

EMPLOYING STANCE DETECTION APPROACH IN IDENTIFYING FAKE TEXT ARTICLES USING SUPERVISED MACHINE LEARNING

¹D. ANIL KUMAR, ²S. SANDHYA, ³T. HARSHA ESWARA TEJA, ⁴K. DIVYA SRI

¹ASSISTANT PROFESSOR, CSE DEPARTMENT, NRI INSTITUTE OF TECHNOLOGY, POTHAVARAPPADU

^{2,3,4}UG STUDENT, CSE DEPARTMENT, NRI INSTITUTE OF TECHNOLOGY, POTHAVARAPPADU

Abstract: Fake News is a real problem nowadays with the advent of social media and the lack of knowledge among the average populace. Fact checking can be a long and drawn out process and not every person has the time to verify the sources of every news article they come across. In such a scenario, our best bet is to develop and perfect the techniques in the areas of Artificial Intelligence and Natural Language Processing. For the purpose of this project, we will be focusing on “Stance Detection” stage of the identifying “Fake News” articles. The problem is quite well-defined, and a considerable amount of work has been done in this area. The input would be a corpus of headline-article pairs and the output will be the classification of the articles on the basis of whether their headline and content agree, disagree, discuss, or are unrelated. Our team focused on using different models for the above-mentioned task and compare the results to the baseline method provided by the organizers of Fake News challenge. This report summarizes our approaches, experiments, and results and we also discuss what are the state-of-the-art techniques for this task.

KEYWORDS: Fake News Challenge, Stance Detection, Support Vector Machines, Random Forests, Natural Language Processing, Machine Learning

1 INTRODUCTION

The Fake News Challenge was introduced with the goal of exploring different techniques in the field of Artificial Intelligence to solve the problem of “fake news”. The task is not trivial, even for humans as there is no specific boundary between a “real news” and a “fake news”. There are many stages in detecting the authenticity of a news article and for the purpose of our project, we have gone for the

first stage (FNC-1); stance detection. Stance Detection involves estimating the relative perspective (or stance) of two pieces of text relative to a topic, claim or issue. It involves estimating 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 [1]. In the following sections, we

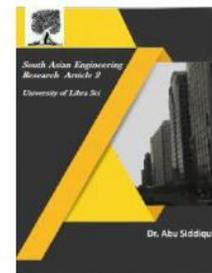


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



give a high-level overview of the data that we worked with.

INPUT AND OUTPUT

The input data was provided by the organizers of Fake News challenge. The training data is divided into two sets; bodies and headlines with stance labels. The set “train_bodies.csv” contains the data in two columns: Body ID and Article Body. Each body has a unique identifier number so that it can be matched up against the headline in the “train_stances.csv” which has three columns: Headline, Body ID and Stance. To create the training corpus, we joined the two sets on the column “Body ID”. In the training corpus that we obtained, the distribution of the 4 stances were as follows:

Stance (for a headline-body pair)	Count	Percentage of the total dataset
Agree	3612	7.3
Discuss	8956	17.8
Disagree	754	1.68

Unrelated	36334	73.13
-----------	-------	-------

2 RELATED WORK

We read up on many different research papers and studied a variety of approaches to tackle the problem at hand. Eventually, we decided to use a few different models and perform our own experiments as attempting the replicate the state-of-the-art technique in this area was not feasible due to the time constraint. The Fake News challenge has been attempted by numerous individuals and groups and there was a wide array of research papers, articles and blogs on the topic. The organizers themselves provided a baseline implementation to get the participants started. This baseline implementation (FNC-1) used Gradient Boosting classifier for the task. The accuracy is 79.52%. We utilized some of the features from FNC-1 and also generated our

own features as we will discuss in the subsequent sections. The work of [2] used a bidirectional LSTM with global features and it was claimed to have outperformed the baseline implementation provided by the organizers. It works almost perfectly on identifying the related/unrelated relationship for the pairs. The work of [2] shows that recurrent neural network models with “memory-enabled” units like LSTMs significantly outperform non-recurrent models on the sequence-based task of stance detection. [3] showed that Conditional Encoding LSTM with Attention gave a better result than the baseline implementation even if it was only by a small margin. After our extensive researching and studying of various techniques used in this field, we have concluded that the state-of-the-art technique to tackle the Fake News challenge is to use bag-of-words followed by a three-layer multilayer perceptron (BoW MLP). According to the work of [4], The BoW MLP model achieved categorical test-set accuracy of 93%, and on the competition-specific FNC-1 metric it achieved a test-set score of 89%—a full ten percentage points above the published baseline [4].

3 METHODOLOGY

After performing the preprocessing steps as mentioned in the above section, the first step we took was the vectorization of the headline-body pairs. Our modeling experiments make use of these vectorized fields for the distance comparison features. We used the pre-trained Google News corpus word vector model for word embeddings. Every word vector had 300 dimensions.

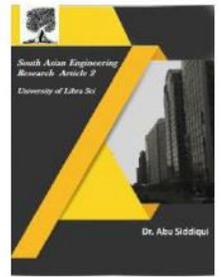


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



FEATURES USED

We used some features from the baseline implementation which were the hand-crafted linguistic features. While we experimented with many different features, eventually we selected the best of the lot and used them in our system. Some example features that we used are listed below:

- Euclidean Distance as the distance comparison feature
- Refuting words from FNC-1
- Word N-grams (n = 2, 3, 4) extracted from the headline and body pair and finding the overlapping grams
- Lengths of headline and its body

Some of the features that we experimented with but did not include in the final version of our system are: polarity features, character grams, binary co-occurrence features, Word Mover's distance[5], Cosine Distance etc. We tried different permutations of all the above listed features and judged some features to be more useful than others. For instance, as a distance feature, the Word Mover's distance simply did not capture the relationship between the headline and its body properly. When two vectors point in the same direction, cosine distance proved to be a better measure. One of the salient features of the cosine distance metric is that two words with similar senses will have their relation represented appropriately by their cosine distance regardless of their frequencies in the document. This makes it highly desirable for our task where we had to deal with rare and uncommon words. We used Euclidian distance metric as well as the cosine metric could not capture some important aspects such as word significance or consistency in word usage and the Euclidean metric captured that better. Moving on, the refuting

features that we used were taken from FNC-1 without change. Refuting words such as “fake”, “fraud”, “hoax”, “not”, “denies” etc. have the potential of providing very useful information about the relationship between the headline and its body. The use of only refuting features, coupled with overlapping words feature and a few other features, performed very well in identifying whether a headline-body pair is related or not. The overlapping ratio is calculated by finding the common token for the headline-body pair divided by the total number of tokens generated by the pair. The overlapping ratio turned out to very different for unrelated stances from the other three classes. This was the first stage of our experimenting.

RANDOM FORESTS

The Random Forest (RF) model grows many classification trees. To classify a new object from an input vector, it puts the input vector down each of the trees in the forest. Each tree gives a classification, and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest). The RF is very robust and not prone to overfitting. When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. This out-of-bag data is used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance. After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, the proximities are

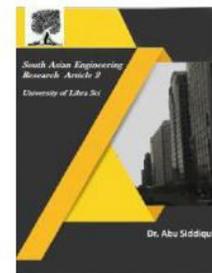


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



normalized by dividing by the number of trees. Proximities are used in replacing missing data, locating outliers, and producing illuminating low-dimensional views of the data. The RF is quite accurate, efficient and also give estimates of what variables are important in the classification. We have provided the following table with the features and their corresponding importance score:

Feature Name	Score
Overlapping Ratio	0.257
Unigram	0.256
Bigram	0.097
Trigram	0.018
Euclidean	0.156
Refuting feature	0.008
Headline size	0.099
Body size	0.102

4 RESULTS

The way to evaluate our model is to use the official metric of the Fake News Challenge. The metric used is a score based on how many predictions the model got right. The different stance classes were weighted. If the model correctly identifies the related/unrelated stance for a headline-body pair, it is awarded 0.25 points and if it correctly predicts any of the other three classes, it is awarded 0.75 points. The reason for this is that the task of determining unrelated/related is trivial compared to determining other classes. The baseline, as mentioned above, achieves a score of 79.53% as per the evaluation scheme described. In the following sections, we present how each of our models did using the same scoring scheme. After performing the selection of features based on their importance as shown above, we received 86% accuracy on the test set. This is higher than the SVM which we will discuss next. We used 200 trees in our forest. We observed no significant difference in performance if we used more than 200 trees.

To improve our accuracy for the “agree” and “disagree” stances, we explored bootstrap, and different maximum depth parameters but no significant improvement in predictive ability was observed. The confusion matrix is given as follows with the gold labels being represented by the column headers.

Stance	Gold			
	Agree	Discuss	Disagree	Unrelated
Agree	199	454	1	68
Discuss	106	1576	8	126
Disagree	27	117	21	27
Unrelated	18	100	1	7146

CONCLUSION

Our experiments turned out to be successfully in terms of beating the baseline score and understanding which features are useful and which are not. The two models that we used are standard models in the machine learning community. In general, methods based on bag-of-words, Tf-Idf, n-grams, and topic models work well for FNC-1. Bidirectional LSTMs (Long short-term memory) are a very popular choice was tasks like the FNC-1. The work of [2] used a conditioned bi-directional LSTM with global features which gave an impressive score of 87.4% on the FNC metric. The state-of-art model, as briefly discussed above is a bag-of-words followed by a three layer multi-layer perceptron (BoW MLP). In the work of [4], the BoW MLP model outperformed other models such as the concatenated multilayer perceptron (Concat MLP), dual GRU, and the bidirectional concatenated stacked LSTM. It gave an accuracy of 89% which is the highest score in this area to our knowledge. All the recurrent neural network models fell short of BoW MLP’s performance even after extensive parameter tuning. In the future, we would like to use BoW MLP and other deep learning models for this task. It was found

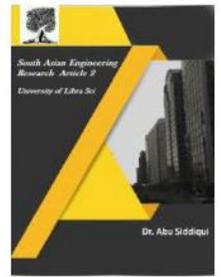


2581-4575

International Journal For Recent Developments in Science & Technology



A Peer Reviewed Research Journal



that word frequency is a very effective feature for these models; so that is something we would use for our models and perform the experiments. We would also like to use other syntactic features such as Part-of-speech-tagging to see how it affects our models' performance. We would like to implement post-facto "truth labelling" from our stance detection system. Such a system would tentatively label a claim or story as true/false based on the stances taken by various news organizations on the topic, weighted by their credibility. Our system gave some us some very useful information regarding the importance of different features for stance detection. We plan to keep these results in our minds as we move on to develop better systems.

REFERENCES

- [1] <http://www.fakenewschallenge.org/>
- [2] Mrowca, D., Wang, E., & Kosson, A. Stance Detection for Fake News Identification.
- [3] Pfohl, S., Triebe, O., & Legros, F. Stance Detection for the Fake News Challenge with Attention and Conditional Encoding.
- [4] Davis, R., & Proctor, C. Fake News, Real Consequences: Recruiting Neural Networks for the Fight Against Fake News.
- [5] Kusner, M., Sun, Y., Kolkin, N., & Weinberger, K. (2015, June). From word embeddings to document distances. In International Conference on Machine Learning (pp. 957-966).
- [6] Chaudhry, A. K., Baker, D., & Thun-Hohenstein, P. Stance Detection for the Fake News Challenge: Identifying Textual Relationships with Deep Neural

Nets.

- [7] Bourgonje, P., Schneider, J. M., & Rehm, G. (2017). From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles. In Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism (pp. 84-89).
- [8] Miller, K., & Oswald, A. Fake News Headline Classification using Neural Networks with Attention.