Detect Fake News and Bias jointly by Multi Task Learning

JSS Sravan Chandra,201501025
Lakshya Agrawal,201530102
Khan Umair,20172060
Venkata Narendra Reddy,2018900009

Abstract:

In this research summary, we explore the application of NLP techniques to the detection of fake news that is, misleading news stories that come from non-reputable sources. The aim of the present work is to develop a high-performance model and compare the performance of models using two distinct feature sets to understand what factors are most predictive of fake news.

Basically Fake news is a term that has been used to describe very different issues, from satirical articles to completely fabricated news and plain government propaganda in some outlets. It is a problem that is heavily affecting society and our perception of not only the media but also facts and opinions themselves.

This project mainly tries predicting whether a given news article is a Fake news or not and the kind of bias the news publisher has using Multi-task architecture.
Existing studies, mainly regard fake news detection and bias classification as separate tasks. Enlightened by the multi-task learning scheme, we implement a joint framework that unifies the two tasks, i.e., fake news detection and bias classification. The idea is to jointly learn shared embeddings and then use them for multi-task prediction.

Dataset:

- The Real and Fake news data is collected from FakeNewsNet.
- Using their source information we extracted the bias of each article using Media Bias/Fact Check.
- Some Sources aren't still reviewed,some articles didn't have any source information so they are labeled as 'No Source'.
- Our final bias labels are Satire, Questionable Source, Right Bias, Left Bias, Conspiracy-Pseudoscience, Least Bias, Left-Center Bias,No source.

Real or Fake Detection(Single Task):

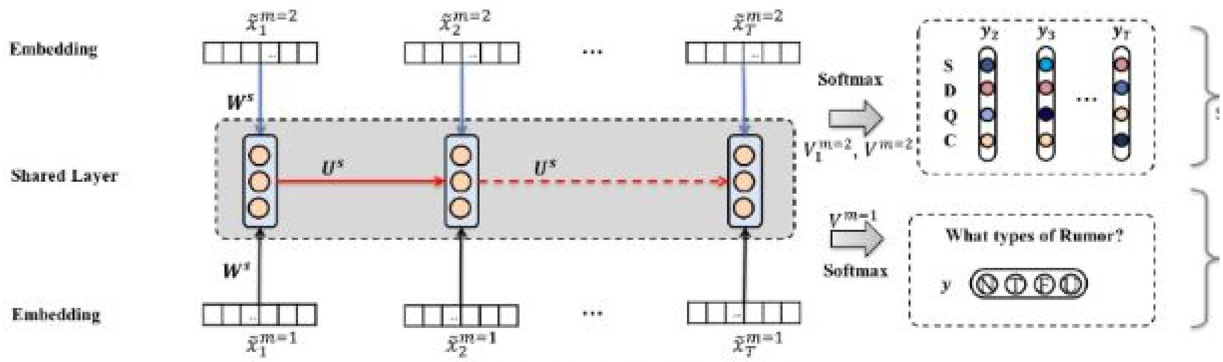
- Libraries used were Keras, Scikit and Numpy.
- Trained a LSTM model to classify a news article into Real and Fake News.
- Initially the given text was tokenized and mapped to numbers.
- The output was a vector and it was padded to maintain consistent size all over.
- This was passed to the embedded layer which creates vectors to our vocabulary.
- This was finally passed to LSTM which used loss function as “Binary Cross Entropy” and “Adam” optimizer.
- The data was initially divided into 80% training and cross validation and remaining 20% for testing.
- 5-fold cross validation was used to get appropriate parameters.
- Then it was finally tested on the test data and the accuracy reported was around 60% training accuracy and 57% testing accuracy.

Multitask Learning using shared layer:

We can motivate multi-task learning in different ways: Biologically, we can see multi-task learning as being inspired by human learning. For learning new tasks, we often apply the knowledge we have acquired by learning related tasks. For instance, a baby first learns to recognize faces and can then apply this knowledge to recognize other objects.

Finally, we can motivate multi-task learning from a machine learning point of view: We can view multi-task learning as a form of inductive transfer. Inductive transfer can help improve a model by introducing an inductive bias, which causes a model to prefer some hypotheses over others. For instance, a common form of inductive bias is ℓ_1 regularization, which leads to a preference for sparse solutions. In the case of MTL, the inductive bias is provided by the auxiliary tasks, which cause the model to prefer hypotheses that explain more than one task.

Uniform Shared Layer Architecture:



(a) Uniform Shared-Layer Architecture

- We train both tasks jointly using weight sharing to extract the common and task-invariant features while each task can still earn its task -specific features.
- This model contains only a single shared layer.
- 5-fold cross validation was used to get appropriate parameters.
- Then it was finally tested on the test data and the accuracy reported was around 60% with a training accuracy of 72% and 25% training accuracy and 33.3% testing accuracy on Bias prediction.

Links:

- https://github.com/Sra1chandra/IRE-FakeNewsBias
- **https://drive.google.com/drive/folders/14sbjaSxXpJ8Hwz0KrpPrRm5g05WKzP0M?usp=sharing**

References:

1) FakeNewsNet:https://github.com/KaiDMML/FakeNewsNet
2) Media Bias/Fact Check:https://mediabiasfactcheck.com/
3) Detect Rumor and Stance Jointly by Neural Multi-task Learning : https://dl.acm.org/citation.cfm?id=3188729
4) An Overview of Multi-Task Learning in Deep Neural Networks : https://arxiv.org/abs/1706.05098