

AIML - Sequential models in NLP - Project 2

The Real Problem

The goal of the Fake News Challenge is to explore how artificial intelligence technologies, particularly machine learning and natural language processing, might be leveraged to combat the fake news problem. We believe that these AI technologies hold promise for significantly automating parts of the procedure human fact-checkers use today to determine if a story is real or a hoax.

Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts. Fortunately, the process can be broken down into steps or stages. A helpful first step towards identifying fake news is to understand what other news organizations are saying about the topic. We believe automating this process, called Stance Detection, could serve as a useful building block in an AI-assisted fact-checking pipeline.

Stance Detection involves estimating the relative perspective (or stance) of two pieces of text relative to a topic, claim or issue. This extends the [work](#) of Ferreira & Vlachos. The task is to estimate the stance of a body text from a news article relative to a headline. Specifically, the body text may agree, disagree, discuss or be unrelated to the headline.

Project Description

Given the huge impact of social networks, online content plays an important role in forming or changing the opinions of people. Unlike traditional journalism where only certain news organizations can publish content, online journalism has given chance even for individuals to publish. This has its own advantages like individual empowerment but has given a chance to a lot of malicious entities to spread misinformation for their own benefit. As reported by many organizations in recent history, this even has an influence on major events like the outcome of elections. Therefore, it is of great importance now, to have some sort of automated classification of news stories.

In this work, we will implement different deep neural architectures using word embeddings for identifying the stance detection of different news stories.

Dataset

The dataset is taken from Fake News Challenge, where the goal of the challenge is not to directly identify whether a headline or article is “fake” or not, which is arguably a highly subjective question, and one that even skilled humans may have difficulty answering.

The challenge is organized around the more well-defined problem of “stance detection,” which involves comparing a headline with a body of text from a news article to determine what relationship (if any) exists between the two.

There are 4 possible classifications:

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

Presumably, a classifier that can solve the stance detection problem with high accuracy might effectively be used either as a tool for humans working to identify fake news (e.g., retrieving articles that agree, disagree and discuss a headline), or as a building block for a more elaborate future AI system that would try to identify the actual truthfulness of news stories (e.g., using credible sources to classify).

Overview

Step 1: Understanding how to convert words into vectors and make them ready for the model. This is an important step and will be the same for almost any problem in NLP when dealing with words.

Step 2: Understanding the working of Sequential Models like RNN, LSTM.. and practically building them in Keras and build the model ready.

Step 3: Train the model and report the accuracy score.

Step 4: Add checkpoints as callbacks to the model and save the model for each epoch. This callback saves the model after each epoch, which can be handy for when you are running long-term training. By loading the saved weight files we can start training the model which starts the learning from there.

Step 5 (Optional): Add attention to the above model and report the change in accuracy. And explain how and why attention has improved (or disturbed) the model.

The breakup for the scores is provided in the notebook.

Reference

Acknowledgements:

<http://www.aclweb.org/anthology/N16-1138>

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

Project submissions and Evaluation Criteria

While we encourage peer collaboration and contribution, plagiarism, copying the code from other sources or peers will defeat the purpose of coming to this program. We expect the highest order of ethical behavior.

Submit the code on Olympus.

Submit the project milestone in a Jupiter notebook and submit it on Olympus for evaluation.

Project Support

You can clarify your queries by dropping a mail to Olympus

Happy Learning!