A Stylometric Inquiry into Hyperpartisan and Fake News

Martin Potthast   **Johannes Kiesel**   **Kevin Reinartz**   **Janek Bevendorff**   **Benno Stein**

Bauhaus-Universität Weimar
`<first name>.<last name>@uni-weimar.de`

Abstract

This paper reports on a writing style analysis of hyperpartisan (i.e., extremely one-sided) news in connection to fake news. It presents a large corpus of 1,627 articles that were manually fact-checked by professional journalists from BuzzFeed. The articles originated from 9 well-known political publishers, 3 each from the mainstream, the hyperpartisan left-wing, and the hyperpartisan right-wing. In sum, the corpus contains 299 fake news, 97% of which originated from hyperpartisan publishers.

We propose and demonstrate a new way of assessing style similarity between text categories via Unmasking—a meta-learning approach originally devised for authorship verification—, revealing that the style of left-wing and right-wing news have a lot more in common than any of the two have with the mainstream. Furthermore, we show that hyperpartisan news can be discriminated well by its style from the mainstream ($F_1 = 0.78$), as can be satire from both ($F_1 = 0.81$). Unsurprisingly, style-based fake news detection does not live up to scratch ($F_1 = 0.46$). Nevertheless, the former results are important to implement pre-screening for fake news detectors.

1 Introduction

The media and the public are currently discussing a new phenomenon called "fake news" and its potential role in swaying recent elections, how it may affect democratic societies, and what can and should be done about it. In a nutshell, "fake news" encompasses the observation that, in social media, a certain kind of 'news' spread much more successfully than others, and that these 'news' are typically extremely one-sided (hyperpartisan), inflammatory, emotional, and often riddled with untruths. Although traditional yellow press has been spreading 'news' of varying degrees of truthfulness long before the digital revolution, the fact that modern social media amplify fake news to outperform *real* news gives many people pause. The fake news

hype caused a widespread disillusionment about social media, and many politicians, news publishers, IT companies, activists, and scientists concur that this is where to draw the line. For all their good intentions, however, it is already obvious that it must be drawn very carefully (if at all), since nothing less than free speech is at stake—a fundamental right of every free society.

Many favor a two-step approach where first fake news items are detected and then countermeasures are implemented to foreclose false rumors and to discourage repeated offenses. The countermeasures aiming at foreclosing range from displaying warnings, the withholding of news items until stakeholders can react, up to their complete removal. Countermeasures aiming at discouraging range from withholding display advertising revenue, the flagging or downranking of a sender's account, up to banning a sender altogether. By comparison, traditional countermeasures that used to work fairly well in offline media are the publication of refutations, either in a different venue or in the same venue as a fake news item by invoking the so-called *right to reply*. In social media, however, the traditional countermeasures are rather ineffective since refutations are typically shared much less than the refuted fake news beforehand, and few jurisdictions have an enforceable right to reply. Once a fake news item spreads virally, the damage is done, and containing it becomes almost impossible: an immediate reaction is crucial.

This paper focuses on fake news detection, but with a twist. While this task is most commonly tackled by automatic fact-checking, we approach it from a different angle by investigating the writing style of fake news in relation to hyperpartisan news. In this regard, we analyze for the first time whether hyperpartisan news can be distinguished by its style from mainstream news (it can), whether satire can be distinguished from both (it can, too), and whether fake news can be detected via style alone (it can't). Furthermore, we introduce a new approach to assess and visualize writing style similarities of text categories based on Unmasking. This way, we assess the style differences between hyperpartisan left-wing news and hyperpartisan right-wing news, showing that the two have sig-

nificant stylistic similarities. All of these experiments are based on a new, publicly shared dataset, comprising annotations whether news are fake or real, and whether they are hyperpartisan, sampled from 9 well-known publishers and annotated by journalists from BuzzFeed.

After a brief review of related work, Section 3 details the dataset and how it was constructed, Section 4 introduces our methodology including our variant of Unmasking, and Section 5 reports the results of the aforementioned experiments.

2 Related Work

Figure 1 organizes the literature on fake news detection in terms of three paradigms: fake news detection based on knowledge, on context, and on style. For each of the paradigms we list specific research areas which supply different methods for solving the task. Knowledge-based fake news detection (also called "fact checking"), is tackled with methods borrowed from information retrieval, semantic web, and linked open data (LOD) research. Context-based fake news detection employs methods from social network analysis where the spread of false information and rumors as well as their containment is studied. Style-based fake news detection relies on computational linguistics and natural language processing, and, more specifically, on methods from deception detection to identify statements at the sentence-level that constitute falsehoods and lies. One field of research that has hardly been considered in the context of fake news detection, yet, is style-based text categorization. In this paper, we will close this gap and analyze the potential of classifying news items by the style of their text body into classes corresponding to fake, real, and satire, as well as to hyperpartisan and mainstream news.

Knowledge-based Fake News Detection. Methods from information retrieval have been proposed early on to determine the truthfulness of a given web resource. For example, Etzioni et al. (2008) propose to use their well-known tool Text Runner (Yates et al., 2007) to extract and index factual knowledge from the web, and to use the same technology to extract factual statements from a given text in question, matching them against the indexed facts to identify inconsistencies. Magdy and Wanas (2010) develop a statistical model to check factual statements extracted from a given document in question, analyzing how frequently they are supported by documents retrieved from the web. Both approaches presume that web resources (or the frequency by which a fact is mentioned) can be used as an indication of its truth. However, the problem
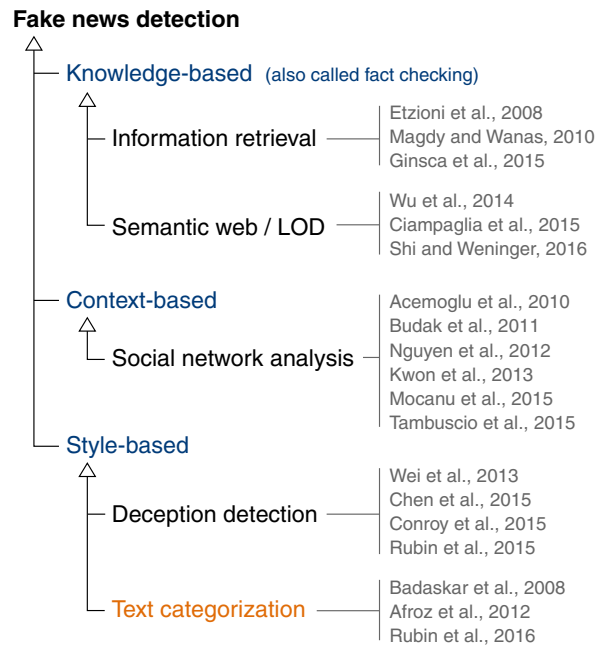
Fake news detection

Figure 1: Taxonomy of paradigms for fake news detection alongside a selection of relevant work.

of this argument is that the reputation and reliability of almost any website can be put into question. In this regard, Ginsca et al. (2015) give a comprehensive overview of challenges and approaches from the literature to assess the various aspects of credibility in information retrieval, namely: expertise, trustworthiness, quality, and reliability.

While the extraction and retrieval of factual information plays an important role when using the web as a knowledge base, knowledge becomes increasingly shared as structured knowledge bases, integrated in the linked open data cloud, inducing the semantic web. In fact, information extracted from the web, too, is typically stored in such knowledge bases. When presuming that factual knowledge is available in sufficient detail for a domain of interest, the task of fake news detection boils down to checking whether a given fact is already known, or whether it can be inferred from other facts. Wu et al. (2014) try to assess the truthfulness of a given fact by "perturbing" it, formulating queries to knowledge bases and interpreting the variation in the results as a sign of whether the fact is either strongly or weakly supported. Ciampaglia et al. (2015) cast fact-checking as a problem of finding shortest paths between concepts in a knowledge graph; they propose a metric to assess the truth of a statement by analyzing path lengths between the concepts in question. Conversely, Shi and Weninger (2016) cast fake news detection as a link prediction task, where a probability is estimated in order to decide whether concepts covered by a to-be-checked statement should be linked.

Context-based Fake News Detection. By considering the mechanisms of social networks, new angles on the problem of fake news dissemination come into reach. Acemoglu et al. (2010) model how (mis-)information is spread in social networks, and Budak et al. (2011) and Nguyen et al. (2012) propose algorithms to limit their spread. Kwon et al. (2013) combine social network analysis and linguistic feature obtained from applying LIWC (Pennebaker et al., 2003) to identify rumors as they spread. Studying the spread of misinformation on Facebook during an election, Mocanu et al. (2015) provide evidence that unsubstantiated claims spread as widely as well-established ones, and that user groups with a predisposition to conspiracy theories are more open to sharing misinformation. Tambuscio et al. (2015) also study the spread of misinformation in social media; however, they also study the efficacy of countermeasures such as debunking sites. In particular, they find that by exceeding a certain threshold in spreading the refutation is sufficient to remove the misinformation from the network, and that this threshold does not depend on the spreading rate but on credulity and forgetfulness.

Style-based Fake News Detection. Another approach to fake news detection is to sidestep fact checking and social network analysis altogether by modeling the *nature of faking* and its manifestation in text. Two branches of research provide the rationale and methodology: deception detection in text and style-based text categorization. Deception detection originates from forensic linguistics and builds on the Undeutsch hypothesis—a result from forensic psychology asserting that memories of real-life, self-experienced events differ in content and quality from imagined events (Undeutsch, 1967). The hypothesis leads to the development of (the now commonly applied) forensic tools to assess testimony at the statement level, such as Criteria-based Content Analysis (CBCA) and Scientific Content Analysis (SCAN). Nowadays, technology is being developed to operationalize deception detection at scale, e.g. in order to identify or to detect uncertainty in social media posts: Wei et al. (2013) propose a model to detect tweets that convey uncertain information. Regarding fake news detection, Chen et al. (2015) point out the need for an "automatic crap detector" for news, but do not report on actual experiments, whereas Rubin et al. (2015) apply, for the first time, deception detection approaches to fake news detection using rhetorical structure theory as a measure of story coherence.

Style-based text categorization was proposed by Argamon-Engelson et al. (1998) as an alternative to topic-based text categorization in order to tackle tasks ranging from author profiling (by age, gender, native language, etc.) to broader style categories such as text genre. Since it is hypothesized that deception has its own style as well, there is an overlap in methodology, using full text classification as another means to assess the truthfulness of a given text as opposed to analyzing individual statements. For example, Afroz et al. (2012) attempt to detect texts whose authors tried to obfuscate their writing style to deflect author identification. As an early precursor to fake news detection, Badaskar et al. (2008) train models to tell real news apart from news that have been automatically generated using a language model, However, Rubin et al. (2016) contributed the first actual attempt at fake news detection by separating satire news as a representative of humorous fakes from real news in a dataset of 180 news articles each, achieving F-Measure values between 0.82 and 0.87 for various variants of a $tf \cdot idf$-weighted lexical vector space model. We employ this dataset in conjunction with our own in our experiments to study the connection of fake news, real news, and satire for the first time.

3 The BuzzFeed-Webis Fake News Corpus 2016

This section introduces the BuzzFeed-Webis Fake News Corpus 2016, detailing its construction and annotation by professional journalists employed at BuzzFeed, as well as key figures and statistics.

3.1 Corpus Construction

The corpus comprises a complete sample of the output of 9 publishers in a week close to the US elections. Among the selected publishers are 6 prolific hyperpartisan publishers (three left-wing and three right-wing) and three mainstream publishers (see Table 1). All publishers earned Facebook's blue checkmark ✓, indicating authenticity and an elevated status within the network. For seven weekdays (September 19 to 23 and September 26 and 27), every post and linked news article of the 9 publishers was fact-checked claim-by-claim by 5 BuzzFeed journalists, including about 10% of posts forwarded from third parties. Silverman et al. (2016) reported key insights as a data journalism article, having checked a total of 2,282 posts, 1,145 of which from mainstream publishers, 471 from hyperpartisan left-wing publishers, and 666 from hyperpartisan right-wing publishers. Alongside the article, the annotations were published as well.[1] However, this data only comprises URLs to the original Facebook posts. To create the corpus, we hence archived the posts, the linked articles, and

---
[1] http://github.com/BuzzFeedNews/2016-10-facebook-fact-check

Table 1: The BuzzFeed-Webis Fake News Corpus 2016 at a glance. ("Paras." short for "paragraphs")

Orientation Publisher	Fact-checking results					Key statistics per article				
	true	mix	false	n/a	Σ	Paras.	Links		Words	
							extern	all	quoted	all
Mainstream	806	8	0	12	826	20.1	2.2	3.7	18.1	692.0
ABC News	90	2	0	3	95	21.1	1.0	4.8	21.0	551.9
CNN	295	4	0	8	307	19.3	2.4	2.5	15.3	588.3
Politico	421	2	0	1	424	20.5	2.3	4.3	19.9	798.5
Left-wing	182	51	15	8	256	14.6	4.5	4.9	28.6	423.2
Addicting Info	95	25	8	7	135	15.9	4.4	4.5	30.5	430.5
Occupy Democrats	59	25	7	0	91	10.9	4.1	4.7	29.0	421.7
The Other 98%	28	1	0	1	30	20.2	6.4	7.2	21.2	394.5
Right-wing	276	153	72	44	545	14.1	2.5	3.1	24.6	397.4
Eagle Rising	106	47	25	36	214	12.9	2.6	2.8	17.3	388.3
Freedom Daily	49	24	22	4	99	14.6	2.2	2.3	23.5	419.3
Right Wing News	121	82	25	4	232	15.0	2.5	3.6	33.6	396.6
Σ	1264	212	87	64	1627	17.2	2.7	3.7	20.6	551.0

attached media as well as relevant meta data to ensure long-term availability. Due to the rapid pace at which the publishers change their websites, we were able to recover only 1,627 articles, 826 mainstream, 256 left-wing, and 545 right-wing. Table 1 gives an overview.

Manual Fact-checking. Five BuzzFeed journalists conducted the manual fact-checks of the news articles. To avoid bias with regard to publishers, news from all publishers and all days were assigned round robin. It became clear very quickly that a binary distinction between fake and real news was infeasible, since hardly any piece of fake news is entirely false, and hardly any piece of real news is flawless. Therefore, posts were rated "mostly true," "mixture of true and false," "mostly false," or, if the post was opinion-driven or otherwise lacked a factual claim, "no factual content." The ratings "mixture of true and false" and "mostly false" had to be justified, and, when in doubt about a rating, a second opinion was collected, whereas disagreements were resolved by a third one. Finally, all news rated "mostly false" underwent a final check by a different rater, to ensure the rating was justified. Raters were given the following guidance:

Mostly true: the post and any related link or image are based on factual information and portray it accurately. This lets the authors interpret the event/info in their own way, so long as they do not misrepresent events, numbers, quotes, reactions, etc., or make information up. This rating does not allow for unsupported speculation or claims.

Mixture of True and False (mix, for short): Some elements of the information are factually accurate, but some elements or claims are not. This rating should be used when speculation or unfounded claims are mixed with real events, numbers, quotes, etc., or when the headline of the link being shared makes a false claim but the text of the story is largely accurate. It should also only be used when

the unsupported or false information is roughly equal to the accurate information in the post or link. Finally, use this rating for news articles that are based on unconfirmed information.

Mostly False: Most or all of the information in the post or in the link being shared is inaccurate. This should also be used when the central claim being made is false.

No Factual Content (n/a, for short): This rating is used for posts that are pure opinion, comics, satire, or any other posts that do not make a factual claim. This is also the category to use for posts that are of the "Like this if you think..." variety.

3.2 Corpus Statistics

Table 1 shows the fact-checking results and some key statistics per article. Unsurprisingly, none of the mainstream articles are mostly false, whereas 8 across all three publishers are a mixture of true and false. Disregarding non-factual articles, a little more than a quarter of all hyperpartisan left-wing articles were found faulty: 15 articles mostly false, and 51 a mixture of true and false. Publisher "The Other 98%" sticks out by achieving an almost perfect score. By contrast, almost 45% of the right-wing articles are a mixture of true and false (153) or mostly false (72). Here, publisher "Right Wing News" sticks out by supplying more than half of mixtures of true and false alone, whereas mostly false articles are equally distributed.

Regarding key statistics per article, it is interesting to note that the articles from all mainstream publishers are on average about 20 paragraphs long with word counts ranging from 550 words on average at ABC News to 800 at Politico. Except for one publisher, left-wing articles and right-wing articles are shorter on average in terms of paragraphs as well as word count, averaging at about 420 words and 400 words, respectively. Left-wing articles quote on average about 10 words more than the mainstream, and right-wing articles 6 words. When articles comprise links, they are usually external ones, whereas ABC News rather uses internal links, and only half of the links found at Politico articles are external. Left-wing news articles stick out by containing almost double the amount of links across publishers than mainstream and right-wing news.

3.3 Operationalizing Fake News

In our experiments, we operationalize the category of fake news by joining the articles that were rated mostly false with those rated a mixture of false and true. Arguably, the latter may not be exactly what is colloquially understood under the term "fake news" (as in: a complete fabrication), however,

practice shows fake news are hardly ever devoid of truth. More often, true facts are misconstrued using argumentative fallacies to influence a person's opinion. In our experiments, we hence call mostly true articles real news, mostly false plus mixtures of true and false, except for satire, fake news, and disregard all articles rated non-factual.

4 Methodology

In this section, we briefly review methodology, including a brief recap of Unmasking by Koppel et al. (2007), for which we investigate for the first time its use in distinguishing genre styles as opposed to authors, and our set of features used to capture writing style. For sake of reproducibility, all our code will be made publicly available.

4.1 Unmasking Style Categories

Unmasking, as proposed by Koppel et al. (2007), is a meta learning approach that was originally intended for authorship verification. In this paper, we study for the first time whether it can be used to assess the similarity of more broadly defined style categories compared to authorial style, such as left-wing versus right-wing versus mainstream news. This way, we attempt uncover relations between the writing styles that people may involuntarily adopt as per their political orientation.

Originally, Unmasking takes two documents as input and outputs its confidence whether they have been written by the same author. Three steps are taken to accomplish this: first, each document is chunked into a set of at least 500-word long chunks; second, reconstruction errors are measured while iteratively removing the most discriminative features of a style model comprising the 250 most frequent words used to separate the two chunk sets with a linear classifier; and third, the resulting reconstruction error curves are analyzed with regard to their slope. A steep decrease is more likely than a shallow decrease if the two documents have been written by the same author, since there are presumably less discriminating features between documents written by the same author than between documents written by different authors. Training a classifier on many examples of error curves obtained from same-author document pairs and different-author document pairs yields an effective authorship verifier—at least for long documents that can be split up into a sufficient number of chunks.

We believe that what applies to the style of authors also applies to more broadly defined styles; in our case hyperpartisanship. We adapt Unmasking by skipping its first step and using two sets of documents (e.g., left-wing articles and right-wing articles) as input. Further, we plot the reconstruction error curves for visual inspection: steeper decreases in these plots indicate style similarity of the two input document sets, just as they do when using chunks obtained from documents written by the same author. This way, we demonstrate that Unmasking can be applied in situations where the question arises whether man-made categories of texts are stylistically discriminative, or not. When applied to document sets sampled from hyperpartisan left, hyperpartisan right, and mainstream publishers, we expect insights into the nature of the writing style unconsciously adopted by people of these orientations.

4.2 Style Features and Feature Selection

Our writing style model incorporates commonly used style features as well as some specific to the news domain. The former are n-grams of characters, stop words (in order of appearance in the text), and parts-of-speech with n in $[1, 3]$. Furthermore, we employ 10 readability scores[2] as well as dictionary features, where each one indicates the frequency of words from a tailor-made dictionary in a given document, using the General Inquirer Dictionaries as a basis (Stone et al., 1966). The domain-specific features include ratios of quoted words and external links, and the number of paragraphs and their average length in a document.

In each of our experiments, we carefully select from the aforementioned features the ones worthwhile using. To avoid overfitting, all features that are hardly represented in the documents of our corpus (i.e., occur in less than 10% of documents) are discarded. This pertains particularly to many individual n-grams as well as some dictionary features, since most low-frequency n-grams found are unique to a certain publisher. We further disregard features that are not represented in at least two of the categories to be distinguished in a given experiment. This way, the performance of our classifiers may not be as high as it could be when using all features, but our results can be used to draw meaningful conclusions.

4.3 Baselines

We employ the standard bag of word model for a topic-based classification baseline. This approach is less practical, however, since topics change frequently and drastically in the news domain. Moreover, we supply naive baselines that classify all items into one of the classes in question, thus relating results to the class distributions.

---

[2] Automated Readability Index, Coleman Liau Index, Flesh Kincaid Grade Level and Reading Ease, Gunning Fog Index, LIX, McAlpine EFLAW Score, RIX, SMOG Grade, Strain Index

5 Experiments

We report on the results of two series of experiments aimed at investigating the style differences and similarities between hyperpartisanship and the mainstream, as well as between fake, real, and satire news. In particular, we shed light on the following three questions: (1) *Is there a common style of hyperpartisanship?* Our working hypothesis is that left-wing and right-wing hyperpartisans have more in common than they themselves would admit, and our results indeed provide first evidence in the affirmative. (2) *Is style-based fake news detection feasible?* Of course, we do not expect to solve fake news detection using style alone, but exploiting style has the advantage that it can be applied for pre-screening in real time, and that authors presumably have little control over their own writing style at large. (3) *Can hyperpartisan fake news be distinguished from satire?* Investigating the special case of satire news is important, as we cannot allow for humor to be sacrificed on the altar of truth. Therefore, any fake news detection approach must be able to deal with satire.

5.1 Hyperpartisanship vs. Mainstream

This series of experiments targets the political orientation, and thus research question (1). We conduct three experiments, where the first two distinguish left-wing, right-wing, and mainstream, and the third one hyperpartisan and mainstream.

A. Unmasking hyperpartisanship. We apply Unmasking as described above onto pairs of the three orientations in question. Figure 2 shows the resulting Unmasking curves (Unmasking is symmetrical, hence three curves). The curves are averaged over 5 runs, where each run comprised sets of 100 documents from each orientation in question. In case of the left-wing orientation, where less than 500 documents are available in our corpus, once all of them had been used, they were shuffled again to select documents for the remainder of the runs.

As all three curves decrease quickly, Koppel et al.'s original hypothesis in the context of authorship verification would be that all articles have been written by the same author. When transferred to our application of Unmasking, one would generally conclude that all pairs of orientations have a common style, which is unsurprising given that the documents have all been sampled from news publishers. They all possess the style of the genre of news articles. However, Unmasking allows for a more fine-grained assessment of style similarity. In this regard, we formulate the Unmasking hypothesis as follows: style similarity is characterized by the slope of a given Unmasking curve,
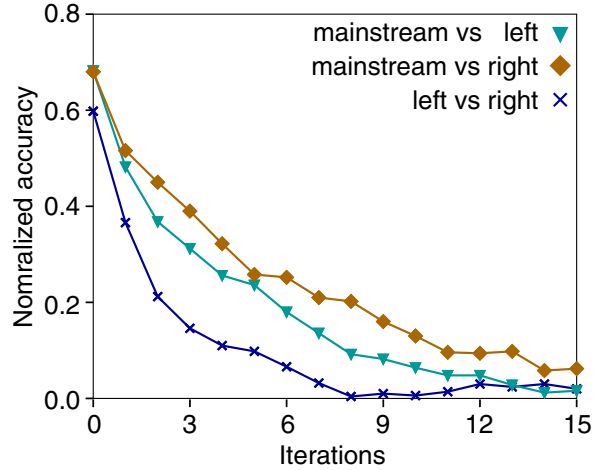

Figure 2: Unmasking applied to pairs of political orientations. The quicker a curve decreases, the more similar the respective styles are.

Table 2: Accuracies of predicting hyperpartisanship vs mainstream under different training sets.

Features	Training set								
	without left			without right			with both		
Test set:	left	right	main.	left	right	main.	left	right	main.
Style	**0.74**	0.84	0.66	0.68	**0.66**	0.74	0.90	0.89	0.62
Topic	0.68	0.77	0.78	0.50	0.48	0.87	0.79	0.85	0.60

where a steeper decrease indicates higher similarity. Marked differences between the curves originating from different pairs of orientations can be observed. If the Unmasking hypothesis is correct, this experiment shows that the left-wing and right-wing documents found in our corpus have a lot more in common, stylistically, than documents from either orientation with mainstream documents.

B. Predicting Hyperpartisanship. Given this initial result, we follow that it should be possible to discriminate hyperpartisan news from the mainstream, even if one has only examples from the respective other hyperpartisan side plus mainstream examples. To test this hypothesis, we train three binary classifiers (random forests) to distinguish hyperpartisan news from mainstream news, while omitting left-wing documents from the training set, omitting right-wing articles from the training set, and retaining both, respectively. The training sets for all three classifiers were balanced between the classes mainstream and hyperpartisan; where left-wing or right-wing documents were omitted, oversampling was applied on the remaining documents from the respective other side.

Table 2 shows performance values as averages of 3-fold cross-validation, where each fold comprises one publisher from each orientation so that the classifier can not learn the publisher style. Besides our style feature model, we also provide performance values of corresponding classifiers trained under

Table 3: Accuracies, class-wise precision, recall, and F-Measure of predicting hyperpartisanship. Including baselines predicting all articles as hyperpartisan or mainstream, respectively

Features	Accuracy	Precision		Recall		F_1	
	all	hyp.	main.	hyp.	main.	hyp.	main.
Style	0.75	0.69	0.86	0.89	0.62	0.78	0.72
Topic	0.71	0.66	0.79	0.83	0.60	0.74	0.68
All-hyp.	0.49	0.49	-	1.00	0.0	0.66	-
All-main.	0.51	-	0.51	0.0	1.00	-	0.68

a topic feature model (bag of words) as baseline for comparison. When omitting left-wing documents from the training set, the style-based classifier still achieves 0.74 accuracy on left-wing test documents. Likewise, it achieves 0.66 accuracy classifying right-wing documents without having trained on them. In both cases, the style-based classifier outperforms the baseline topic feature model. This behavior supports our observation that the style of left-wing documents is not so different from that of right-wing ones. Unsurprisingly, not omitting any documents from the training data improves accuracy significantly, however, only for the class of hyperpartisan documents. The mainstream class can be best discriminated when omitting right-wing documents, at the expense of accuracy for the class of hyperpartisan documents.

Table 3 shows performance values of the binary classifier, trained with both hyperpartisan sides present in its training set to discriminate hyperpartisan and mainstream news. The best classification accuracy of 0.75 at a remarkable 0.89 recall for the hyperpartisan class is achieved by the style-based classifier, outperforming the topic feature model.

C. Predicting Orientation. As a last experiment in this series, we trained a classifier to predict the orientation of an individual news article. Table 4 shows the performance values of a random forest in terms of accuracy, precision, and recall, again, as averages of 3-fold cross-validation, compared to the same baselines. Interestingly, in the three-class setting, the topic baseline outperforms the style-based model with regard to accuracy, whereas the results for class-wise precision and recall are a mixed bag. The left-wing documents are apparently significantly more difficult to be identified compared to documents from the other two orientations. When looking at the confusion matrix, it turns out that 66% of misclassification of left-wing documents are falsely classified as right-wing documents. The reverse, however, is not true: 60% of all misclassified right-wing documents are classified as mainstream documents, whereas misclassified mainstream documents spread almost equally across the other classes.

Table 4: Accuracies and class-wise precision, recall and F-Measure of predicting orientation.

Features	Accuracy	Precision			Recall			F_1		
	all	left	right	main.	left	right	main.	left	right	main.
Style	0.60	0.21	0.56	0.75	0.20	0.59	0.74	0.20	0.57	0.75
Topic	0.64	0.24	0.62	0.72	0.15	0.54	0.86	0.19	0.58	0.79
All-left	0.16	0.16	-	-	1.00	0.0	0.0	0.27	-	-
All-right	0.33	-	0.33	-	0.0	1.00	0.0	-	0.50	-
All-main.	0.51	-	-	0.51	0.0	0.0	1.00	-	-	0.68

Table 5: Accuracies, class-wise precision, recall and F-Measure of predicting fake news by features and classifier.

Features	Accuracy	Precision		Recall		F_1	
	all	fake	real	fake	real	fake	real
Generic classifier							
Style	0.55	0.42	0.62	0.41	0.64	0.41	0.63
Topic	0.52	0.41	0.62	0.48	0.55	0.44	0.58
Orientation-specific classifier							
Style	0.55	0.43	0.64	0.49	0.59	0.46	0.61
Topic	0.58	0.46	0.65	0.45	0.66	0.46	0.66
All-fake	0.39	0.39	-	1.00	0.0	0.56	-
All-real	0.61	-	0.61	0.0	1.00	-	0.76

5.2 Fake vs. Real (vs. Satire)

This series of experiments targets research questions (2) and (3). Again, we conduct three experiments, where the first is about predicting veracity, and the last two about discriminating satire.

A. Predicting veracity. Given the encouraging results in predicting hyperpartisan news—in particular the high recall of 0.89 at a reasonable precision of 0.69—we are confident that, with some further effort, a practical classifier can be built that detects hyperpartisan news at scale. When taking into account that the mainstream news publishers in our corpus did not publish any news that are mostly false, and only very few instances that are mixtures of true and false, we can safely disregard them for the task of fake news detection. In this connection, a classifier that reliably distinguishes hyperpartisan news from the mainstream can act as a pre-filter for a subsequent, more in-depth fake news detection approach, which may in turn be tailored to a much more narrowly defined classification task. We hence use only the left-wing documents and the right-wing documents of our corpus as a basis for training our fake news classifier.

Table 5 shows the performance values for three classifiers, a generic classifier that predicts fake news across sides, and orientation-specific classifiers that have been individually trained on documents from either side, left-wing and right-wing news. In both cases, values are again averages of 3-fold cross-validation. Although all classifiers outperform the naive baselines of classifying everything as one class in terms of precision, the slight increase comes at the cost of a large decrease in recall. While the orientation-specific classifiers are
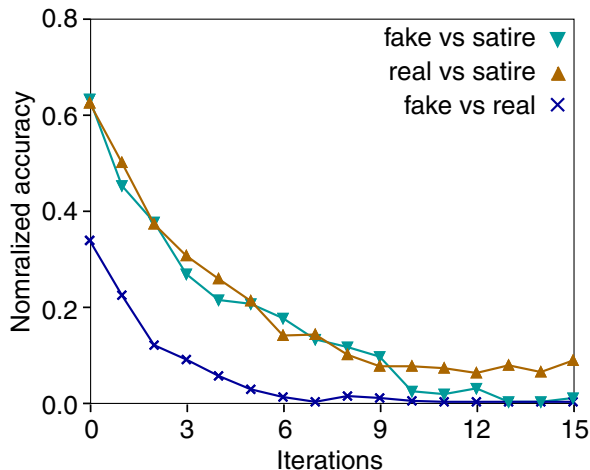
Figure 3: Unmasking applied to pairs of sets of news that are fake, real, and satire.

slightly better for most metrics, none of them can outperform the naive baselines with regard to the F-Measure.

B. Unmasking satire. As mentioned above, a special kind of harmless fake news is satire, which takes the form of news but lies more or less obviously to amuse its readers. Regardless the problems that fake news spread may cause, satire and other forms of humor should never be filtered. We assess the style similarity of satire compared to fake news and real news for the firs time, again applying Unmasking to compare pairs of the three categories of news as described above. For satire news, we use the 180 articles from the S-n-L News DB (Rubin et al. (2016), Section 2). Figure 3 shows the resulting Unmasking curves. The slope for the pair of fake vs real news drops somewhat slower compared to the other two pairs. Apparently, the style of fake news has much more in common with that of real news than either of the two has with satire (cf. Section 5.1 A). These results are encouraging: satire is apparently not very similar to either fake news or real news, so that, just like with hyperpartisan news compared to mainstream news, it should be fairly easy to discriminate between the two. This is indeed the case, as our last experiment confirms.

C. Discriminating satire. Finally, we evaluate the use of style features for predicting satire. Table 6 shows the performance values of our random forest classifier in the satire-detection setting used by Rubin et al. (2016). This setting uses a balanced 3:1 training-to-test set split over 360 articles (180 per class). As can be seen, our style-based model significantly outperforms all baselines across the board, achieving an accuracy of 0.82, and an F score of 0.81. While it does not outperform the classifier based on topic, absurdity, grammar, and punctuation features by Rubin et al.

Table 6: Accuracies, class-wise precision, recall, and F-Measure of predicting satire. Also shows the results for the best classifier of Rubin et al. (2016).

Features	Accuracy	Precision		Recall		F_1	
	all	sat.	real	sat.	real	sat.	real
Style	0.82	0.84	0.80	0.78	0.85	0.81	0.82
Topic	0.77	0.78	0.75	0.74	0.79	0.76	0.77
All-sat.	0.50	0.50	-	1.00	0.0	0.67	-
All-real	0.50	-	0.50	0.00	1.00	-	0.67
Rubin et al.		0.90		0.84		0.87	

(2016), it clearly improves over a pure topic classification. We argue that incorporating topic into satire detection is not appropriate, since the topics of satire change as the topics of news in general do. A classifier with topic features would therefore not generalize. Apparently, a style-based model is competitive, and we believe that satire can be detected at scale this way, so as to prevent other fake news detection technology from falsely filtering it.

6 Conclusion

Fake news detection poses an interdisciplinary challenge: technology needs to be in place to extract factual statements from text, to match facts with a knowledge base, to dynamically retrieve and maintain knowledge bases from the web, to reliably assess the overall truthfulness of an article rather than individual statements, to do so in real time as news events unfold, to monitor the spread of fake news within and across social media, to measure the reputation of information sources, and to raise awareness in readers. These are perhaps only the most obvious things that can be done to tackle the problem, and as our cross-section of related work shows, many of them are tackled already. Given the recent hype around fake news, many more contributions will surface in the future—hopefully including our own.

Contemplating the task in its whole breadth, we sought ways to support these developments that are specific and can be reasonably addressed within the scope of a paper. Since many already attack fake news head on, by developing one way or another of fact-checking, we thought it worthwhile to mount our attack from another angle: writing style. Although difficult to be captured in a computer representation, it is just as difficult to be manipulated, let alone in a consistent way. While any knowledge base of sufficient size is prone to manipulation, authors of fake news may unwittingly leave traces of their hatred, political predisposition, or even disinterest, in case earning money is their only driving force, right within their fabrications.

In this paper, we have tried and succeeded to uncover at least some of these traces. News articles conveying a hyperpartisan world view can fairly

easily be distinguished from more balanced news. Moreover, we have shown that the writing styles of otherwise opposing orientations, namely left-wing and right-wing, are in fact very similar: there appears to be a common style of extremism. We further show that satire can be distinguished well from other news, making at least sure that humor will not be outcast by fake news detection technology. Alas, we cannot claim to have solved fake news detection via style analysis alone.

Acknowledgements

References

Daron Acemoglu, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. Spread of (Mis)Information in Social Networks. *Games and Economic Behavior*, 70(2):194–227.

Sadia Afroz, Michael Brennan, and Rachel Greenstadt. 2012. Detecting Hoaxes, Frauds, and Deception in Writing Style Online. In *2012 IEEE Symposium on Security and Privacy*, pages 461–475, May.

Shlomo Argamon-Engelson, Moshe Koppel, and Galit Avneri. 1998. Style-based text categorization: What newspaper am i reading. In *Proc. of the AAAI Workshop on Text Categorization*, pages 1–4.

Sameer Badaskar, Sachin Agarwal, and Shilpa Arora. 2008. Identifying real or fake articles: Towards better language modeling. In *Third International Joint Conference on Natural Language Processing, IJCNLP 2008, Hyderabad, India, January 7-12, 2008*, pages 817–822. The Association for Computer Linguistics.

Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 665–674, New York, NY, USA. ACM.

Yimin Chen, Niall J. Conroy, and Victoria L. Rubin. 2015. News in an Online World: The Need for an "Automatic Crap Detector". In *Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community*, ASIST '15, pages 81:1–81:4, Silver Springs, MD, USA. American Society for Information Science.

Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis M Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational Fact Checking from Knowledge Networks. *PloS one*, 10(6):e0128193.

Oren Etzioni, Michele Banko, Stephen Soderland, and Daniel S. Weld. 2008. Open Information Extraction from the Web. *Commun. ACM*, 51(12):68–74, December.

Alexandru L. Ginsca, Adrian Popescu, and Mihai Lupu. 2015. Credibility in Information Retrieval. *Found. Trends Inf. Retr.*, 9(5):355–475, December.

Moshe Koppel, Jonathan Schler, and Elisheva Bonchek-Dokow. 2007. Measuring differentiability: Unmasking pseudonymous authors. *J. Mach. Learn. Res.*, 8:1261–1276.

Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent Features of Rumor Propagation in Online Social Media. In *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, pages 1103–1108. IEEE.

Amr Magdy and Nayer Wanas. 2010. Web-based Statistical Fact Checking of Textual Documents. In *Proceedings of the 2Nd International Workshop on Search and Mining User-generated Contents*, SMUC '10, pages 103–110, New York, NY, USA. ACM.

Delia Mocanu, Luca Rossi, Qian Zhang, Marton Karsai, and Walter Quattrociocchi. 2015. Collective Attention in the Age of (Mis)Information. *Comput. Hum. Behav.*, 51(PB):1198–1204, October.

Nam P. Nguyen, Guanhua Yan, My T. Thai, and Stephan Eidenbenz. 2012. Containment of Misinformation Spread in Online Social Networks. In *Proceedings of the 4th Annual ACM Web Science Conference*, WebSci '12, pages 213–222, New York, NY, USA. ACM.

James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology*, 54:547–577.

Victoria Rubin, Niall Conroy, and Yimin Chen. 2015. Towards News Verification: Deception Detection Methods for News Discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium*, Kauai, Hawaii, USA, January.

Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. In *Proceedings of the Second Workshop on Computational Approaches to Deception Detection*, pages 7–17, San Diego, California, June. Association for Computational Linguistics.

Baoxu Shi and Tim Weninger. 2016. Fact Checking in Heterogeneous Information Networks. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 101–102, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Craig Silverman, Lauren Strapagiel, Hamza Shaban, Ellie Hall, and Jeremy Singer-Vine. 2016. Hyperpartisan Facebook Pages are Publishing False and Misleading Information at an Alarming Rate. https://www.buzzfeed.com/craigsilverman/partisan-fb-pages-analysis. BuzzFeed.

Philip J. Stone, Dexter C. Dunphy, and Marshall S. Smith. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT press.

Marcella Tambuscio, Giancarlo Ruffo, Alessandro Flammini, and Filippo Menczer. 2015. Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 977–982, New York, NY, USA. ACM.

Udo Undeutsch. 1967. Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der Psychologie*, 11:26–181.

Zhongyu Wei, Junwen Chen, Wei Gao, Binyang Li, Lanjun Zhou, Yulan He, and Kam-Fai Wong. 2013. An empirical study on uncertainty identification in social media context. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 58–62, Sofia, Bulgaria, August. Association for Computational Linguistics.

You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2014. Toward Computational Fact-checking. *Proc. VLDB Endow.*, 7(7):589–600, March.

Alexander Yates, Michele Banko, Matthew Broadhead, Michael Cafarella, Oren Etzioni, and Stephen Soderland. 2007. TextRunner: Open information extraction on the web. In *Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 25–26, Rochester, New York, USA, April. Association for Computational Linguistics.