Fake News Identification Using Stance Classification

By Jason K. Love
Id: z5181392
Course: Comp 9417, Machine Learning

Aim

The goal of this project is to build a multi-classification model to judge the stance of a news headline against a variety of article bodies.

Introduction

In our lives and in the world around us, whether the real or online, we come across many items of news. However, it can sometimes be hard to tell if the news is real, gossip, propaganda, or just plain fiction. An example of this is the infamous radio broadcast of H.G. Wells' War of the Worlds whose listeners reportedly could not tell the veracity of claims.

The New York times defined Fake News[4] to be "a made-up story with the intention to deceive". This is especially an issue in the modern era where 'news' can bypass verification procedures used by '*traditional* outlets'. The 'deception' can in turn lead to a great deal of confusion as the fake news gets popularised and re-posted, whether unknowingly or not, as was found in the study [2] 64% of Americans across ages, income groups, race, and political inclination.

A way to help in dealing with this issue, and one of the main objectives of computational journalism, is to make use of machine learners to preform automated fact-checking of the news stories. [1], Due to the inherent open nature of the task and limitations in current natural-language-processing methods, however, deducing the veracity of news articles is still too challenging at this stage.

We can, though, make headway into solving the issue by breaking down and looking at smaller parallel tasks. This would allow us at the very least produce automated systems that can be of aid and simplify or quicken the process for human fact checkers. The process taken by the Fake News Challenge in stage 1 of their competition[4], was to use a process called Stance Deduction. Stance Detection involves comparing two pieces of texts in relation to a topic, claim or issue and deducing their relative stance. In the case of the challenge it checks if they agree or disagree, are unrelated to each other or discuss each other without taking sides.

Table 1: example of stances

Headline: "Robert Plant Ripped up \$800M Led Zeppelin Reunion Contract"	
Snippets from body texts	Correct Classification
"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."	AGREE
"... No, Robert Plant did not rip up an \$800 million deal to get Led Zeppelin back together. ..."	DISAGREE
"... Robert Plant reportedly tore up an \$800 million Led Zeppelin reunion deal. ..."	DISCUSSES
"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."	UNRELATED

Method

For the coding aspect of the project the feature extraction was designed from scratch then put through a machine learner taken from the third party software sci-kit learn, via stratified folding.

The data set was made up in two sets; one containing the article bodies, one containing the stances to the headline. After collating the training data between the stances and articles, the training/testing split was made in a 80/20 ratio. This was retained so as to do the following model selection

It was decided to use 10-fold-Cross Validation to run the ML across the 80% training set. Stratified was not done given the nature of the dataset design.

To select a suitable machine learner runs were done using the baseline feature extraction on LinearSVC, Gaussian Naive Bayes, Stochastic Gradient Descent Classifier, K-Nearest Neighbour, Multi-Layer Perceptron classifier, and an ensemble method RandomForestClassifier. These classifiers were chosen due to the inherent capability of dealing with multi-class data.

To compare with methods used by [1] and[4] Logistic regression and Gradient Decent Classifiers were utilised.

A selection was then made based on the highest CV-score as calculated by the FNC-1 scorer. As well as the highest score against the initially removed 20% test set.

The chosen classifier was then tested on the competition set to compare the results to the leader board

Finally, the chosen ML was re-fitted with redesigned feature extraction and run.

Edited features include Word-1-skip-n-grams, a basic sentiment analyser and Term Frequencies.

Results

It was found that the ensemble method worked best giving a score of 3526.75 out of 4448.5 (or 79.27%) on the remaining 20% hold test set. This is a touch less then the FNC-1 base line found using the ensemble method Gradient Decent Classifier which gave a result of (79.53%).

The RandomForestClassifier was then trained on the entire set and tested on the competition test set. This produced a result of 8628.0 out of 11651.25 (or 74.05%). This result gives the place of 30^{th} on the top 50 leader board of the competition. As a baseline the minimal score was found by predicting everything as unrelated (i.e. the “all-false” baseline) this produced a result of 4460.25 out of 11651.25 (or ~40%). But while this is better then the research by [1] this is also bellow the FNC-1 baseline of 75.2%. (as can be seen in Appendix 2)

So carrying on for the next stage the Baseline GradientBoostingClassifier was used. Although ML was sent to refit, this was done too late and system was too slow in finding TF-IDF’s resulting in not being able to get final results of personally designed feature extraction on time.

Related Work

The top three of FNC-1 have their work open sourced on git-hub as per requirement of FNC. The methods used by the top two incorporate ensemble with Deep-Learning, and Multi-Threading-Feed-Forward respectively. With the third using a novel approach with Term-Frequency vectors and one-hidden-layer Multi-Layer-Perceptron. The third approach scored highly due to the close to perfect binary classification of related and unrelated (accuracy: 96.55%) [3]

Conclusion

In conclusion, the task is one that a basic ML on a basic feature set can not do better than the given baseline . But it is not necessary to make a complex Machine Learner Model as from editing the backend NLP performance can improve drastically as shown by [3] . However they are slower tasks.

References:

[1]W. Ferreira and A. Vlachos, "Emergent: a novel data-set for stance classification", *Aclweb.org*, 2016. [Online]. Available: http://aclweb.org/anthology/N/N16/N16-1138.pdf. [Accessed: 30- May- 2018].

[2]M. Barthel, A. Mitchell and J. Holcomb, "Many Americans Believe Fake News Is Sowing Confusion", *Pew Research Center's Journalism Project*, 2016. [Online]. Available: http://www.journalism.org/2016/12/15/many-americans-believe-fake-news-is-sowing-confusion/. [Accessed: 31- May- 2018].

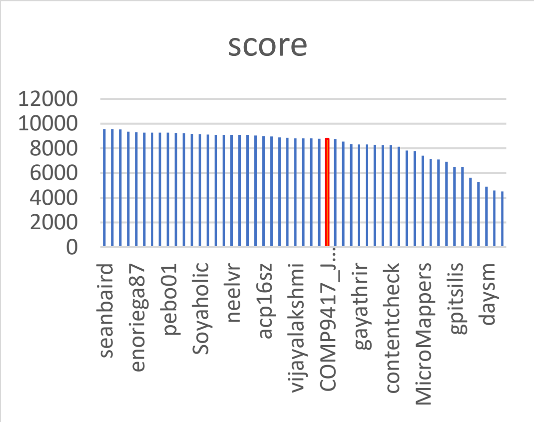
[3]B. Riedel, I. Augenstein, G. Spithourakis and S. Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task", *Arxiv.org*, 2018. [Online]. Available: https://arxiv.org/pdf/1707.03264.pdf. [Accessed: 01- Jun- 2018].

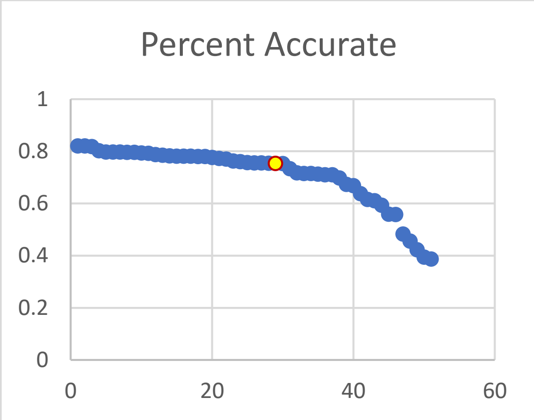
[4]"Fake News Challenge", *Fakenewschallenge.org*, 2016. [Online]. Available: http://www.fakenewschallenge.org. [Accessed: 30- May- 2018].

[5]"Cisco-Talos/fnc-1/tree_model", *GitHub*, 2018. [Online]. Available: https://github.com/Cisco-Talos/fnc-1/tree/master/tree_model. [Accessed: 02- Jun- 2018].

APPENDIX:

1. **Result placement (**Highlighted Yellow-red is where I would be)





2. **Result Confusion Matrix**

<table>
<tr>
<td>My-Results</td>
<td>Scores on the dev set

| | agree | disagree | discuss | unrelated |

| agree | 153 | 1 | 517 | 91 |

| disagree | 23 | 3 | 116 | 20 |

| discuss | 114 | 5 | 1481 | 200 |

| unrelated | 9 | 0 | 106 | 6783 |

Score: 3526.75 out of 4448.5 (79.27953242666067%)</td>
<td>Scores on the test set

| | agree | disagree | discuss | unrelated |

| agree | 204 | 5 | 1379 | 315 |

| disagree | 47 | 2 | 403 | 245 |

| discuss | 351 | 8 | 3381 | 724 |

| unrelated | 29 | 4 | 345 | 17971 |

Score: 8628.0 out of 11651.25 (74.05214032829096%)</td>
</tr>
<tr>
<td>FNC-1 Baseline</td>
<td>Scores on the dev set

| | agree | disagree | discuss | unrelated |

| agree | 118 | 3 | 556 | 85 |

| disagree | 14 | 3 | 130 | 15 |

| discuss | 58 | 5 | 1527 | 210 |

| unrelated | 5 | 1 | 98 | 6794 |

Score: 3538.0 out of 4448.5 (79.53242666067214%)</td>
<td>Scores on the test set

| | agree | disagree | discuss | unrelated |

| agree | 173 | 10 | 1435 | 285 |

| disagree | 39 | 7 | 413 | 238 |

| discuss | 221 | 7 | 3556 | 680 |

| unrelated | 10 | 3 | 358 | 17978 |

Score: 8761.75 out of 11651.25 (75.20008582770089%)</td>
</tr>
</table>