Read Between the Lies:

Verbal Deception Detection with Natural Language Processing

Victor Palacios

Summary

Question:

Do deceptive statements differ from truthful statements in terms of linguistic markers? If yes, can machine learning predict deception?

Method:

Scrape political statements that are truthful or deceptive from the website politifact.com and perform NLP and ML

Results:

Statistically significant markers exist for lying and truth-telling, and machine learning accuracies on holdout data were 57-58%.

Background: Linguistic Markers in General

Depression, Smirnova, et al., 2018

Patients medical depression have an "increased use of personal and indefinite pronouns, and verb use in continuous/imperfective and past tenses"

Narcissism, Rathner, et al., 2018

"[D]epressive individuals use less social words and more anxiety-related words, narcissists do the opposite"

Background: Linguistic Markers in Deception

In Pamela Meyer's Ted Talk, "How to Spot a Liar",

she identifies two indicators of deception: "non-contracted denial" (i.e., "did not" versus "didn't") and distancing language (i.e., "that woman").

Method - Population

Data Gathering

16,611 political statements were scraped from https://www.politifact.com

About Politifact

- A not-for-profit national news organization
- fact checks in a nonpartisan manner.
- Statements are found from: transcripts, speeches, new stories, press releases, viral posts, and campaign brochures.

Statement Classification

3 editors determine: *True, Mostly True, Half True, Mostly False, False,* and *Pants on Fire*.

Method - Sample

Of the 16,611 statements only a small subset was viable for analysis

1,147 viable statements fulfilled the following criteria:

(1) Only true (50%) or false (50%) statements (*Mostly True, Half True, Mostly False,* and *Pants on Fire* were excluded),

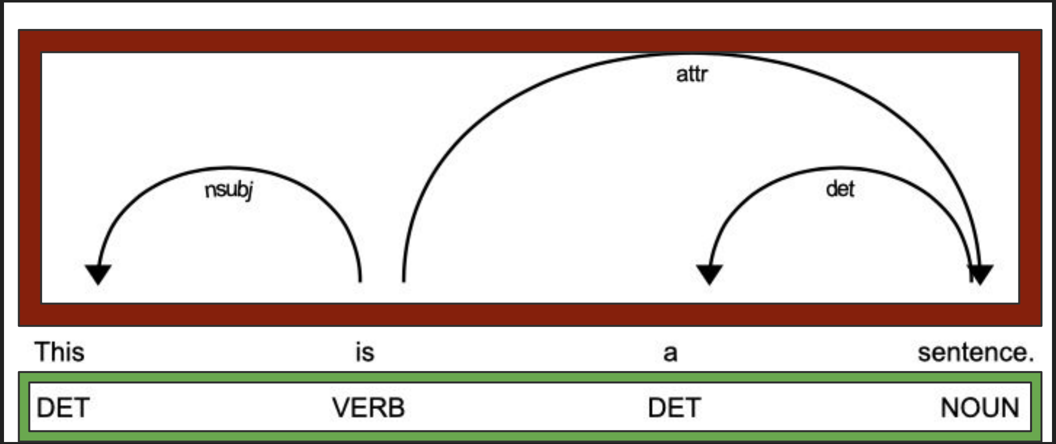
(2) Only direct quotes (direct quotes are defined as beginning and ending with a quotation mark), and

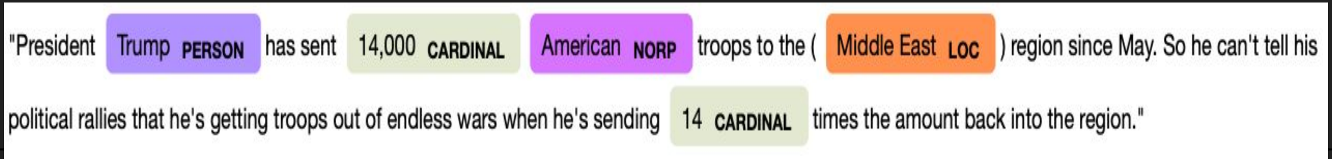
(3) Only 1-3 statement per person were allowed

These restrictions allow for a more controlled examination of deception and ensure equal representation of different speech patterns

Method - Text Tagging

spaCy was used to tag parts of speech (adjective, verb, etc.), syntactic dependency, and named entity recognition





Method - Other Engineered Variables

1. Number of characters
2. Number of words
3. Average length of words
4. Readability score

   Gunning Fog Score depends on sentence length and syllable amount. It scores as follows:

      17 = College Graduate

      …

      6 = 6th grade

Method - Normalization

	Text
Original Statement	"@ ! i DON'TTT, won't, a féél can't not USING the NLP 27x maaaah?"
Lightly Normalized	"@ ! i don'ttt, won't, a feel can't not using the nlp 27x maaaah?"
Fully Normalized	'feel use nlp maaaah'

Method - Machine Learning

1. Datasets

   2 Datasets: Lightly Normalized Text and Fully Normalized Text

2. Classifiers

   Multinomial Naive Bayes, Logistic Regression, Random Forest

3. Vectorizers

   Bag of Words, n-grams, Term Frequency-Inverse Document Frequency

Method - Machine Learning

4. Validation

   Data split into a train (70%) and hold-out set (30%). The train set was grid-searched and cross- validated (n_folds = 3)

5. Optimization

   Accuracy used because the data was balanced (50% lies and 50% truth)

   Note: ROC-AUC was also used with less success

Results - Statistics - Part of Speech

	Mean Number in ***Truthful*** *Statements*	*Mean Number in* ***Deceptive*** *Statements*	*p-value*
Adjective	***1.42***	*1.23*	*.006*
Noun	***4.35***	*3.95*	*.002*
Verb	*1.72*	***1.89***	*.040*
Number	***0.99***	*0.72*	*.000*
Adverb	***2.33***	*2.05*	*.002*
Determiner	***1.87***	*1.64*	*.006*

Results - Statistics - Various Counts

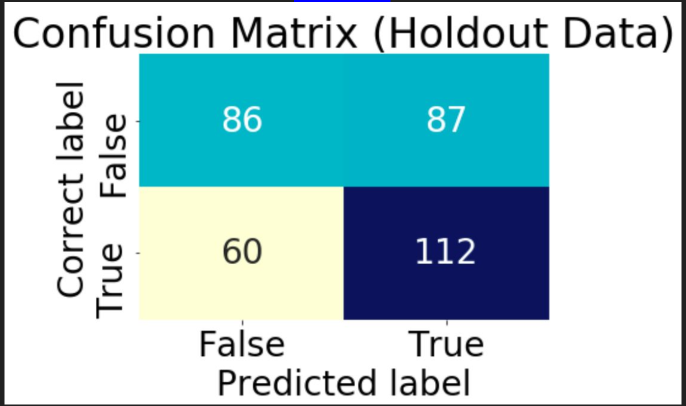
	*Mean Number in **Truthful** Statements*	*Mean Number in **Deceptive** Statements*	*p-value*
Character Count	***108.8***	*103.4*	*.042*
Word Count	***18.4***	*17.3*	*.021*
Average Length of Word	5.0	**5.2**	*.003*

Results - Machine Learning

Best Accuracy on Training Set: Light normalization / TF-IDF / Logistic Regression model

Cross-validated accuracy score on train set: 63%

Holdout Accuracy: 57%



Interestingly, the model was much better at predicting truthful statements than it was at predicting deceptive statements with recall of .65 and .50, respectively.

Conclusion

Statistics

Liars appear to rely on a more complex vocabulary centered on action-based words because they lack a true experience. Conversely, their truthful counterparts appear to add descriptions and thus are more verbose in their prose.

Machine Learning

Lightly normalized text delivered higher accuracy scores implying that the removal of stopwords is not a fruitful endeavor in deception detection, but accuracy was below 60% (maximum of 58%).

Deception may be more complex than any model can capture.