

Automatic Identification of Political Ideology in Online News Articles

Vikki FIORE

Montclair State University, Montclair, NJ, USA, fiorev3@montclair.edu

ABSTRACT: The following research discusses text analysis approaches to automatically categorize news articles based on their political ideology. In this case, ideology is defined as a writer expressing either a liberal or a conservative point of view. This classification is done at both the document and the phrase level, as previous research has indicated that doing so increases classifier performance over using a “bag of words” approach. Linguistic features related to lexical richness are extracted from the articles via Python, and features related to emotions and values are extracted via the Linguistic Inquiry and Word Count software. The machine learning software Weka is then used to apply various classification algorithms on the numeric features. Additionally, Amazon Mechanical Turk is used to measure human accuracy and inter-rater agreement on identifying the ideology of the same texts. In all, the trained classifiers perform well above the baseline and outperform the human annotators on the same tasks.

KEYWORDS: text analysis, natural language processing, applied linguistics, politics, bias detection, machine learning, artificial intelligence

1. Introduction

Overall, recent research indicates that it is possible for bias detection models to perform equally as well as humans, if not better. Various methods exist to create these models, including crowdsourcing, sentiment analysis, and machine learning techniques such as neural networks. Most research has been done at the phrase or sentence level as opposed to the word level, with some research even making use of the document level in analysis. Using some form of stemming or lemmatization has also been rather consistent in research of this type. All human writers have opinions and bias, and therefore it is virtually impossible to find a completely bias-free source of information. However, the degree to which a source is biased is worth continued analysis, as are the potential ways in which bias can be detected and reduced.

2. Literature Review

Opinions and biases are expected in media, news, and other such information sources. Even writing which claims to come from neutral, non-politically-affiliated source will be prone to some type of bias simply based on its writer’s word choice and opinion. In their paper, Iyyer, Enns, Boyd-Graber, and Resnik (2014) from University of Maryland denoted that, “Many of the issues discussed by politicians and the media are so nuanced that even word choice entails choosing an ideological position”. In the case of this research, a sentence was said to show ideological bias and opinion if the writer’s political ideology was evident from the text. While much existing research toward bias detection used “bag of words” classifiers, Iyyer, Enns, Boyd-Graber, and Resnik (2014) focused at the level of whole documents by applying recursive neural networks to political ideology detection. This model outperformed existing models on two separate

datasets. However, the researchers also noted the challenge of defining opinion and opinion bearing language. Their research used the Linguistic Inquiry and Word Count Lexicon (LIWC) to extract features.

Previously, Kim and Hovy (2005) had created an automatic method for obtaining opinion-bearing words, which was effectively used to identify opinionated sentences. This research proposed that, “a profitable approach to opinion requires a system to know and/or identify at least the following elements: the topic (T), the opinion holder (H), the belief (B), and the opinion valence (V)”. Kim and Hovy (2005) noted that this did not provide very satisfactory results on arbitrary text. However, if the relative frequency of a word in opinion-bearing texts compared to non-opinion bearing text was known, statistical information could be used in place of lexical information. This reaffirms the position of Iyyer, Enns, Boyd-Graber, and Resnik on the need to use sentence level analysis in addition to lexicons.

Recasens, Danescu-Niculescu-Mizil, and Jurafsky (2013) of Stanford University followed a similar approach. These researchers looked at the before and after versions of articles that had been revised to eliminate bias. In this case, the context of a word was defined as a 5-gram, which is similar to other research that analyzed at the phrase level. Lemmatization was also incorporated. The performance of the regression model was then compared to that of humans. Overall, the research found that humans could correctly identified a biased word 30% of the time, with an inter-rater reliability of 40.73% agreement amount human raters. The linguistically informed model for bias detection performed almost as well as these humans tested on the same task. These Stanford researchers were not the first to use crowdsourced human intelligence as a research tool. The idea of crowdsourcing, or gathering information by utilizing the input of a group of people, was first popularized in a 2006 Wired magazine article written by Jeff Howe (2006). Crowdsourcing can be described as a scenario in which combining and averaging the input of a group of people will likely yield a correct result to a posed question. In her review on the growth of crowdsourcing, Kerri Wanzky (2017) noted that, “if a million individuals were to contribute towards answering a problem via crowdsourcing, there would be a 97.7% likelihood that the crowd would arrive at the correct answer”. Wanzky notes that intelligent crowds require diversity, independent to limit the influence of one person’s opinion on another, decentralization, and aggregation to combine the opinions of the crowd. The growth of technology over the last ten years has also had a huge impact in this phenomenon, due to an abundance of new data and connections between people worldwide.

3. Methodology

3.1 Data Collection

Collections of online news documents were gathered. The first collection represents a liberal point of view and contains documents from five openly liberal news sources: Slate, The New Yorker, The Huffington Post, Jacobin, and Alternet. The second collection represents a conservative point of view and contains documents from five consistently conservative news sources: Breitbart, The Federalist, Fox, Daily Wire, and New York Post. For each news source, documents were collected on the following topics of political controversy: gun control, abortion, immigration, tax reform, and same sex marriage. In total, each point of view collection contained twenty-five documents from five different sources on five different political issues. This methodology was used in order to prevent one topic or one news source from skewing the results.

3.2 Linguistic Features

In order to compare point of view differences, a total of 71 features were extracted from the text data set and analysed for significance.

3.2.1 Lexical Richness

A simple script was created using the LexicalRichness Python library to extract the six following lexical diversity features from each document.

Table 1. Lexical Richness Features

ttr	type-token ratio computed as t / w (Chotlos 1944, Templin 1957)
rttr	root TTR computed as t / \sqrt{w} (Guiraud 1954, 1960)
cttr	corrected TTR computed as $t / \sqrt{2w}$ (Carrol, 1964)
msttr	mean segmental TTR (Johnson 1944)
mattr	moving average TTR (Covington 2007, Covington and McFall 2010)
mtld	measure of Lexical Diversity (McCarthy 2005, McCarthy and Jarvis 2010)

3.2.2 Linguistic Inquiry and Word Count (LIWC)

Sixty-five additional features were gathered via the Linguistic Inquiry and Word Count tool. Commonly referred to as LIWC, this software calculates and quantifies text features that contain certain words, parts of speech, and word categories such as emotionality, cognitive processes, and many others. It was originally created for psychological analysis of texts, but more recently LIWC2015 has been used for research across many different fields. This is the same software used in the research of Iyyer, Enns, Boyd-Graber, and Resnik (2014) on opinion bearing language. Most similar research of this nature obtained the best results when using phrase or document level analysis. For this reason, I chose to do the LIWC feature extraction and analysis at the document level.

3.2.3 N-Grams

The fifty documents were combined into two corpora; one containing all the articles from a liberal point of view and the other containing the conservative point of view articles. Since previous researchers seemed in agreement that using phrase, or sentence level analysis increased classification performance over word-level analysis, I decided on 5-grams and 6-grams for this research. I extracted a list of the top fifteen 5-grams and top fifteen 6-grams from each corpus by using a simple python script along with the AntConc software. This created a total of thirty phrases for each point of view.

4. Results

4.1 Significant Features

The Waikato Environment for Knowledge Analysis machine learning software, commonly referred to as Weka, was used to extract the most significant LIWC and Lexical Richness features from the liberal and conservative corpora via the software's Correlation Ranking Filter. The top eight features were discovered to be as follows:

- 0.503 rttr
- 0.498 cttr
- 0.329 Comma
- 0.308 Dash
- 0.303 prep
- 0.275 swear
- 0.241 family
- 0.236 Analytic

Interestingly, most of these features are indicative of writing style rather than the sentiment of an article. The top four features in particular should have little influence on content, yet they are significant in distinguishing point of view in the document set.

4.2 Document Point of View Classification

The significant features were then analyzed using machine learning algorithms in Weka. The SMO supportVector algorithm obtained a 75% correct classification rate using a 10-fold cross validation. The best results were obtained using the Naïve Bayes classification algorithm, also with a 10-fold cross validation. This resulted in a 77.08% correct classification rate, much improved when compared to the baseline of 50%. More details on these results are shown in the table below.

Table 2. LIWC Document Level Results

	TP Rate	FP Rate	Precision	Recall	F-Measure
Conservative	0.833	0.292	0.741	0.833	0.784
Liberal	0.708	0.167	0.810	0.708	0.756
Weighted Avg.	0.771	0.229	0.775	0.771	0.770

4.3 N-Gram Sentiment Analysis

4.3.1 Human Intelligence

Amazon Mechanical Turk was used to gain insight into human performance on political ideology classification. The Turkers were asked to view the collection of sixty n-grams and decide if the phrase was more likely to be from a liberal or a conservative news article. Five different Turkers reviewed each phrase. The overall correct classification rate was 43.6%, compared to a baseline of 50%. The Fleiss Kappa score for inter-rater agreement among Turkers was calculated using a simple Python script and came to 0.16, showing slight agreement among raters. This further proves the difficulty of this classification task for humans.

4.3.2 Automatic Classification

The thirty liberal and thirty conservative n-gram phrases were then labeled as to their ideology, and a sentiment analysis classification was performed using StringToWordVector classification in Weka. The best results showed a correct classification rate of 75.4% using a Naïve Bayes 10-fold cross validation, again using a 50% baseline.

Table 3. LIWC N-Gram Results

	TP Rate	FP Rate	Precision	Recall	F-Measure
Conservative	0.600	0.097	0.857	0.600	0.706
Liberal	0.903	0.400	0.700	0.903	0.789
Weighted Avg.	0.754	0.251	0.777	0.754	0.748

5. Discussion and Future Directions

Overall, the automatic classifications achieved results well above the baseline at both the phrase and the document level. The phrase level analysis outperformed humans by an even higher margin. In future, it would be beneficial to replicate this research using a larger dataset. Using an even wider variety of new sources and topics could prove useful

as well. Previous research has focused on the content of articles, such as specific collocations, bigrams, or emotions related to political ideology. However, based on the eight features of the data set that were found to be most significant, it would seem as though writing style may have distinct variations between liberal and conservative news articles. This is something that should be explored in depth at a future time.

Additionally, due to the nature of Wikipedia writing, it lends itself well to analysis and has been used as a resource in other research projects on bias detection. Unlike most written encyclopedia and news sources, Wikipedia is unique in that each article is written and edited by a number of people from varying ideologies and backgrounds. In this way, Wikipedia is a prime example crowdsourcing at use, due to the website's policy on neutral point of view (NPOV), according to which all articles should, "fairly, proportionately, and as far as possible without bias, all significant views that have been published by reliable sources" (Wikipedia, 2013). Additionally, Wikipedia's style guide also asks writers and editors to use nonjudgmental language, to take note of opposing points of view, to avoid writing off controversial topics as mere opinion, and to also avoid stating opinions as fact.

The initial goals of this research had been to also build a three-way classifier to investigate if Wikipedia is truly neutral or if a political bias still exists, and a collection of twenty-five total Wikipedia documents on the same political issues as the liberal and conservative articles was gathered. However, due to an abundance of stylistic differences between news and encyclopedia writing, the Wikipedia dataset was not used in this analysis. In future, these stylistic differences would need to be accounted for, perhaps by comparing Wikipedia to other encyclopedias with known political affiliations, such as Conservapedia. Additionally, news articles from sources considered to be more neutral could be compared to the openly liberal and conservative news sources. In all, the task of assessing the neutrality of Wikipedia is possible and would make for interesting future discoveries.

References

- Carroll, J. B. 1964. "Language and thought." *Reading Improvement* 2(1): 80.
- Chotlos, J. W. 1944. IV. "A statistical and comparative analysis of individual written language samples." *Psychological Monographs* 56(2): 75.
- Covington, M. A. 2007. MATTR user manual (CASPR research report 2007-05). Athens, GA.
- Covington, M. A., & McFall, J. D. 2010. "Cutting the Gordian knot: The moving-average type-token ratio (MATTR)." *Journal of Quantitative Linguistics* 17(2): 94-100.
- Guiraud, P. 1954. *Les caractères statistiques du vocabulaire*. Presses universitaires de France.
- Guiraud, P. 1959. *Problèmes et méthodes de la statistique linguistique*. D. Reidel Publishing Company.
- Howe, J. 2006. "The rise of crowdsourcing." *Wired Magazine* 14(6): 1-4.
- Iyyer, M., Enns, P., Boyd-Graber, J., & Resnik, P. 2014. "Political ideology detection using recursive neural networks." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, pp. 1113-1122.
- Johnson, W. 1944. "Studies in language behavior: A program of research." *Psychological Monographs* 56(2): 1-15.
- Kim, S. M., & Hovy, E. 2005. "Automatic detection of opinion bearing words and sentences." In *Companion Volume to the Proceedings of Conference including Posters/Demos and tutorial abstracts*. Jeju Island.
- McCarthy, P. M. 2005. An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). Doctoral dissertation, The University of Memphis.
- McCarthy, P. M., & Jarvis, S. 2010. "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment." *Behavior Research Methods* 42(2): 381-392.
- Recasens, M., Danescu-Niculescu-Mizil, C., & Jurafsky, D. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, pp. 1650-1659.
- Templin M. 1957. *Certain language skills in children*. Minneapolis: University of Minnesota Press.
- Wazny, K. 2017. "'Crowdsourcing' ten years in: A review." *Journal of Global Health*, vol 7, No 2. P. 020601.