SYNOPSIS

Automatic Stance Detection

Mentor:- Prof. Md. Zeeshan Ansari

By:Abhinav Kumar Jha(15BCS0002)

Shaz Akhtar(15BCS0039)

Automatic Stance Detection

The issue of “fake news” has arisen recently as a potential threat to high-quality journalism and well-informed public discourse.In this project we are performing “stance detection” -- i.e. identifying whether a particular news headline “agrees” with, “disagrees” with, “discusses,” or is unrelated to a particular news article -- in order to allow journalists and others to more easily find and investigate possible instances of “fake news”.

Problem Statement

The problem is about “stance detection,” which involves comparing a headline with a body of text from a news article to determine what relationship (if any) existsbetween the two. There are 4 possible classifications:

- The article text agrees with the headline.
- The article text disagrees with the headline.
- The article text is a discussion of the headline, without taking a position on it.
- The article text is unrelated to the headline (i.e. it doesn’t address the same topic).

DataSet Overview

The datasets for our task are provided by the Fake News Challenge organization. The complete training set consists of just under 50,000 “stance” tuples, with each tuple consisting of:

- A headline that is to be compared against an article to determine its stance toward the article. Word counts for the headlines range from 2 to approximately 40, with an average length of ~11.
- The (integer) ID of an article against which the headline is to be compared, which can be used to find the text of the article body in a separate file. Article lengths range from 2 to nearly 5000 words,with an average length of around 360 words.
- The true stance of the headline with respect to the article. (This is one of the four classes outlined earlier: agree, disagree, discuss, and unrelated.)

https://github.com/FakeNewsChallenge/fnc-1

Modelling Techniques

Our goal in approaching the stance detection problem was to experiment with a wide range of the deep learning and NLP techniques including:

- Dense vector embeddings of words, tokens, and sequences
- Multi-layer feed-forward networks
- Recurrent neural networks, including LSTM and GRU “cells”
- Attention mechanisms

Programming Tools

Language :Python
Associated Libraries:Pandas,SciPy,Numpy,TensorFlow,Keras.

References

- [1] Augenstein, Isabelle, et al. "Stance detection with bidirectional conditional encoding." arXiv preprint arXiv:1606.05464,2016.
- [2] Bahdanau, Dzmitry, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473, 2014.
- [3] Bowman, Samuel R., et al. "A large annotated corpus for learning natural language inference." arXivpreprint arXiv:1508.05326, 2015.
- [4] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555, 2014.
- [5] Ferreira, William, and Andreas Vlachos. "Emergent: a novel data-set for stance classification." Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational .
  Linguistics: Human Language Technologies. ACL, 2016.