

Fake News vs Satire: A Dataset and Analysis

By Jennifer Golbeck et al.

Presented by Travis Reid
CS 895: Web Archiving Forensics
September 24, 2020

Introduction

- Fake news is an important societal issue and is a technical challenge to identify
- Fake news can be:
 - Intentionally false stories that are deceptive
 - Satire
 - Stories that are disliked by others
 - Legitimate news that unintentionally contains false information
- To address fake news, there needs to be clear definitions and examples
- Need to develop mechanisms that can be used to automatically classify and differentiate fake news from satire and legitimate news

Motivation

- To create a dataset that can help the research community answer the questions:
 - Are there differences in the language used in fake news articles vs satirical articles on the same topic such that word based classification approach can be used?
 - Are there significant thematic differences between the fake news articles and the satirical articles on the same topic?

Challenges Of Identifying Fake News

Many information items can have some characteristics that are similar to fake news stories.

- Satire stories
 - Sometimes are presented as news even though it is factually incorrect
 - The intent for satire is not to deceive, but to call out certain behavior that can be considered "bad"
- Legitimate stories
 - Can have factual errors, but should not be considered fake news because it was not intentionally deceptive

Fake News Definition

Even though fake news is not a technical term, a clear definition is needed to identify it.

Definition of fake news used by the authors:

"Fake news is information, presented as a news story that is factually incorrect and designed to deceive the consumer into believing it is true."

Dataset

- The dataset was hand coded and it includes fake news, satirical articles, and links to the original stories
 - 283 fake news articles
 - 203 satirical stories
- In the dataset, each fake news article has a rebutting article from a reliable source
- All of the articles focuses on American politics
 - For consistency of topics for all articles
 - Minimizes the effect on the classifier from topical differences between fake and satirical stories
- Articles were posted between January 2016 and October 2017

Dataset: Spreadsheet

Article Number	URL of article	Fake or Satire?	URL of rebutting article	Fake or Satire?
375	http://www.redflagnews.com/headlines-2016/cdc-proposes-rule-to-apprehend-and-detain-anyone-anywhere	Fake	http://www.snopes.com/cdc-forced-vaccination	Fake
376	http://www.redflagnews.com/headlines-2016/outrage-what-obama-just-did-to-the-white-house-logo-will-m	Fake	http://www.snopes.com/white-house-logo-char	Fake
377	http://www.redflagnews.com/headlines-2016/white-house-cancels-all-obama-appearances-at-hillary-camp	Fake	http://www.snopes.com/obama-veterans-mone	Fake
378	http://www.redflagnews.com/headlines-2016/obama-cuts-26-billion-from-veterans-while-allocating-45-billio	Fake	http://www.snopes.com/obama-veterans-mone	Fake
379	http://www.redflagnews.com/headlines-2016/california-introduces-law-to-jail-anyone-who-questions-man-n	Fake	http://www.snopes.com/california-to-jail-climat	Fake
8	http://president45donaldtrump.com/breaking-study-proves-trump-right-every-citizen-voter-betrayed/	Fake	http://www.snopes.com/illegal-immigrants-200	Fake
9	http://president45donaldtrump.com/0-23-0/	Fake	http://www.snopes.com/trump-undocumented-	Fake
10	http://president45donaldtrump.com/1-9892-2/	Fake	http://www.snopes.com/james-comey-seth-ric	Fake
13	http://usinfofnews.com/index.php/2017/09/10/newly-leaked-report-shows-hillary-not-win-new-hampshirean	Fake	http://www.snopes.com/leaked-videos-clinton-	Fake
14	http://usinfofnews.com/index.php/2017/09/06/katie-couric-speaks-canceling-clinton-interview-called-n-wor	Fake	http://www.snopes.com/hillary-trash-talk-obarr	Fake
15	http://usinfofnews.com/index.php/2017/08/14/shock-claim-hillary-indicted-criminal-charges/	Fake	http://www.snopes.com/clinton-federal-rackete	Fake
16	http://usinfofnews.com/index.php/2017/08/11/boom-president-trump-sent-michelle-obama-bill-shell-never-	Fake	http://www.snopes.com/trump-michelle-obama	Fake
17	http://usinfofnews.com/index.php/2017/07/12/obamas-birth-brother-breaks-silence-reveals-baracks-true-id	Fake	http://www.snopes.com/politics/obama/birthe	Fake
18	http://24wpn.com/index.php/2017/04/29/hillary-clinton-busted-in-the-middle-of-huge-pedophilia-ring-cover-	Fake	https://www.whitehouse.gov/the-press-office/2/	Fake
19	http://24wpn.com/index.php/2017/04/29/breaking-sharia-law-is-gone-the-house-made-a-decision/	Fake	http://abcnews.go.com/Politics/us-house-vote	Fake
20	http://24wpn.com/index.php/2017/08/21/breaking-wasserman-schultz-aide-imran-awan-under-investigation	Fake	https://www.nytimes.com/2017/07/28/us/politi	Fake
21	http://24wpn.com/index.php/2017/08/19/bombshell-new-evidence-suggests-we-are-not-being-told-everythi	Fake	http://www.snopes.com/were-police-told-stand	Fake
22	http://24wpn.com/index.php/2017/06/02/covfefe-mystery-solved/	Fake	http://www.snopes.com/covfefe-arabic-antedilu	Fake
23	http://americannews.com/24812-2/	Fake	http://abcnews.go.com/Politics/republicans-br	Fake
24	http://americannews.com/24759-2/	Fake	http://www.snopes.com/un-official-john-ashe-k	Fake
25	http://americannews.com/breaking-islam-prohibited-public-schools-supreme-court-voted-thanks-trump/	Fake	http://www.factcheck.org/2017/04/scotus-didnt	Fake
26	http://americannews.com/obama-signs-executive-order-banning-the-pledge-of-allegiance-in-schools-natio	Fake	http://www.factcheck.org/2016/09/obama-did-f	Fake
27	http://americannews.com/what-donald-trump-did-for-200-marines-shows-who-he-really-is/	Fake	https://www.cnn.com/2016/12/30/read-all-abc	Fake
38	http://americanprides.com/2017/08/24/wikileaks-massive-exposure-here-are-6-republicans-who-took-bribe	Fake	http://abcnews.go.com/Politics/republicans-br	Fake
40	http://americanprides.com/2017/08/29/obama-stole-the-presidency-new-report-shows-millions-of-illegals-f	Fake	http://www.snopes.com/illegal-immigrants-200	Fake
41	http://americanprides.com/2017/08/08/ben-carson-dies-un-obamas-filthiest-skeleton-500-billion-fraud/	Fake	http://www.factcheck.org/2017/04/carson-didnt	Fake

Dataset: Full Text

ALIENS MEDDLED WITH U.S. ELECTION CLAIMS CONTROVERSIAL EXPERT
<http://worldnewsdailyreport.com/aliens-meddled-with-u-s-election-claims-controversial-expert/>

The famous publisher, producer and TV show host, Giorgio A. Tsoukalos, says he has some "irrefutable evidence" that an extraterrestrial civilization influenced the American presidential election in order to help Trump win the White House. Mr Tsoukalos was invited on the show L.A. Tonight to discuss various topics, when he made an astonishing declaration concerning the federal election.

He said that he'd found some evidence that aliens had used an elaborate "mind-control device" to make Americans vote for Donald Trump.

"We know this isn't the first time that aliens meddle with elections in the United States," claimed Mr Tsoukalos. "They've helped Abraham Lincoln, John F. Kennedy and Ronald Reagan in

Dataset: Restrictions

- No more than 5 articles from a single website
 - This was done so that the sources for the articles are diverse
 - Also reduces the chance of the classifier detecting the language or style used on a certain site
- No borderline cases
 - There is a spectrum from fake to satirical news
 - Some fake news websites have disclaimers that the stories are satire, even though there is nothing satirical about the articles
- The articles not included in this dataset are:
 - Articles that are considered fake news because some dislike the story
 - Legitimate news with unintentional errors

Classification

- Goal: Determine if the word patterns differed enough to do accurate classification
- A model was built to classify an article based on the language used in the article
 - Naive Bayes Multinomial algorithm was used to train the model
 - The accuracy of the classification model is 79.1% and it has a ROC AUC of 0.880
- Since the model is high performing, this suggests that the language differs between fake news and satire in the dataset

Article Themes

- A code book was developed based on 7 major themes:
 1. H - Hyperbolic position against a person or group
 2. F - Hyperbolic position supporting a person or group
 3. D - Discredit a credible source
 4. S - Sensationalist crimes and violence
 5. R - Racist messaging
 6. P - Paranormal theories like Aliens and Flat Earth
 7. C - Conspiracy theories
- Some articles are unlabeled if a code does not apply
- Articles could be labeled with multiple codes

Common And Uncommon Themes

Common Themes:

> 66.67% Hyperbolic criticism (H)

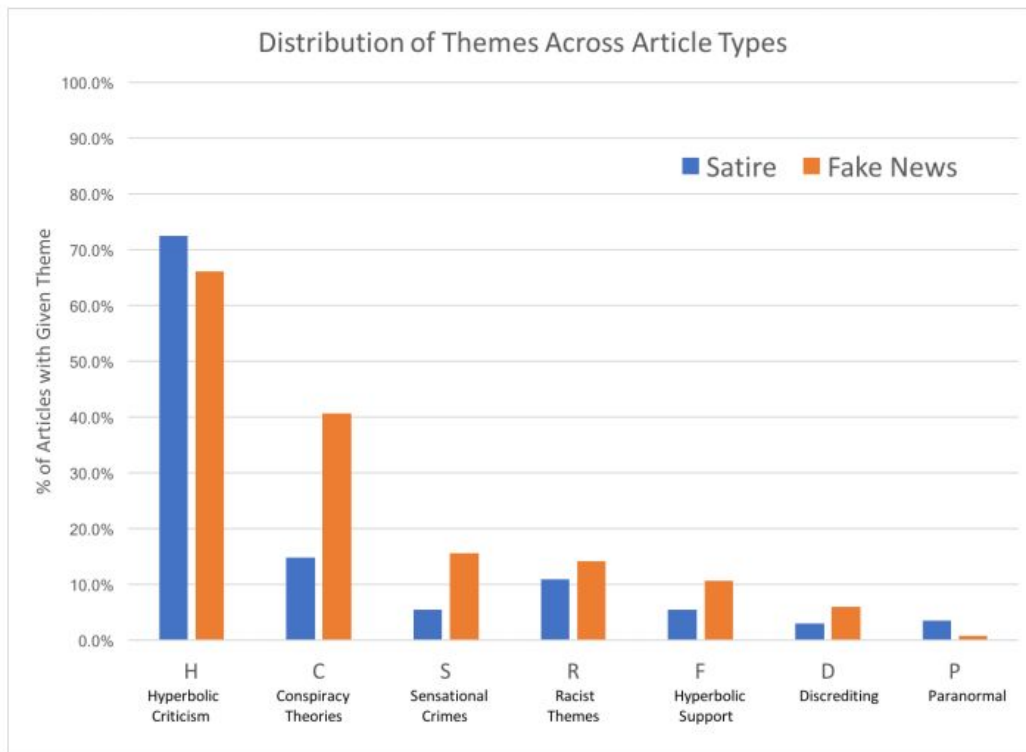
30% Conspiracy theories (C)

Uncommon Themes:

< 5% Discrediting credible sources (D)

< 5% Paranormal theories (P)

Distribution Of Themes



Multiple Themes

- Articles can have more than one theme
 - 43.8% of the articles had multiple themes
- Multiple themes were more common in fake news articles
 - 55.5% of fake news articles had multiple themes
 - 27.6% of satire articles had multiple themes
- Most common pair of themes are conspiracy theories (C) and Hyperbolic criticism (H)
 - 19.5% of all articles
 - 26.9% of fake articles
 - 9.4% of satire

Themes And Classification

- In the dataset, the themes were assigned manually to each article
- They were also able to use the words in an article to detect the theme for an article
 - The model used achieved a high accuracy for most themes except for the hyperbolic criticism of a person or group (H)
- This thematic analysis gives insight that can be useful for automated and qualitative analysis of fake news
 - Since hyperbolic criticism (H) and conspiracy theories (C) are common in fake news, this could be useful for automatically detecting fake stories
- Themes do not provide any differentiating information that can help improve the classification of an article

Contributions

- A dataset of fake news and satire stories
- Content analysis of articles that contain fake news
- This paper also includes an analysis of the dataset and shows results based on language that is useful for classification
 - From the thematic analysis, it has been shown that it is possible to automatically distinguish between fake news and satire