**Text Stance Detection and Classification**

Xiaowei Wu

wux@stanford.edu

**Introduction**

Automatic text stance detection and classification is very important in many applications including fake news detection. Fake news has becoming a great threat to our society nowadays with its intention to deceive. They are widely spread via online channels such as social media, causing a lot of confusion or even deliberately misleading public opinions. One of the most notable examples involves fake news on the 2016 US presidential election. Fake news can be easily identified with high accuracy by human fact checker. However, with the astronomical volume of online information, manual checking becomes almost impossible and the fact checking process has to be automated. Recent development in artificial intelligence provides a promising way to automate fake news identification. However, due to the complexity of natural language and the ways fake news is fabricated, accuracy of fake news identification is yet to be improved. As social media gaining more popularity and the effect of fake news becoming more detrimental, automatic fake news identification becomes a very important topic.

**Task Definition**

While fake news identification is a very complex process, understanding what other news sources are saying about the topic is an important part of process. In this project, the focus will be on this fundamental building block – stance detection, to classify the opinions on correlation between a given news body text and its headline. The data set provided Fake News Challenge [1] will be used for this task. It consists of 49972 instances, each with a headline and a body text as input and the stance as output. The input headline/body pairs are either from the same news article or from different articles, while four stance output categories denoting the correlation between headline and body are “agrees”, “disagrees”, “discusses”, and “unrelated”. Multiple models, such as support vector machine and multilayer perceptron are used for training and prediction. Based on the progress, the method developed here may also be extended to other applications such as online review classification such as Yelp review sentiment analysis [2].

**Approach**

In order to achieve stance detection, the Fake News Challenge organizer implemented a baseline algorithm using gradient boosting classifier. The stance included in the data set was created by accredited journalist, making it both high quality and credible. In order to achieve performance as close as possible to oracle, several learning methods are developed in this project. Specifically, support vector machines and neural network has been tested so far in this project.

Support vector machines are supervised learning algorithms used for classification and regression. SVM can perform linear binary classification. During learning stage all the features from the examples mapped into points in high dimensional space and these points are divided by a gap as wide as possible (Figure 1). During predicting stage, new examples can be mapped into the same high dimensional space and classified into one of the two categories by the model learned earlier. With non-linear kernel applied, SVM can also perform non-linear classification.

In this project, linear kernel and non-linear kernels are tested and compared for stance detection on this data set.

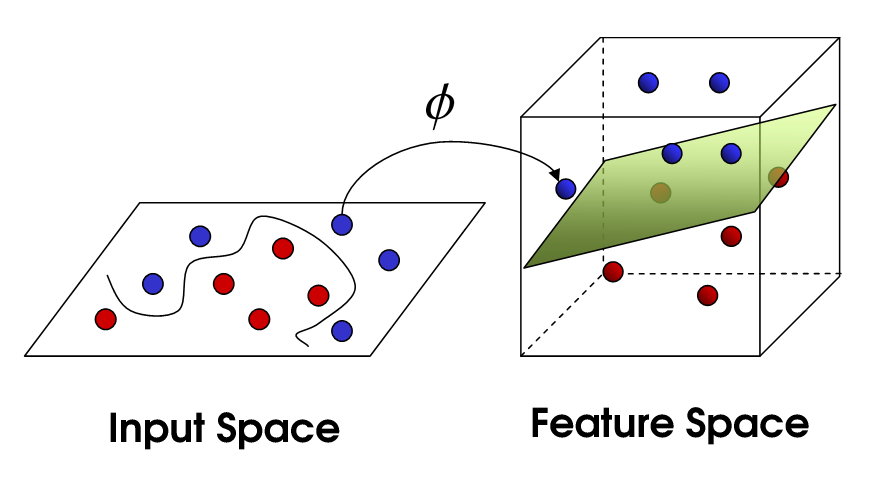


Figure 1: Features of the examples are mapped to points in high dimensional space (image from https://commons.wikimedia.org)

Inspired by biological neural network, neural network models are another powerful supervised learning algorithm. A neural network model consists of an input layer, one to several hidden layers and an output (Figure 2). Each hidden layer also consists of several nodes called neurons which use non-linear activation function. A multilayer perceptron is used in this project. The model takes all the features of a headline/body pair as the input, stance classification as the output and backpropagation for training.



Figure 2: schematic for artificial neutral network

**Data, Experiments, and Analysis**

The data downloaded from www.fakenewschallenge.org is (headline, body, stance) instances stored in two csv files. Each stance is one of {unrelated, discuss, agree, disagree}. The “train\_bodies.csv” includes news article body text and its body ID while the “train\_stances.csv” contains headlines and their corresponding body ID and stances. Each text body can correspond to multiple headlines and stances. Below is an example of the (headline, body, stance) instance.

**Body ID**: 4

**Headline**: It Begins: HazMat-Wearing Passenger Spotted At Airport

**Body**:

Last week we hinted at what was to come as Ebola fears spread across America. Today, we get confirmation. As The Daily Caller reports, one passenger at Dulles International Airport outside Washington, D.C. is apparently not taking any chances. A female passenger dressed in a hazmat suit - complete with a full body gown, mask and gloves - was spotted Wednesday waiting for a flight at the airport.

Source: The Daily Caller

We particularly liked the JCPenney bag - maybe that's a new business line for the bankrupt retailer...

**Stance**: discuss

Multiple features are extracted from each headline/body pair to generate a feature vector. For this progress report, only the baseline feature extraction is used (such as removing stop words and n-gram) and not optimized.

40350, 9622, and 25413 instances are randomly selected as the training, dev, and test sets, respectively.

SVM is implemented using sklearn [4]. Both linear kernel and a non-linear radial basis function kernel (RBF kernel) are used. The table below is the result on the test set using linear kernel:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label \ Predict | Agree | Disagree | Discuss | Unrelated |
| Agree | 71 | 6 | 1526 | 300 |
| Disagree | 15 | 0 | 443 | 239 |
| Discuss | 79 | 0 | 3689 | 696 |
| Unrelated | 2 | 0 | 413 | 17934 |

The results are evaluated based on a weighted scoring system proposed in Fake News Challenge site [1] – 25% score weighting from classifying headline and body text as related/unrelated and 75% from classifying related pairs as agrees, disagrees, or discuss. The above SVM with linear kernel got a score of 75.19%

The table below is the result on the test set using RBF kernel with a score of 73.54%:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label \ Predict | Agree | Disagree | Discuss | Unrelated |
| Agree | 61 | 2 | 1487 | 353 |
| Disagree | 12 | 0 | 407 | 278 |
| Discuss | 102 | 1 | 3496 | 865 |
| Unrelated | 0 | 0 | 314 | 18035 |

From the above two methods, the majority of the unrelated articles are corrected separated from the rest of the articles. Among the related articles, a large portion of discuss category is also separated from the rest. However, the agree and disagrees categories identification is not quite successful as the many instances in the agree and disagree categories are mis-classified as discuss or unrelated. This may be because the feature selection hasn’t been optimized yet. Also, the SVM parameters also need to be optimized on the learning set in the next few weeks to achieve better classification on the test score.

Neural network model is also tested on this data set. Multilayer perceptron is implemented using sklearn as well. ReLU activation function is used for this model. The table below is the result on the test set with a score of 74.51%:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Label \ Predict | Agree | Disagree | Discuss | Unrelated |
| Agree | 0 | 0 | 1575 | 328 |
| Disagree | 0 | 0 | 423 | 274 |
| Discuss | 0 | 0 | 3679 | 785 |
| Unrelated | 0 | 0 | 337 | 18012 |

Interestingly, all the agree and disagree categories are classified incorrectly although the unrelated category gets separated mostly correctly. Feature selection will be optimized in the next few weeks to solve this issue. Model parameters will also be adjusted using the dev set.

**Reference:**

[1] http://www.fakenewschallenge.org

[2] https://www.yelp.com/dataset/challenge

[3] Sobhani P., Mohammad S.M., and Kiritchenko S. (2016) Detecting stance in tweets and analyzing its interaction with sentiment. *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (\*SEM 2016)*,159–169.

[4] http://scikit-learn.org/stable/#