Significant Paper Report

Paper title: Computationally Detecting and Quantifying the Degree of Bias in Sentence-Level Text of News Stories
(https://pdfs.semanticscholar.org/3722/40da2a416abfb406426af71f084c988ff7d9.pdf)
Authors: C.J. Hutto, Dennis Folds, Scott Appling
Presented: HUSO 2015: The First International Conference on Human and Social Analytics

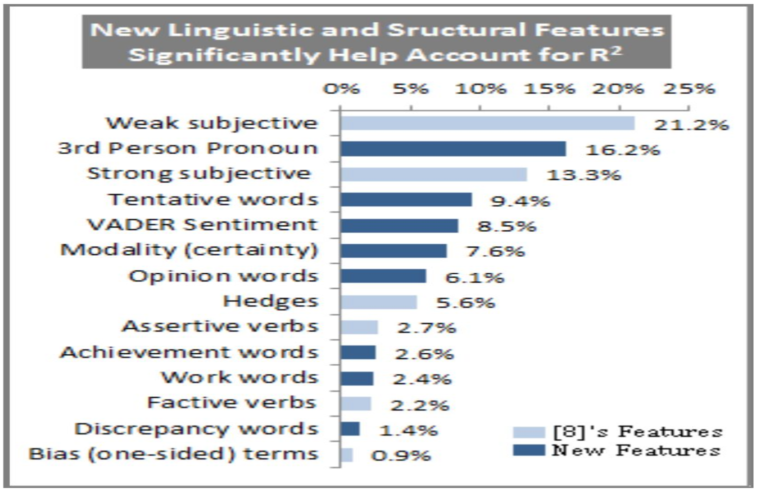
Summary:

Although the given paper is about detecting fake news, it perfectly fits the case of solving the issue of fake review detection from the customer review text. The paper discusses 26 common linguistic and structural cues of biased language, incorporating sentiment analysis, subjectivity analysis, modality (expressed certainty), hedge phrases, and many other features to build a model for detecting the degree of bias in the given text news.

These features can be broadly divided as follows:

I. **Structural analysis at the sentence level**
   This consists of Sentiment score, Subjectivity, Modality, Mood and Readability of the text.
II. **Linguistic analysis at the sentence level**
   This aims at detecting either epistemological bias(knowledge contained in the text) or framing bias(avoiding risks while writing the review). Includes features such as Factive verbs, Implicative verbs, Assertive verbs, Hedges, Strong subjective intensifiers, weak subjective intensifiers, bias terms, Opinion words, Degree Modifiers, Coherence Markers, Causation words, Certainity words, Tentative words, 3rd person pronoun, Achievement words, Work words, Discrepancy words, Conjunctions, prepositions, Adverbs and Auxiliary verbs.

Using the R package of Akaike Information Criterion (AIC) relative quality of each feature is measured for characterizing the degree of bias in the text. For example, the sentence-level measure for modality is a stronger indicator of bias than the linguistic cues associated with LIWC certainty words, so the certainty words feature was removed from the model. Similarly, it was found that measures for implicative verbs, degree modifiers, coherence markers, causation words, conjunctions, prepositions, adverbs, and auxiliary verbs were all relatively poor indicators of sentence-level bias; therefore, all these features were removed from the final model.



The above graph depicts the relative importance of each of the features that were considered. Features were selected accordingly.

TABLE II: COEFFICIENTS, ERROR, T-VALUES, AND *P*-VALUES FOR THE IMPROVED MODEL. $F(14,26) = 11.3$, $P = 1.04E-07$.

	b	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.56	0.19	-3.02	0.006
Strong subjective	5.10	1.07	4.74	0.000***
3rd Person Pronoun	8.36	1.95	4.30	0.000***
Weak subjective	4.87	1.19	4.08	0.000***
Modality (certainty)	0.52	0.15	3.42	0.002**
VADER Sentiment	0.35	0.11	3.13	0.004**
Tentative words	4.60	1.65	2.79	0.010**
Opinion words	-2.05	0.95	-2.16	0.040*
Achievement words	5.74	2.66	2.16	0.040*
Factive verbs	-16.64	8.39	-1.98	0.058`
Work words	9.81	5.20	1.89	0.070`
Hedges	3.06	1.75	1.75	0.092`
Assertive verbs	-3.58	2.16	-1.66	0.110
Discrepancy words	5.66	3.62	1.56	0.130
Bias (one-sided) terms	-0.95	0.74	-1.30	0.206

Signif. level codes: $p < 0.001$*** $p < 0.01$** $p < 0.05$* $p < 0.1$`

The above table depicts the preliminary results of the linear regression model for the 14 features selected in the previous step.

Key Learnings and Implementations from the paper:

This paper clearly lays down the features that can help us build the classifier for biased/unbiased text. The two types of features (Structural and linguistic) are the two key kinds of factors that capture major possible factors that can help in biased text review detection. It also teaches how we can Identify the most important features out of the ones we identified in the beginning.

In our project, we used this paper as a basis and identified certain features like Polarity, word count, Modality, etc of the review text to create a model using ensemble classifier giving an overall accuracy of ~78%.