Team Harry Friberg at SemEval-2019 Task 4: Identifying Hyperpartisan News through Editorially Defined Metatopics

Nazanin Afsarmanesh, Jussi Karlgren, Peter Sumbler, Nina Viereckel
Gavagai
Stockholm , Sweden
{nazanin, jussi, peter, nina}@gavagai.io

Abstract

This report describes the starting point for a simple rule based hypothesis testing excercise on identifying hyperpartisan news items carried out by the Harry Friberg team from Gavagai. We used manually crafted *metatopics*, topics which often appear in hyperpartisan texts as rant conduits, together with tonality analysis to identify general characteristics of hyperpartisan news items. While the precision of the resulting effort is less than stellar— our contribution ranked 37th of the 42 successfully submitted experiments with overly high recall (95%) and low precision (54%)—we believe we have a model which allows us to continue exploring the underlying features of what the subgenre of hyperpartisan news items is characterised by.

1 Hyperpartisanism

Hyperpartisan news are news items that are strongly argumentative and one-sided. However, being biased is not enough to be characterised as being hyperpartisan, neither is it enough for a news item to use strong language. Confounders for this task includes items that use strong language without being partisan, items that are subjective but not "hyper", and items that report dispassionately on typically hyperpartisan topics.

We hypothesise that authors of hyperpartisan texts are in the process of performing a sub-genre of their own, intended not as much to convey the reader information about some state of the world but to mobilise sentiment and affect in the readership, and establishing a shared attitudinal space. Taking this point of departure, we assume that the linguistic items employed by the authors of hyperpartisan text are not only related to the topics under discussions, nor to argumentation, but also include some genre-specific features to explicitly signal hyperpartisanness. This report describes an experiment based on these starting points, performed on data from the 2019 SemEval task on Hyperpartisan News Detection. (Kiesel et al., 2019)

2 Gavagai Explorer

The Gavagai Explorer is a commercially available tool which provides an end-to-end solution for the analysis of unstructured text data (Espinoza et al., 2018). We have in these experiments made use of its components for topic clustering, sentiment analysis, and concept modelling.

2.1 Trigger Topics

The *topic clustering* is based on lexical cues, and can be used to detect what themes and topics are prevalent in some set of e.g. customer feedback messages. Here, we used the topic clustering to establish what sort of themes were frequent in the hyperpartisan training set.

We postulate that many *metatopics* turn out to become lightning rods for hyperpartisan argumentation, somewhat (but not entirely) unpredictably. A characteristic of some of the more extreme sample items was that a hyperpartisan rant will bring in additional only marginally related topics into an argumentation. We identified a small set of potential rant metatopics using the topic clustering mechanism in the Gavagai Explorer. A breakdown of these topics with some example terms can be found in Table 1.

2.2 Trigger Attitudes

The *concept modeling* tool allows an analyst to define measures based on lexical items. *Sentiments* are a special case, applied to the palette of human emotion. On entering some seed words, the user is presented with semantically similar terms acquired from a distributional model (Sahlgren et al., 2016). The user accepts terms which are relevant,

Topic	Example terms
American	across america, america first, god bless america, ...
Elites	elite, establishment, oligarchy, ...
Freedom of Speech	first amendment rights, freedom of speech, press freedom, ...
Islam	islamist, islamism, muslim, ...
Media	cable news, mainstream media, twitter, ...
People	people, these people, the people, ...
Political Movements	alt-right, bolshevik, marxist
Politicians	party leaders, political party, politician, ...
Populism	populism, populist, ...
Public Safety	felon, incriminating, predator, ...
Race	black lives matter, black people, white people, ...
Sexual Rights	reproductive rights, same sex marriage, transgender, ...
Support	support, supportive, supported, ...
Woman	woman, women, ...

Table 1: Trigger topics and some of terms that indicate them. The full list with all terms is available at https://www.gavagai.io/blog/2019/06/06/gavagai-identifying-hyperpartisan-news/ .

Concept	Example terms
Certainty	blatantly, undeniably, hands-down, ...
Cynical	bizarre, far-fetched, ludicrous, ...
Exasperation	and once again, for some reason, yet again, ...
Failure	catastrophe, complete failure, disaster, ...
Nonsense	arrogant, babble, claptrap, ...
Puffery	landmark, pioneering, visionary, ...
Weasel Words	recent study, widely acknowledged, been claimed, ...
Win and Lose	bruised, humiliated, vanquished, ...

Table 2: Trigger attitudes and some of terms that indicate them. The full list with all terms is available at https://www.gavagai.io/blog/2019/06/06/gavagai-identifying-hyperpartisan-news/ .

which are in turn used to provide more suggestions in the following iteration. Here, we used the concept modelling tool to define *trigger attitudes* such as those shown in Table 2.

We find that strongly expressed attitudes not necessarily mean that an article is hyperpartisan, but that the combination of a trigger topic together with negative sentiment appears to be indicative of hyperpartisanism. The sentiment analysis component identifies several types of polar language, and measures the intensity of expression in each item using both presence of polar terms and of amplifier terms such as "extremely" and "very". In addition to standard polar sentiments we used the concept modeling tool to build a set of concepts tailored to observable presence in hyperpartisan texts (Karlgren et al., 2012).

2.3 Trigger Styles

Besides topical specificity we expect hyperpartisan texts to be couched in specific styles, as already established in previous studies (Potthast et al., 2018). We compared some stylistic features known to us to have discriminative power in other contexts, such as counts of exclamation marks, question marks, digits, capital letters, capitalised words, type token ratio, word length, sentence length etc. We found that the strongest single stylistic feature was the presence of many exclamation marks, in conjunction with trigger topics, while most other features on their own were less indicative. This is an indication that the authors of hyperpartisan texts appear to adhere to most stylistic conventions of the news genre.

3 Rule Based Fusion

We combined the above evidence in a rule based model, to achieve reasonably high explanatory power of results for downstream application. Through analysis of the training data, we distilled the results into the following pieces of reasoning, applied in the order given here:

1. Presence of many trigger topics (> 3) in an article, indicates it is hyperpartisan.

2. Presence of at least one trigger topic and a negative sentiment score for an article indicates it is hyperpartisan.

Figure 1: K W Gullers and Stieg Trenter, 1950s

3. A positive sentiment score combined with lack of trigger topics indicates a non-hyperpartisan article.

4. Presence of at least one trigger topic together with a high type token ratio *or* high ratio of questions in an article indicates it is hyperpartisan.

5. A high trigger attitude score (given in Table 2) indicates an article is hyperpartisan.

4 Results

The end results of our experiment on the by-article test set were decidedly underwhelming, with our contribution ranked 37th of 42 experiments. Our combined experimental pipeline yielded high recall (95%) and low precision (54%), meaning that it turned out to be overly sensitive to the features it was trained on. The rule set given above triggered for too many non-hyperpartisan items, with the last rule being the most permissive. We still believe that informed and hypothesis-driven analysis of content, rather than an end-to-end learning models, will result in a model of greater generality and greater explanatory power, but that the rule based combination should have been done using some learning scheme. While the precision of the resulting effort is less than stellar, we believe we have a model which allows us to continue exploring the underlying features of what the sub-genre

of hyperpartisan news items is characterised by, and we also believe that the explicit representation of what features are in play will afford end users greater trust in the system's classification results.

5 Namesake

Harry Friberg was a fictional photojournalist and the protagonist of a series of crime novels by Stieg Trenter (1914-1967). The character first appeared in the novel *Farlig fåfänga*, 1944, and continued in a series of novels which have since become popular classics for their depiction of Stockholm in the 1950s. Harry Friberg was modeled on the internationally recognised photojournalist K W Gullers (1916—1998), a friend of the author.

References

Fredrik Espinoza, Ola Hamfors, Jussi Karlgren, Fredrik Olsson, Per Persson, Lars Hamberg, and Magnus Sahlgren. 2018. Analysis of open answers to survey questions through interactive clustering and theme extraction. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*, pages 317–320. ACM.

Jussi Karlgren, Magnus Sahlgren, Fredrik Olsson, Fredrik Espinoza, and Ola Hamfors. 2012. Usefulness of sentiment analysis. In *European Conference on Information Retrieval*, pages 426–435. Springer.

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.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A Stylometric Inquiry into Hyperpartisan and Fake News. In *56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)*, pages 231–240. Association for Computational Linguistics.

Magnus Sahlgren, Amaru Cuba Gyllensten, Fredrik Espinoza, Ola Hamfors, Jussi Karlgren, Fredrik Olsson, Per Persson, Akshay Viswanathan, and Anders Holst. 2016. The Gavagai living lexicon. In *Language Resources and Evaluation Conference*. ELRA.