Stance Detection for the Fake News Challenge Dataset using Deep Learning

Anak Agung Ngurah Bagus Trihatmaja and Shishir Kurhade

Abstract

The problem of fake news has arisen recently as a threat to high-quality journalism and well-informed public disclosure. The goal of 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 [1]. The goal of our project is to develop machine learning models to predict a stance label (Agrees, Disagrees, Related, Unrelated) with respect to the title for a respective news article. For this purpose, we will use the gated recurrent unit (GRU) to predict the labels. As a baseline to measure performance we will also solve the problem using logistic regression method.

1 Introduction

In this project, we try to combat a serious problem in our media using machine learning techniques. In a poll conducted by Pew Research Center, 64% of US adults said that fake news has caused a great deal of confusion about the basic facts of current issues and events [2]. This problem views the task of fake-news detection as a stance detection problem which is a labeling task. We want to automatically classify a news into four labels, which are unrelated, agrees, disagrees, and discusses.

A reasoning for these labels is as follows:

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

The classifier that we build could later be used as a base of a fake news detection tool that can automatically categorize the news into the stances given.

2 Proposed Project

2.1 Dataset Overview

The data provided by the Fake News Challenge consists of headline, body, and stace. For training, there are two csv files:

1. *train_bodies.csv*: contains the body text of articles with its ID
2. *train_stances.csv*: contains labeled stances for pairs of article headlines and article bodies, in which the article bodies refer to the bodies in train_bodies.csv

The distribution of the data is as follows:

Rows	Unrelated	Discuss	Agree	Disagree
49972	0.73131	0.17828	0.0736012	0.0168094

We will roughly use 4000 samples as our development set, for choosing hyperparameter and performance evaluation, and use the rest for our training data.

2.2 Methods

2.2.1 Pre-processing

The text from the corpus will be converted to tokens using *nltk* package and then be mapped to corresponding vectorized forms using pre-trained *GloVe* representations freely available on the Stanford NLP group website [3, 4]. Since the text sequences observed will be of variable length we will pad all sequences to the length of the maximum length text sequence before inputting it to our model. The data pre-processing includes normalizing the case, handling the punctuation and non-alphabetic symbols.

2.2.2 Learning Model

We will use a learning model that is RNN variant to predict the stances. The reason why it is based on RNN instead of CNN is that, for language modeling, RNN models are still the best approach [5]. The difficulty with RNN models is that they are hard to train because they suffer from the vanishing gradient problem [6]. The solution to this problem is to use RNN variant called long short-term memory (LSTM). However, recently a model called gate recurrent unit (GRU) was introduced by Cho et al. in 2014 [7]. GRU is similar with LSTM. Unlike LSTM, GRU combines forget and input gates into a single update gate and merges the cell state and hidden state. This makes GRU computationally more efficient than LSTM and GRU model has been increasingly popular [8]. For this project, we will use TensorFlow API for GRU, *tf.nn.rnn_cell.GRUCell* [9].

It is often said, that neural network performs better for a case like this. To compare the performance of GRU RNNs we will also perform logistic regression to generate stances as a baseline for our project by building it from scratch.

2.2.3 Scoring

To measure the accuracy, we will use the evaluation tool provided by Fake News Challenge. The tool will evaluate our model and output a score. The score is weighted as follows:

1. Classifying pair of body and headline as related and unrelated is weighted 25%
2. Classifying related pairs as agrees, disagrees, or discusses is weigted as 75%

The related/unrelated classification task is expected to be much easier while classifying the agrees, disagrees or discuss is more difficult and more relevant to fake news detection.

References

[1] Dean Pomerleau and Delip Rao. Fake news challenge.

[2] Michael Barthel, Amy Mitchell, and Jesse Holcomb. Many americans believe fake news is sowing confusion, Dec 2016.

[3] Steven Bird, Ewan Klein, and Edward Loper. *Natural Language Processing with Python*. O'Reilly Media, Inc., 1st edition, 2009.

[4] Jeffrey Pennington, Richard Socher, and Christopher D Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[5] Tomas Mikolov, Anoop Deoras, Stefan Kombrink, Lukás Burget, and Jan Cernocký. Empirical evaluation and combination of advanced language modeling techniques. In *INTERSPEECH 2011, 12th Annual Conference of the International Speech Communication Association, Florence, Italy, August 27-31, 2011*, pages 605–608, 2011.

[6] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

[7] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *CoRR*, abs/1406.1078, 2014.

[8] Rajib Rana. Gated recurrent unit (GRU) for emotion classification from noisy speech. *CoRR*, abs/1612.07778, 2016.

[9] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.