data

We plan to use FNC-1 dataset for this project and the dataset is available on Github.
https://github.com/FakeNewsChallenge/fnc-1

4 Evaluation

We plan to use the weighted accuracy indicated in the paper(Bhatt et al., 2017).

$$Score_1 = Accuracy_{Related,Unrelated}$$

$$Score_2 = Accuracy_{Agree,Disagree,Discuss}$$

$$Score_{FNC} = 0.25 * Score_1 + 0.75 * Score_2$$

5 Timeline and work breakdown

By the milestone, we plan to finish the implementation of baseline method indicated above and be in the process of finding better solutions.
We plan to finish every part together and share our ideas immediately to get better results.

References

Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. 2017. On the benefit of combining neural, statistical and external features for fake news identification. *arXiv preprint arXiv:1712.03935*.