

Vernon-fenwick at SemEval-2019 Task 4: Hyperpartisan News Detection using Lexical and Semantic Features

Vertika Srivastava
Sudeep Kumar Sahoo

Ankita Gupta
Rohit R.R

Divya Prakash
Yeon Hyang Kim

Samsung R&D Institute India, Bangalore

{v.srivastava, gupta.ankita, p.divya,
sudeep.sahoo, rohit.r.r, purine.kim}@samsung.com

Abstract

In this paper, we present our submission for SemEval-2019 Task 4: Hyperpartisan News Detection. Hyperpartisan news articles are sharply polarized and extremely biased (one-sided). It shows blind beliefs, opinions and unreasonable adherence to a party, idea, faction or a person. Through this task, we aim to develop an automated system that can be used to detect hyperpartisan news and serve as a pre-screening technique for fake news detection. The proposed system jointly uses a rich set of handcrafted textual and semantic features. Our system achieved 2nd rank on the primary metric (82.0% accuracy) and 1st rank on the secondary metric (82.1% F1-score), among all participating teams. Comparison with the best performing system on the leaderboard¹ shows that our system is behind by only 0.2% absolute difference in accuracy.

1 Introduction

Today in the age of digitization, a smartphone has become an indispensable tool for information sharing and consumption. It is much more convenient for users to read news through online articles and social media platforms. These platforms provide them quick and easy access to information almost everywhere. However, it is highly possible that the shared information is unverified and may bias the reader's opinions. This issue is exacerbated by the fact that most of the people find it difficult to distinguish between what's real and objective, what's fake and what's partisan.

Although these online news articles are expected to be written well-balanced without any prejudices, the authors/media houses may at times spill their standpoints and beliefs. Instead of providing a holistic view to the readers, the author

¹<https://pan.webis.de/semeval19/semeval19-web/leaderboard.html>

tries to convey a picture with which he agrees and thus making the audience biased towards a party or a faction. When these news articles are extremely polarized towards one side of the argument, they are referred to as "Hyperpartisan News Articles".

This extreme polarization can leave users vulnerable to detrimental arguments and cloud their judgment to make objective decisions. These hyperpartisan articles may also carry some common elements of fake news. They are typically used to spread propaganda and manipulate readers. It targets human psychology by creating confirmation bias and echo chambers and therefore impairing their ability to dispel the hyperpartisan articles in favor of neutral articles.

SemEval-2019 Task 4 aims to solve this issue of hyperpartisanship. The objective of this task is to detect if a given news article has hyperpartisan arguments (extremely one-sided). In this paper, we have described our system that automates the process of identifying and annotating news articles as hyperpartisan or not.

2 Related Work

The problem of fake news and hyperpartisan has been discussed earlier by Potthast et al. (2017). They followed a style based approach to tackle the problem and have also suggested that writing style of left-wing and right-wing news are quite similar. Apart from this, there hasn't been much work in hyperpartisan news detection. Our approach is inspired by some recent work in the domain of sentiment analysis (Pontiki et al., 2016) and bias detection (Patankar et al., 2018; Recasens et al., 2013; Patankar and Bose, 2017; Baly et al., 2018). Jian and Wilson (2018) have explored linguistic signals embedded in news articles. They have used these linguistic clues to detect the spread of misinformation via social media. Iyyer et al. (2014) have

studied the impact of words in identifying people’s ideology and have proposed RNN to capture semantic features of a sentence.

3 System Description

Our hyperpartisan news detection system consists of three phases: 1) preprocessing, 2) generating article representation, and 3) training a classifier. An overview of our system is shown in Figure 1.

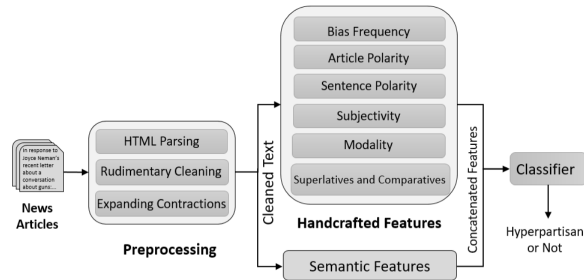


Figure 1: System Pipeline.

3.1 Preprocessing

The original dataset consisted of news articles along with HTML tags. Hyperpartisan news detection cleaner² was used to convert original text to plain text by removing tags. In the next step, further rudimentary cleaning of articles was done. We have also expanded contractions in the dataset like ‘shan’t’ was converted to “shall not”, ‘don’t’ to “do not” etc.

3.2 Article Representation

Hyperpartisan news articles are extremely one sided and exhibits blind beliefs as compared to a neutral news articles. We have observed that writers of such articles generally, manifest usage of harsh tone and inflammatory language, they even exaggerate and convey opinions to stress their ideology. Hyperpartisan news articles tend to use superlatives and comparatives frequently to dramatize or exaggerate situations.

Polarity at an article level can capture emotions and sentiments of the article, it may also capture contextual polarity. Polarity at sentence level helps to identify bias localized to a sentence which might not be perceptible at an article level. It helps to shift the focus on bias-heavy sentences. By combining polarity at both levels, we can capture the tone, overpraise and sentiment in an arti-

²<https://github.com/webis-de/semEval19-hyperpartisan-news-detection-article-cleaner>

cle. Subjectivity, modality and bias lexicons help to discover the attitude, prejudice and beliefs expressed by the author. Building from these insights, we created a set of handcrafted textual features (HF), which are based on writing style, linguistics, and lexicons.

The problem with handcrafted features is that they don’t capture the semantic relationship among the sentences. To solve this problem, we incorporated semantic features (SF) that can capture long-range dependencies of sentences and bring out the semantics of the article. The drawback of semantic features obtained via word based embeddings like Glove, is that it ignores word sequencing. In our approach, we have also explored features generated via distributed document representation (Universal Sentence Encoder or Doc2Vec) that are agnostic to word orderings and captures the semantics of an article.

Consider a set of N news articles $A = \{a_1 \dots a_N\}$. Each article a_i has a set of $S = \{s_1 \dots s_m\}$ sentences and a set of $W = \{w_1 \dots w_l\}$ words, where l is the length of the article. We jointly used HF and SF to obtain article representation ($ArtRep$), where \oplus is concatenation operator:

$$ArtRep = HF \oplus SF \quad (1)$$

Handcrafted features used in our system is described in Section 4 and semantic features are discussed in Section 5.

4 Handcrafted Features

Bias Score: For identifying bias words in an article, bias lexicon built from NPOV corpus of Wikipedia articles (Recasens et al., 2013) was used. Wikipedia advocates Neutral Point of View policy (NPOV), articles falling under NPOV dispute category were used to build this corpus. Bias score is the frequency of article words that occur in bias lexicon.

Article Level Polarity: Polarity of an article ($APol$) was extracted using MPQA Subjectivity lexicon (Wilson et al., 2005) ($SLex$), which lists around 8000 words with their prior polarity and their subjectivity type. Let a set of prior positive polarity words in an article be $PLex_i$ and negative polarity words be $NLex_i$. We computed positive ($APol_i^+$) and negative polarity score ($APol_i^-$) of

an article a_i , where $\mathbb{1}$ is an indicator function:

$$APol_i^+ = \frac{1}{l} \sum_{j=1}^l \mathbb{1}(w_j \in (PLex_i \cap SLex)) \quad (2)$$

$$APol_i^- = \frac{1}{l} \sum_{j=1}^l \mathbb{1}(w_j \in (NLex_i \cap SLex)) \quad (3)$$

Sentence Level Polarity: Polarity was further fine-grained to generate features for sentences of a news article using Pattern toolkit for English³. A sentence s_j was given as an input to the toolkit and a polarity score α_j in the range of $[-1.0, 1.0]$ was obtained. Positive ($PolScore_i^+$), negative ($PolScore_i^-$) and neutral polarity score ($PolScore_i^{Neu}$) of an article a_i was computed with $|0.1|$ as a threshold as it gave the best results for our system:

$$PolScore_i^+ = \sum_{j=1}^m \mathbb{1}(\alpha_j > 0.1) \quad (4)$$

$$PolScore_i^- = \sum_{j=1}^m \mathbb{1}(\alpha_j < -0.1) \quad (5)$$

$$PolScore_i^{Neu} = \sum_{j=1}^m \mathbb{1}(-0.1 \leq \alpha_j \leq 0.1) \quad (6)$$

Subjectivity and Modality: Subjectivity score is computed using Sentiment module of Pattern toolkit for English³. Toolkit gives a score based on adjectives and their context in the range of $[0.0, 1.0]$. Modality is a measure of the degree of certainty. It was computed using Modality module of Pattern toolkit for English³.

Superlatives and Comparatives : Intensifying lexicons like adjectives and adverbs in superlative and comparative degree were used. We ran POS tagger from NLTK (Bird and Loper, 2004) on the article text to identify Subjective and Comparative adjectives and adverbs and their corresponding frequencies in the text were used as a feature.

5 Semantic Features

5.1 Glove

Glove (Pennington et al., 2014) provides distributional vector representations of words in the

³<https://www.clips.uantwerpen.be/pages/pattern-en>

Dataset	Articles
Hyperpartisan	238
Non-Hyperpartisan	407

Table 1: ByArticle Dataset statistics.

semantic space. We have used 300-dimensional Glove embeddings trained on Common Crawl data of 2.2 million words and 840 billion tokens. An article was tokenized into sentences and further into words to obtain its article representation. Each of these words was vectorized using Glove pre-trained embeddings. Article representation was generated by averaging (Wieting et al., 2015) these 300-dimensional word embeddings.

5.2 Doc2Vec

Doc2Vec (D2V) (Le and Mikolov, 2014) is an unsupervised algorithm to learn distributed representation of multi-word sequences in semantic space. We have used Python implementation of Doc2Vec provided by gensim to learn embeddings for the news articles. For our experiments, we have used 512-dimensional embeddings generated from D2V on article text.

5.3 Universal Sentence Encoder

Universal Sentence Encoder (USE) (Cer et al., 2018) is a pretrained model to generate embeddings for sentences, phrases and much larger multi-word sequences. It has shown good performance on diverse NLP tasks for e.g., phrase level opinion extraction and sentiment classification. USE takes English text as an input and generates 512-dimensional embedding. In our best system, we have used these 512-dimensional article embeddings, generated on feeding article text to USE.

6 Experiments

6.1 Dataset

We have used *ByArticle* dataset provided in the task, which is labeled through crowdsourcing. Dataset consists of 645 news articles and a label to denote if it is hyperpartisan or not, details of the dataset are provided briefly in Table 1. More information on the dataset can be found in (Kiesel et al., 2019).

Model	Acc.	Prec.	Recall	F1
Baseline	46.18	46.03	44.27	45.13
1st ranked system	82.17	87.13	75.48	80.90
SM	58.76	58.51	60.19	59.34
SC	65.76	70.37	54.46	61.40
B	68.63	73.31	58.60	65.13
SP	68.63	72.24	60.51	65.86
AP	65.13	64.90	65.92	65.40
HF	70.22	72.76	64.65	68.47
D2V+HF	73.41	69.60	83.12	75.76
Glove+HF	78.34	82.01	72.61	77.03
USE+HF	82.01	81.50	82.80	82.15

Table 2: SemEval-2019 Task 4 Performance comparison on hidden test data (SM: Subjectivity and Modality, SC: Superlatives and Comparatives, B: Bias Score, AP: Article Level Polarity, SP: Sentence Level Polarity, HF: set of all Handcrafted Features). All results reported in the table are in percentage and rounded to 2 decimal places.

6.2 Results and Analysis

Handcrafted features were concatenated with semantic features, to generate a rich article representation which was fed to the classifier as an input. A L2-regularized logistic regression (Pedregosa et al., 2011) classifier was trained with 10-fold cross-validation. Since the training dataset is unbalanced, we have used class weighted logistic regression by weighing classes inversely proportional to their frequency.

For evaluation, a balanced hidden test data (*ByArticle*) was provided by the organizers through TIRA (Potthast et al., 2019). We have used accuracy for performance evaluation, which was a primary performance measure in the task. Apart from this, we have also reported precision, recall, and F1-score on the dataset.

Table 2 shows the performance of our various approaches against the baseline results (Task 4 semeval-pan-2019-baseline on TIRA⁴). From the results, we can observe that our best performing system has outperformed the baseline by a huge margin, recording an absolute jump of 35.83% accuracy. We can also see that our system is as good as the best system (bertha-von-suttner⁴) submitted for the task, which has 82.17% accuracy, showing that we were behind by only 0.16%.

In order to assess the importance of each hand-

⁴<https://www.tira.io/task/hyperpartisan-news-detection/dataset/pan19-hyperpartisan-news-detection-by-article-test-dataset-2018-12-07/>

crafted feature, we performed experiments by using individual features and their combination (HF). In fact, Table 2 highlights that all the features jointly perform well with 70.22% accuracy. We found that bias lexicon and polarity based features are the most informative ones.

In an attempt to further improve our model’s performance, we experimented by combining semantic features with handcrafted features as described in Section 3.2. Our results have also shown that handcrafted features alone aren’t sufficient and better performance can be achieved by combining them with semantic features. USE+HF has beaten all our other models and showed an absolute improvement of 11.79% accuracy over the HF model. It also attained 1st rank on the leaderboard (based on F1-score) and 2nd overall rank (based on Accuracy).

The poor performance of D2V+HF can be attributed to the training of D2V as typically D2V requires large data for training. Results have also shown that USE+HF performed better than Glove+HF, and thus validating our earlier claim of limitations of word-based embeddings.

7 Conclusion and Future Work

In this paper, we proposed a novel approach to detect hyperpartisan arguments in a news article. Our system ranked 2nd in SemEval-2019 Task 4. Our approach leverages rich semantic and handcrafted textual features. In the paper, we have also studied the importance of capturing semantic relationship among sentences of an article. Our system employed linguistic and lexical features to detect polarity, sentiments and blind beliefs exhibited in an article. Experiments with various model configurations demonstrated the effectiveness of our approach.

Detecting hyperpartisan in news articles should also, involve incorporation of world knowledge, as statements as an individual may not be extremely biased but when seen from a global perspective, they turn out to be hyperpartisan. As a future work, we would like to exploit the use of external knowledge. We would also like to investigate the role of credibility of news sources (Baly et al., 2018; Popat et al., 2018) in detecting hyperpartisan news articles.

References

- Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018. Predicting factuality of reporting and bias of news media sources. *arXiv preprint arXiv:1810.01765*.
- Steven Bird and Edward Loper. 2004. Nltk: the natural language toolkit. In *Proceedings of the ACL 2004 on Interactive poster and demonstration sessions*, page 31. Association for Computational Linguistics.
- Daniel Cer, Yinfei Yang, Sheng-yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, et al. 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.
- Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1113–1122.
- Shan Jiang and Christo Wilson. 2018. Linguistic signals under misinformation and fact-checking: Evidence from user comments on social media. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):82.
- Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 Task 4: Hyperpartisan News Detection. In *Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval 2019)*. Association for Computational Linguistics.
- Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *International Conference on Machine Learning*, pages 1188–1196.
- Anish Anil Patankar and Joy Bose. 2017. Bias discovery in news articles using word vectors. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 785–788. IEEE.
- Anish Anil Patankar, Joy Bose, and Harshit Khanna. 2018. A bias aware news recommendation system. *arXiv preprint arXiv:1803.03428*.
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.
- Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, AL-Smadi Mohammad, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, pages 19–30.
- Kashyap Papat, Subhabrata Mukherjee, Andrew Yates, and Gerhard Weikum. 2018. Declare: Debunking fake news and false claims using evidence-aware deep learning. *arXiv preprint arXiv:1809.06416*.
- Martin Potthast, Tim Gollub, Matti Wiegmann, and Benno Stein. 2019. TIRA Integrated Research Architecture. In Nicola Ferro and Carol Peters, editors, *Information Retrieval Evaluation in a Changing World - Lessons Learned from 20 Years of CLEF*. Springer.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylistometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1650–1659.
- John Wieting, Mohit Bansal, Kevin Gimpel, and Karen Livescu. 2015. Towards universal paraphrastic sentence embeddings. *arXiv preprint arXiv:1511.08198*.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*, pages 347–354. Association for Computational Linguistics.