Hyperpartisian News detection

Florian Biebel   Daan Middendorp

July 8, 2019

Table of contents

Introduction

- Hyperpartisian News
    - "Extremely biased in favor of a political party"
- SemEval 2019 competition
- Datasets
    - By article
    - By publisher

Attempts

- Classifiers
  - Logistic Regression
  - SGD Classifier
  - Random Forest Classifier
  - Naive Base Classifier
  - CNN Keras
- Feature Unions
  - Length of article
  - Number of capitalized words
  - Number of exclamation marks

Length of article

- Hyperpartisian
  - Average of 596 words per article
  - Average of 7 capitalized words
- Non-Hyperpartisian
  - Average of 415 words per article
  - Average of 5 capitalized words

- How to tokenize?

Scores (20 runs)

Method	Acc.	Prec.	Rec.	F1
Logistic regression	0,70-0,82	0,65-0,81	0,38-0,60	0,51-0,70
SGD Classifier	0,62-0,76	0,51-0,79	0,52-0,78	0,59-0,70
Random Forest	0,63-0,78	0,60-0,88	0,27-0,47	0,41-0,59
Naive Bayes	0,65-0,78	0,50-0,75	0,57-0,80	0,59-0,71
Log. reg. FU	**0,71-0,82**	**0,67-0,92**	**0,39-0,64**	**0,49-0,70**
SGD FU	0,34-0,70	0,0-0,40	0,0-1,0	0,0-0,58

Scores