A Stylometric Inquiry into Hyperpartisan and Fake News

(ACL 2018)

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

Chiyu Zhang

Oct. 31st, 2018

Fake New Detection

- Approaches of fake news detection divide into three categories.

- Information retrieval: veracity of web documents
- Semantic web and lined open data: knowledge graph
- Social network analysis: meta information, spread patterns, author information and conversation
- Deception detection: rhetorical structure theory, single statements
- Text categorization: author profiling and genre classification, satire detection, assess entire texts

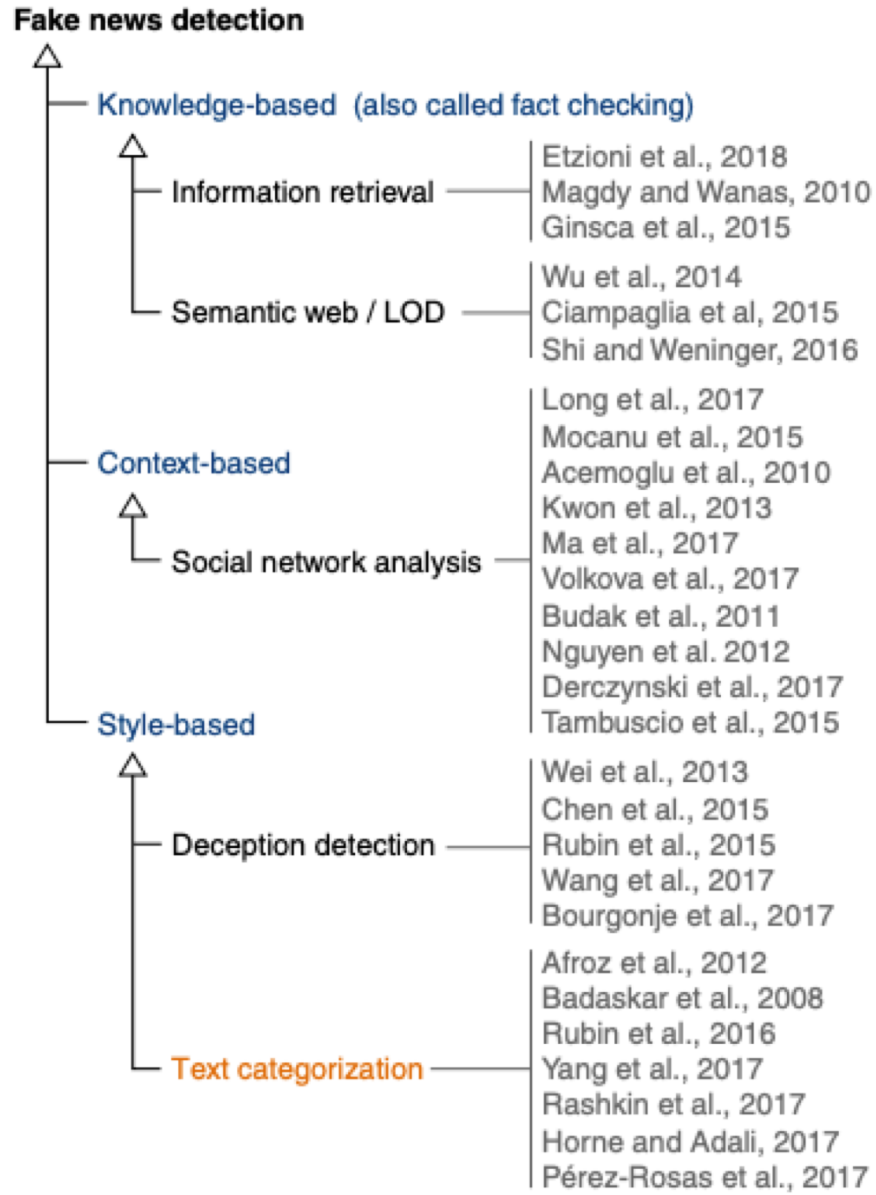


Figure 1: Taxonomy of paradigms for fake news detection alongside a selection of related work.

Data: BuzzFeed-Webis Fake New Corpus

- Annotated into mostly true, mixture of true or false, mostly false and no factual contents
- Each article have been fact-checked by 4 journalists.
- True news: Mostly true
- False news: Mixture of true or false and mostly false
- Disregard: no factual contents

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	55	23	6	0	91	10.9	4.1	4.7	29.0	421.7
The Other 98%	32	3	1	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	107	47	25	36	214	12.9	2.6	2.8	17.3	388.3
Freedom Daily	48	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

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

Methodology

Style Features and Feature Selection

Style Features：

- Common feature: n-grams of characters, stop words and parts of speech
- 10 readability scores: 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
- Dictionary features: the frequency of words and General Inquirer Dictionaries
- Domain-specific features: ratios of quoted words, external links, the number of paragraphs and average length

Feature selection

Prevent overfitting and improve generalization

- Discard the words that occur in less than 2.5% of documents
- Discard the n-grams that occur in less than 10% of documents

Unmasking Genre Styles

- a meta learning approach for authorship verification
- 1. pad document to 500-word long chunks
- 2. measure classification error of linear classifier while iteratively removing the most discriminative features
- 3. analyze the accuracy curve of classification

A steep decrease is more likely than a shallow decrease, if the two documents have been written by the same author.

Training a classifier on many error curves obtained from same-author and different author yields an effective authorship verifier.

Unmasking Genre Styles

- In this research, Unmasking is used to uncover relations between the writing styles and people's political orientation.
- 1. Use two sets of documents (e.g., left-wing articles and right-wing articles) as input.
- 2. Get error curve of linear classifier while iteratively removing features
- 3. Plot error curves for visual inspection, steeper decreases in these plots indicate higher style similarity of the two input document sets

Baseline

- 1. a topic-based bag of words model
- 2. a model using only the domain-specific news style features
- 3. naïve baselines
- Measures

Accuracy, precision, recall and F1

Experiments

- Classifier: WEKA's random forest
- Measurement: perform 3-fold cross-validation where each fold comprises one publisher from each orientation
- Answer the following questions:
- 1. Can (left/right) hyperpartisanship be distinguished from the mainstream?
- 2. Is style-based fake news detection feasible?
- 3. Can fake news be distinguished from satire?

Hyperpartisanship vs. Mainstream

- They hypothesized that maybe the writing style of the hyperpartisan left and right are more similar to one another than to the mainstream.
- A. classify to three class: right-wing, left-wing or mainstream

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
News style	0.39	0.09	0.35	0.59	0.14	0.36	0.49	0.11	0.36	0.53
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 2: Performance of predicting orientation.

Hyperpartisanship vs. Mainstream

- B. binary classification: hyperpartisanship or mainstream
- an extremist should not use a different style dependent on political orientation.

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
News style	0.56	0.54	0.58	0.65	0.47	0.59	0.52
All-hyp.	0.49	0.49	-	1.00	0.0	0.66	-
All-main.	0.51	-	0.51	0.0	1.00	-	0.68

Table 3: Performance of predicting hyperpartisanship.

Hyperpartisanship vs. Mainstream

- C. leave-out classification
- what class would a classifier assign to a left-wing article, if left-wing were remove out in training set, and vice versa?
- Full style-based classifiers have a tendency of classifying left as right and right as left.

Features	**Left**		**Right**	
Trained on:	right+main.	all	left+main.	all
Style	0.74	0.90	0.66	0.89
Topic	0.68	0.79	0.48	0.85
News style	0.52	0.61	0.47	0.66

Table 4: Ratio of left articles misclassified right when omitting left articles from training, and vice versa.

Hyperpartisanship vs. Mainstream

- D. Validation using Unmasking
- Style similarity will be characterized by the slope of a unmasking curve, where a steeper decrease indicates higher similarity

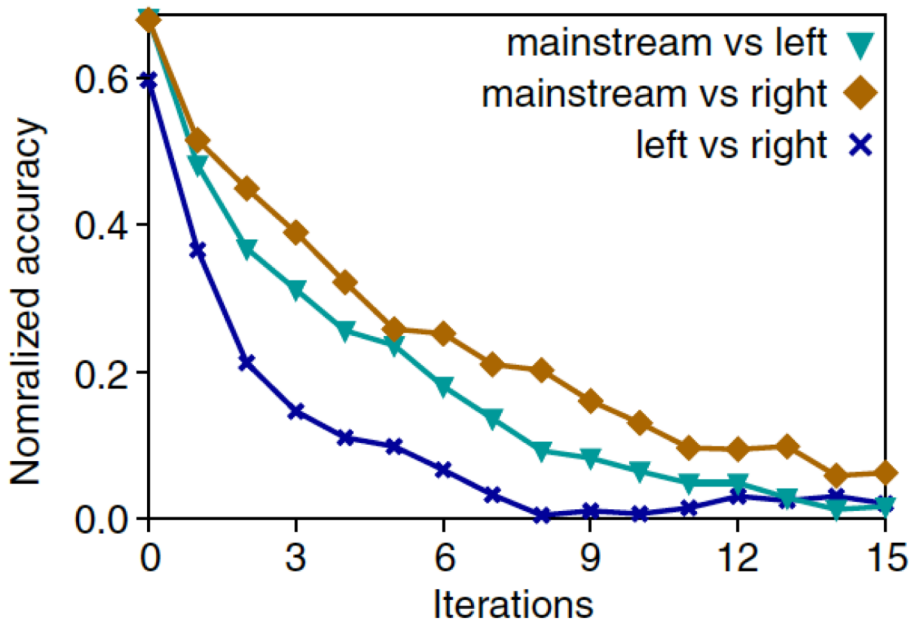


Figure 2: Unmasking applied to pairs of political orientations. The steeper a curve, the more similar the respective styles.

Hyperpartisanship vs. Mainstream

- Based on the experiment B, C and D:

The hyperpartisan left and the right have more in common in writing style than any of the two have with the mainstream

Fake vs. Real (vs. Satire)

- A. Predicting veracity
- Used only the left-wing and right-wing articles for style-based fake news classifier.
- Trained a generic classifier that predicts fake news across orientations, and trained orientation-specific classifiers
- Conclude that style-based fake news classification simply does not work in general.

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

Table 5: Performance of predicting veracity.

Fake vs. Real (vs. Satire)

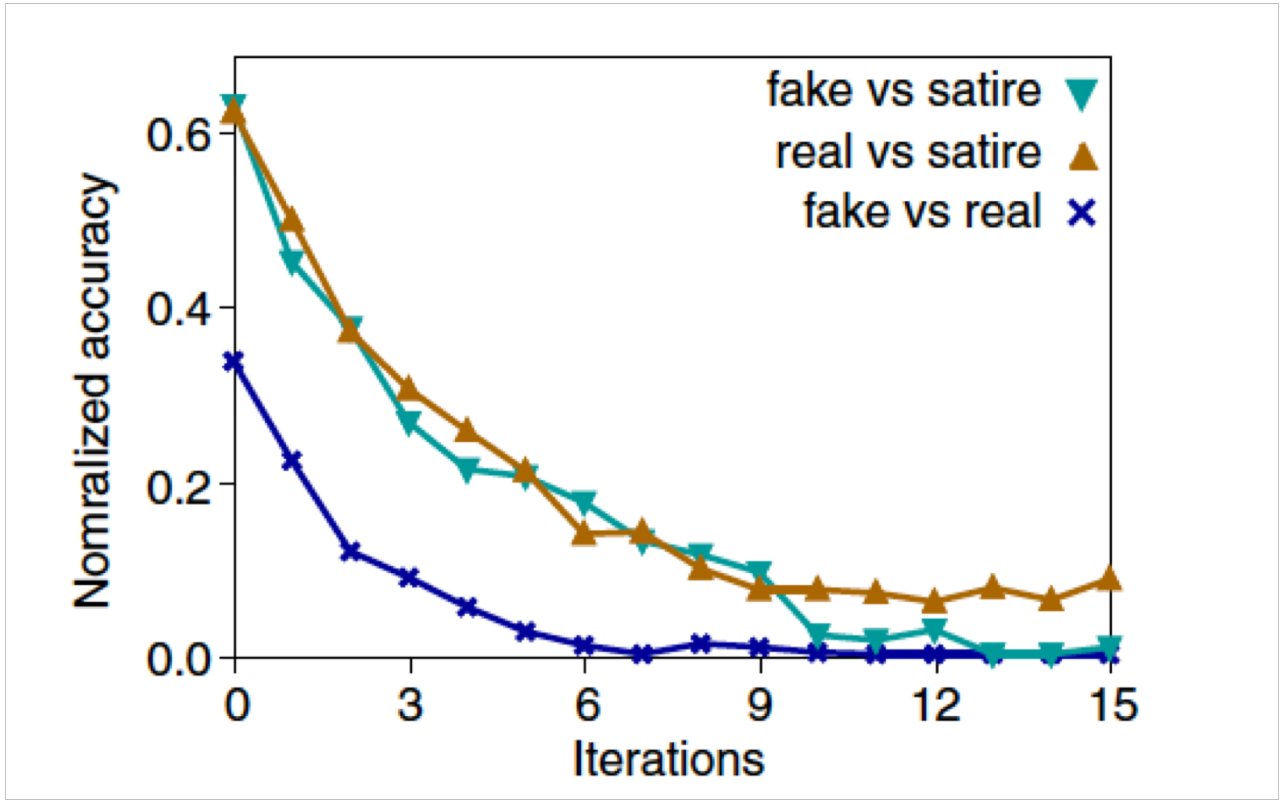
- A. Predicting satire
- Satire should never be filtered, but be discriminated from others.
- Data: satire-detection used by Rubin et al.
- A style-based model is competitive
- Rubin et al.'s classifier is better, which includes features based on topic, absurdity, grammar, and punctuation.

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.	n/a	0.90	n/a	0.84	n/a	0.87	n/a

Table 6: Performance of predicting satire (sat.).

Fake vs. Real (vs. Satire)

- C. Unmasking satire
- Unmasking to compare pairs of the three categories of news (real, fake and satire)
- The style of fake news has more in common with that of real news than either of the two have with satire.



Conclusion

- The writing styles of news of the two opposing orientations are in fact very similar: there appears to be a common writing style of left and right extremism.

- Satire can be distinguished well from other news based on writing style.

- Employed as pre-filtering technologies to separate hyperpartisan news from mainstream news, this approach allows for directing the attention of human fact checkers to the most likely sources of fake news.

Thanks!