Fake News Challenge

Prajwal Rao (5176504) & Julian Blair (3463793)  
COMP9417 2018s1 – Assignment 3, Topic 1.5

# Background

* Brief description of challenge
* Baseline provided
  + Explain how it works

# Implementation

## Problem structure hypothesis

* Divide and conquer approach – break up multi-class problem into several two-class problems

## Data Munging

* Common word filter:
  + Looking for words with high variance and low frequency (avoid common, homogeneous words)
  + Related to stop words in English, but specialised for entropy of words in articles

## Related vs. Unrelated

* Simplified baseline to two features (word overlap, headline-in-body word counts), added common word filter

## Multinomial Bayes Stacking

* Mention stuff here

## Overfitting Issues

* Attempted punctuation counts, estimations on opinion lexicons, all overfit and reduced competition test data score

# Results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Actual╲Predict | Agree | Disagree | Discuss | Unrelated |
| Agree | 971 | 26 | 748 | 158 |
| Disagree | 293 | 17 | 262 | 125 |
| Discuss | 924 | 22 | 3149 | 369 |
| Unrelated | 157 | 6 | 400 | 17786 |

* Some tables containing baseline results + divide-and-conquer results
* Would have come 13th out of 80 participants

# Acknowledgements

The Fake News Challenge was hosted in 2017 by a group of academic and industry volunteers. Thanks to the same team for providing a baseline implementation, which was used as a starting point for our project.

# References

CodaLab (2017) *Fake News Challenge Stage 1 (FNC-I) - Stance Detection*, June, [Online], Available: <https://competitions.codalab.org/competitions/16843#results> [29 May 2018].

Galbraith, B., Iqbal, H., van Veen, H., Rao, D., Thorne, J. and Pan, Y. (2017) *Baseline FNC implementation*, 25 July, [Online], Available: <https://github.com/FakeNewsChallenge/fnc-1-baseline> [28 May 2018].

Pomerleau, D. and Rao, D. (2017) *Fake News Challenge*, [Online], Available: <http://www.fakenewschallenge.org/> [28 May 2018].