Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking

by: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017.

presentation by: Alec Braynen, Logan Fields, Ben Kreiger

Problem: **Predicting Truthfulness** (of facts and news stories)

- language analysis(comparison) of *satire, hoaxes, propaganda* and *real news*
- prediction of *satire, hoaxes, propaganda* and *real news*
- automatic political fact-checking(Politifact.com 6-point scale)

The truthfulness of stories problem

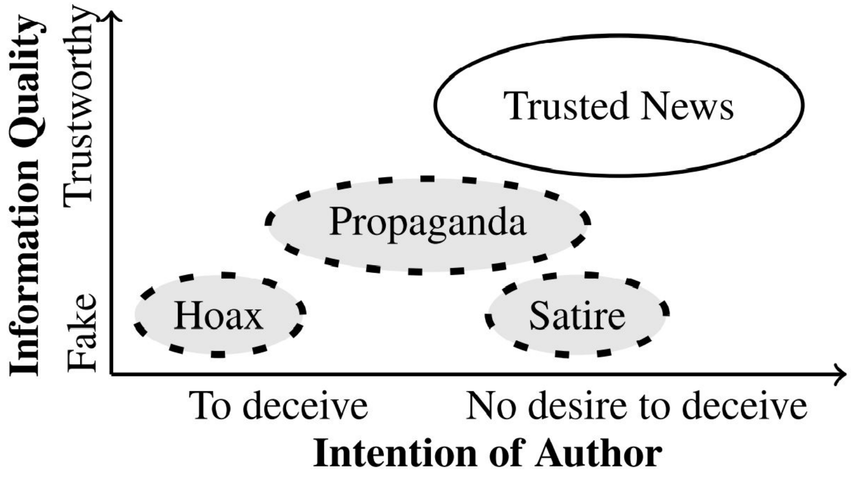

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Problem: **Predicting Truthfulness (of facts and news stories)**

- language analysis(comparison) of *satire, hoaxes, propaganda* and *real news*
- prediction of *satire, hoaxes, propaganda* and *real news*
- automatic political fact-checking(Politifact.com 6-point scale)

The graded notion of truthfulness problem in fact-checking

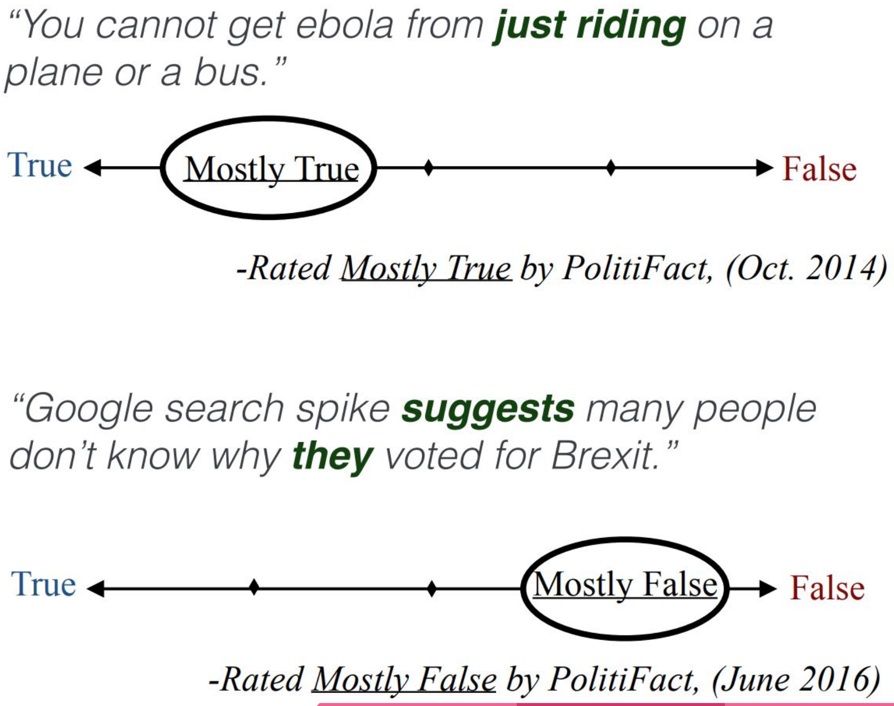

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Datasets

- Unreliable News Dataset:
    - *English Gigaword News* corpus,
    - *The Onion (crawled),*
    - *The Borowitz (crawled),*
    - *Clickhole (crawled),*
    - *American News (crawled),*
    - *DC Gazette (crawled),*
    - *Natural News (crawled),*
    - *Activist Report (crawled),*
- Politifact Dataset *(from Politifact):*
    - 6 labels: True, Mostly True, Half True, Mostly False, False, Pants-on-fire
    - 2 labels: More True, More False

Politifact Dataset Key:

- **TRUE** - The statement is accurate and there's nothing significant missing.
- **MOSTLY TRUE** - The statement is accurate but needs clarification or additional information.
- **HALF TRUE** - The statement is partially accurate but leaves out important details or takes things out of context.
- **MOSTLY FALSE** - The statement contains an element of truth but ignores critical facts that would give a different impression.
- **FALSE** - The statement is not accurate.
- **PANTS ON FIRE** - The statement is not accurate and makes a ridiculous claim.

Datasets: Unreliable News Dataset

- Labels:
  Trusted, Satire, Hoax, Propaganda

- Pros:
  - *Over 60,000 documents*
  - *Lexical Analysis supported scientifically (social science)*
  - *Rational hypothesis for lexicon addition (intensifying lexicon in fake news)*

- Cons:
  - *Fake news sources seem picked by random*

News Type	Source	# of Doc	# Tokens per Doc.
Trusted	Gigaword News	13,995	541
Satire	The Onion	14,170	350
	The Borowitz Report	627	250
	Clickhole	188	303
Hoax	American News	6,914	204
	DC Gazette	5,133	582
Propaganda	The Natural News	15,580	857
	Activist Report	17,869	1,169

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Datasets: Unreliable News Dataset

- Ratio: Truthful News < 1 < Fake News
- S = Satire, H = Hoax, P = Propaganda

Lexicon Markers	Ratio	Example Source	Example Text	Max
Swear (LIWC)	7.00	Borowitz Report	... Ms. Rand, who has been **damned** to eternal torment ...	S
2nd pers (You)	6.73	DC Gazette	**You** would instinctively justify and rationalize **your** ...	P
Modal Adverb	2.63	American News	... investigation of Hillary Clinton was **inevitably** linked ...	S
Action Adverb	2.18	Activist News	... if one **foolishly** assumes the US State Department ...	S
1st pers singular (I)	2.06	Activist Post	**I** think its against the law of the land to finance riots ...	S
Manner Adverb	1.87	Natural News	... consequences of **deliberately** engineering extinction.	S
Sexual (LIWC)	1.80	The Onion	... added that his daughter better not be **pregnant**.	S
See (LIWC)	1.52	Clickhole	New Yorkers ... can bask in the **beautiful image** ...	H
Negation(LIWC)	1.51	American News	There is **nothing** that outrages liberals more than ...	H
Strong subjective	1.51	Clickhole	He has one of the most **brilliant** minds in basketball.	H
Hedge (Hyland, 2015)	1.19	DC Gazette	As the Communist Party USA website **claims**...	H
Superlatives	1.17	Activist News	Fresh water is the single **most** important natural resource	P
Weak subjective	1.13	American News	... he made that **very clear** in his response to her.	P
Number (LIWC)	0.43	Xinhua News	... **7 million** foreign tourists coming to the country in **2010**	S
Hear (LIWC)	0.50	AFP	The prime minister also **spoke** about the commission ...	S
Money (LIWC)	0.57	NYTimes	He has proposed to lift the state **sales tax** on groceries	P
Assertive	0.84	NYTimes	Hofstra has **guaranteed** scholarships to the current players.	P
Comparitives	0.86	Assoc. Press	... from fossil fuels to **greener** sources of energy	P

Datasets: Politifact.com Dataset

- Collection of rated statements from Politifact Fact-Checkers (March 2016)
  - 6 labels: True, Mostly True, Half True, Mostly False, False, Pants-on-fire
  - 2 labels: More True, More False

	More True			More False		
	True	Mostly True	Half True	Mostly False	False	Pants-on-fire
6-class	20%	21%	21%	14%	17%	7%
2-class		62%			38%	

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

- Pros:
  - *From a reputable source*
  - *Labelled and Cleaned with some methodology*
- Cons:
  - *Small dataset*

Datasets: Politifact.com Dataset

Speaker	Statement	Rating
Gil Kerlikowske	More people are driving under the influence of drugs than are driving under the influence of alcohol. A recent roadside survey showed that 16 percent of the people tested, tested positive for illicit or licit drugs. That's significantly higher than alcohol.	2
Tom Coburn	The American people will be appalled to learn the health care bill exempts (congressional) leadership and committee staff.	1
Tom Coburn	Medicare has at least $80 billion worth of fraud a year. That's a full 20 percent of every dollar that's spent on Medicare goes to fraud.	2
Tom Coburn	In 2010, everybody said you can't dare let guns go into the national parks, and of course the rapes, murders, robberies and assaults are down about 85 percent since we did that.	4
Tom Coburn	The government's twice the size it was 10 years ago. It's 30 percent bigger than it was when (Barack) Obama became president.	2

Label Distribution in the Dataset Splits

	0	1	2	3	4	5
Train	499	527	530	364	463	193
Dev	144	147	156	93	125	47
Test	227	252	217	169	160	50

Applying the Datasets: Unreliable News Dataset

1. Analysis:
   a. *First and Second Person pronouns correlate with deceptive news*
   b. *Subjectives, superlatives and modal adverbs correlate with deceptive news*
   c. *Assertive words correlate with truthful news*

2. News Reliability Prediction:
   a. Balanced Training Set and Test Set (20K and 3K articles respectively)
   b. 20% of the training set used as an in-domain dev set
   c. Trained a Max-Entropy classifier with L2 regularization on n-gram tf-idf vectors

Applying the Datasets: Unreliable News Dataset

News Reliability Prediction Results:

Data	Sources	Random	MaxEnt
Dev	in-domain	0.26	0.91
Test	out-of-domain	0.26	0.65

Table 3: F1 scores of 4-way classification of news reliability.

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Applying the Datasets: Politifact Dataset

1. Predicting Truthfulness:
   a. Dataset Split: train: 2575, dev: 712, test: 1074
   b. Trained a LSTM model that takes a sequence of words (the fact) as input and predicts the Politifact rating
      i. Word sequences as input
      ii. LSTM output is concatenated with LIWC feature vectors
   c. LSTM word embeddings initialized with 100-dimensional GLOVE embeddings
      i. 300D hidden state | batch size of 64 | ADAM optimizer | 10 Epoch | CCE Loss
   d. Results compared against Maximum Entropy and Naive Bayes Model
      i. TF-IDF vectors as input
      ii. LIWC measurements concatenated to TF-IDF vectors

Applying the Datasets: Politifact Dataset

Politifact Truthfulness Prediction Results (F1 Score):

	2-CLASS		6-CLASS	
	text	+ LIWC	text	+ LIWC
Majority Baseline	.39	-	.06	-
Naive Bayes	.44	.58	.16	.21
MaxEnt	.55	.58	.20	.21
LSTM	.58	.57	.21	.22

Table 5: Model performance on the Politifact validation set.

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Applying the Datasets: Politifact Dataset

Politifact Truthfulness Prediction Results (F1 Score):

Model	Feature	2-class	6-class
Majority Baseline		.39	.06
Naive Bayes	text + LIWC	.56	.17
MaxEnt	text + LIWC	.55	.22
LSTM	text + LIWC	.52	.19
LSTM	text	.56	.20

Table 6: Model performance on the Politifact test set.

image source: Rashkin, Hannah, et al. "Truth of varying shades: Analyzing language in fake news and political fact-checking." *Proceedings of the 2017 conference on empirical methods in natural language processing*. 2017

Logistics

A. How the workload was split
   - Alec: presentation, dropout (hyperparameter) tuning, testing, bart-summarized input experiment
   - Logan: hyperparameter tuning, testing of final test set, presentation
   - Ben: getting initial code functional, initial testing, testing with speaker name, presentation

B. What software/servers/tools we used
   - Google Colab
   - Anaconda
   - Pytorch
   - Cuda
   - Ubuntu VM

Experiments and Improvements: Hyperparameters

Learning Rate:

	ROC AUC	Macro-F1	MCC	Accuracy
5e-4	0.563491	0.059908	0	0.219101
1e-4	0.598239	0.167008	0.0773	0.26264
5e-5	0.624029	0.225923	0.091139	0.268258
1e-5	0.641404	0.233598	0.1185	0.294944
5e-6	**0.64976**	**0.277445**	**0.137417**	**0.304775**

Experiments and Improvements: Hyperparameters

Dropout:

	ROC AUC	Macro-F1	MCC	Accuracy
0.1	**0.64976**	**0.277445**	**0.137417**	**0.304775**
0.15	0.633009	0.217378	0.118426	0.292135
0.2	0.644552	0.189149	0.111894	0.282303
0.25	0.635464	0.223705	0.109312	0.283708
0.3	0.633937	0.24163	0.111202	0.279494

Experiments and Improvements: Hyperparameters

Weight Decay:

	ROC AUC	Macro-F1	MCC	Accuracy
0	**0.64976**	**0.277445**	**0.137417**	**0.304775**
1e-5	0.643207	0.205765	0.099712	0.279494
1e-4	0.640692	0.236672	0.112374	0.286517
1e-3	0.634457	0.242044	0.103662	0.285112
1e-2	0.64112	0.231837	0.101782	0.279494

Experiments and Improvements: Adding Speaker Name

- Speaker: Statement
  - Test F1: 0.19
- Speaker said, “Statement”
  - Test F1: 0.18

Speaker	Statement	Rating	speakerstatement
Gil Kerlikowske	More people are driving under the influence of drugs than are driving under the influence of alcohol. A recent roadside survey showed that 16 percent of the people tested, tested positive for illicit or licit drugs. That's significantly higher than alcohol.	2	Gil Kerlikowske: More people are driving under the influence of drugs than are driving under the influence of alcohol. A recent roadside survey showed that 16 percent of the people tested, tested positive for illicit or licit drugs. That's significantly higher than alcohol.
Tom Coburn	The American people will be appalled to learn the health care bill exempts (congressional) leadership and committee staff.	1	Tom Coburn: The American people will be appalled to learn the health care bill exempts (congressional) leadership and committee staff.
Tom Coburn	Medicare has at least $80 billion worth of fraud a year. That's a full 20 percent of every dollar that's spent on Medicare goes to fraud.	2	Tom Coburn: Medicare has at least $80 billion worth of fraud a year. That's a full 20 percent of every dollar that's spent on Medicare goes to fraud.
Tom Coburn	In 2010, everybody said you can't dare let guns go into the national parks, and of course the rapes, murders, robberies and assaults are down about 85 percent since we did that.	4	Tom Coburn: In 2010, everybody said you can't dare let guns go into the national parks, and of course the rapes, murders, robberies and assaults are down about 85 percent since we did that.
Tom Coburn	The government's twice the size it was 10 years ago. It's 30 percent bigger than it was when (Barack) Obama became president.	2	Tom Coburn: The government's twice the size it was 10 years ago. It's 30 percent bigger than it was when (Barack) Obama became president.

Experiments and Improvements: Bart-Summarized Input

- Feed input to a BART transformer that summarizes the input
- Feed these summarized inputs with labels to distibert
  - F1 Score: 0.1865483
- Summarization is slow

```
summarized_list = []
dictionary_representation = {}

for i in traindf["Statement"]:
  bart_summarized = summarizer(i, max_length=50, min_length=10, do_sample=False)
  newL = dict()
  newL['summary'] = bart_summarized['summary_text']
  summarized_list.append(summarizer(i, max_length=50, min_length=10, do_sample=False))
```

Final Testing

F1: 0.1687, Accuracy: 22%

Epoch	Training Loss	Validation Loss	Roc auc	Mcc	Macro-f1	Accuracy
1	No log	1.714664	{'roc_auc': 0.612694840642986}	{'matthews_correlation': 0.032536707794801216}	{'f1': 0.10072989227665487}	{'accuracy': 0.2247191011235955}
2	1.741600	1.685455	{'roc_auc': 0.6312237741680021}	{'matthews_correlation': 0.06611935367327114}	{'f1': 0.18444185979558406}	{'accuracy': 0.2570224719101235}
3	1.741600	1.667408	{'roc_auc': 0.6359250630900846}	{'matthews_correlation': 0.08782470639734205}	{'f1': 0.19072568763852057}	{'accuracy': 0.2640449438202247}
4	1.684000	1.659217	{'roc_auc': 0.6389031606447513}	{'matthews_correlation': 0.08944916299030413}	{'f1': 0.21167567491356812}	{'accuracy': 0.2710674157303371}
5	1.595700	1.655409	{'roc_auc': 0.6395041415723319}	{'matthews_correlation': 0.1006228991762991}	{'f1': 0.22637571616079566}	{'accuracy': 0.2794943820224719}
6	1.595700	1.662016	{'roc_auc': 0.6387182531157638}	{'matthews_correlation': 0.09676561450700082}	{'f1': 0.22738546471826837}	{'accuracy': 0.2752808988764045}

[135/135 00:03]

```
PredictionOutput(predictions=array([[ 0.9887807 ,  1.446709  ,  0.2872495 , -0.82790744, -0.23150116,
        -1.8411783 ],
       [ 0.9333766 ,  0.7932639 ,  0.10226732, -0.57166034,  0.01741575,
        -1.5018946 ],
       [ 0.8713634 ,  0.6651125 ,  0.22604308, -0.22939667, -0.06033457,
        -1.29253   ],
       ...,
       [ 0.7864591 ,  1.1178788 ,  0.61243993, -0.6561254 , -0.2934225 ,
        -1.8323919 ],
       [ 0.786129  ,  0.5268515 ,  0.0891403 , -0.64283395,  0.69543165,
        -1.1025784 ],
       [-0.18240988, -0.41855165,  0.3546308 ,  0.42936867,  0.29357514,
        -0.44964722]], dtype=float32), label_ids=array([5, 1, 0, ..., 2, 2, 4]), metrics={'test_loss': 1.7254387140274048, 'test_ROC AUC': {'roc_auc': 0.5776803780379732}, 'test_MCC':
{'matthews_correlation': 0.022932915719616274}, 'test_Macro-F1': {'f1': 0.16867278334591787}, 'test_Accuracy': {'accuracy': 0.21767441860465117}, 'test_runtime': 4.2014,
'test_samples_per_second': 255.866, 'test_steps_per_second': 32.132})
```

Questions?