Computer Science • 16 (2) 2015 http://dx.doi.org/10.7494/csci.2015.16.2.145

Bartłomiej Balcerzak
Wojciech Jaworski

APPLICATION OF LINGUISTIC CUES IN THE ANALYSIS OF LANGUAGE OF HATE GROUPS

Abstract *Hate speech and fringe ideologies are social phenomena that thrive on-line. Members of the political and religious fringe are able to propagate their ideas via the Internet with less effort than in traditional media. In this article, we attempt to use linguistic cues such as the occurrence of certain parts of speech in order to distinguish the language of fringe groups from strictly informative sources. The aim of this research is to provide a preliminary model for identifying deceptive materials online. Examples of these would include aggressive marketing and hate speech. For the sake of this paper, we aim to focus on the political aspect. Our research has shown that information about sentence length and the occurrence of adjectives and adverbs can provide information for the identification of differences between the language of fringe political groups and mainstream media.*

Keywords hate speech, natural language processing, propaganda, machine learning

Citation Computer Science 16 (2) 2015: 145–156

1. Introduction

In cyberspace, millions of people send out and review terabytes of data. At the same time, countless political, religious, and ideological agendas can be propagated with nearly no limitations. This, in turn, leads to an increased risk of successfully spreading various types of ideologies. Many tools can be introduced in an attempt to combat this process and promote a critical approach to information available online. With the ever-growing plethora of websites and forums, an automated method of recognizing such material seems to be a viable solution. In this paper, we propose an approach to this problem that can lead to the development of such tools. By applying methods of natural language processing, we aim to construct a classifier for identifying the language of political propaganda. For the sake of this research, we decided to focus on written text that constitutes the language of entities from the political extremes. Applying methods that concentrate on the structure of text rather than word count may allow for a more generalized method of automated text processing. In this paper, we want to use a selection of verbal cues as a method of distinguishing the language of political propaganda. These characteristics include basic information about sentence length and its structure, and will be extracted with the use of part of speech taggers already available for the English language. Unlike the bag-of-words approach (which focuses on word concurrence within a set of documents), our approach aims to focus on a deeper level of analysis, one involving cues that describe the style of the text. By using this information, we plan to construct robust classifiers that are able to identify the language of fringe elements of the political spectrum.

In order to perform our research, we collected two corpora representing the extremes of the political spectrum. One corpus includes texts gathered from a set of websites connected with American groups promoting radical nationalistic ideologies (i.e., national socialism, white and black supremacy, racism, anti-immigrant combatants). The other consists of text from websites that promote communism. We decided to split the fringe material into two corpora in order to take account of potential differences between extreme ideologies. This collection will be compared with the language of mainstream political news, both from national and local news agencies and newspapers.

If this model of linguistic cues based on part of speech occurrence provides viable results with such differing materials, this would indicate that it can also be used in more subtle scenarios. Analysis of the given material has been conducted both at the article and sentence levels. This involved the application of machine learning algorithms in order to identify sentences and articles belonging to political propaganda. Our main assumption is: since fringe ideologies are bent on changing the world in a radical way, materials that endorse such sentiments would be mostly emotional and filled with statements that evaluate the current situation as well as the desired ideal world (the end game of fully implementing the tenets of the ideology). Therefore, when speaking in linguistic terms, texts belonging to a fringe ideology corpus would have higher amounts of adjectives and adverbs than strictly-informative

sources. Hence, the following hypotheses regarding the role of each of the selected linguistic attributes have been proposed for fringe ideology detection:

Hypothesis 1. Text belonging to fringe ideologies contain more adjectives and adverbs than news texts. This hypothesis can also be rephrased as: Sentences belonging to fringe ideologies contain more adjectives and adverbs than strictly-informative texts.

We also propose an additional hypothesis regarding the average length of a sentence in both fringe and informative sources.

Hypothesis 2. Texts that endorse fringe ideologies contain longer sentences than strictly-informative sources.

In our current research, we focus on both the sentence and article levels. Extracted features will serve as attributes used by the algorithm for training. This means that standard procedures of machine-learning evaluation will be applied. We also propose a third hypothesis:

Hypothesis 3. When predicting whether a text belongs to fringe ideological or informative sources, information about the frequency of fringe sentences within provides the highest measures of performance (classification accuracy).

In order to test these hypotheses, three main tasks have been conducted. Firstly, machine-learning techniques were used to identify whether articles from a collected corpus belong to either the fringe or informative class. The second task would be similar, but rather focused on making predictions for sentences in the corpus. Thirdly, predictions for articles were conducted based on the frequency of propaganda sentences in an article. After the performance of algorithms used in each task is evaluated, attribute-extraction methods are used in order to discover which attributes are crucial for determining whether the source is fringe or informational.

2. Related work

Various fields of natural language processing can be connected with the problem of identifying the language of political fringe. One of these fields would be text genre detection, where NLP tools are used in order to identify the genre of a given text. Such tools involve a bag-of-words approach as well as part of speech n-grams as used in [17]. Other applications of genre classification include the use of both linguistic cues and html structure of the web page [16].

In our work, however we focus not only on detecting text genres, but also on identifying a very specific type of narration, which can be classified as manipulative or deceptive.

This leads to a second field of study that can be applied to the study of materials from the political fringe. Analysis of such sources on the Internet had been present in the field of computer science for some time. The research done so far can be divided into three main areas of investigation. First of them can be described as studies into behavioral patterns of propagandists on-line. This approach focuses mostly on

the agent distributing the content, rather than the content itself, it is very strongly inspired by research into spam detection [10]. Such research was mostly focused on micro-blogging platforms such as Twitter or open source projects such as Wikipedia [3]. Wikipedia was also a topic of research into diagnosing controversial content, and for improving team collaboration by diagnosing conflict among editors [18, 19].

What is notable is the fact that in these papers the main source of traits that could identify propaganda are mostly meta information, referring to frequency of commenting on a material, and presenting the same content repeatedly. The other main type of research is the field of deception detection which deals with the problem of distinguishing whether the author of a text intended to fool the recipient. Researchers working in this field wanted to identify linguistic cues related with deceptive behavior. Most of these cues such as the ones used by [12] in experimental setups or [8] for fraudulent financial claims emphasized basic shallow characteristics such as sentence or noun phrase length that showed to be indicative of deceptive materials.

A more sophisticated method was proposed by [4]. In their paper they introduced a concept of stylometry (analyzing larger syntactical structures) for deception detection and opinion spam. Their work focuses on the structure of syntax, the relation between larger elements of the sentence (ie. Noun phrases).

A slightly different approach was tested by [11] in their essay experiment. With the aid of LIWC tool [14] they aimed to identify words that appear most often in deceptive materials.

In our work we want to extend and test the findings of deception detection for the task of identifying textual materials forming the bulk of fringe ideologies.

It is also worth noting that most of the research in this field was done for the English language. The only instance of deception detection in another language that we managed to find was study [5] that was done for Italian.

3. Hypothesis verification

3.1. Dataset

In order to test our hypothesis, we gathered a corpus consisting of both radical ideological material and balanced informative text. All of the corpora represent modern American English.

We prepared a collection of texts taken from three distinguished sources:

- **Nazi corpus**: This contains articles from websites belonging to groups in the US that promote national socialism and ideas of racial and ethnic superiority. When collecting these websites, we used a list of hate groups provided by the Southern Poverty Law Center. Noticeable sources include the American National Socialist Party, National Socialist movement, Aryan Nations, etc. In all, 100 web pages were extracted. An example of a text from this corpus is shown below:
  To a National Socialist, things like PRIDE, HONOR, LOYALTY, COURAGE, DISCIPLINE, and MORALITY – actually MEAN something. Like our fore-

fathers, we too are willing to SACRIFICE to build a better world for our children whom we love deeply, and like them as well – we are willing to DO ANYTHING NECESSARY to ACHIEVE THAT GOAL. We are your brothers and your sisters, we are your fathers and your mothers, your friends and your co-workers – WE ARE WHITE AMERICA, just like YOU! Your enemies in control attempt with their constant "anti-nazi" propaganda – to persuade you that "they" are your "real friends" – and that WE, your own kin, are your lifelong enemies. Yet, ask yourself THIS – "WHO" have you to "THANK" for ALL the PROBLEMS FACING YOU? The creatures who HAVE been in total CONTROL – or – those of us resisting the evil? The TRUTH is right there in front of you – DON'T be afraid to understand it and to ACT upon it! You hold the FUTURE – BRIGHT or DARK – in YOUR hands, and TIME IS RUNNING OUT!

- **Communist Corpus**: This corpus contains pages from websites of American groups and parties that describe themselves as communist. These include among others: the Progressive Labour Party and the Communist Party of the United States of America. As with the Nazi corpus, 100 web pages were extracted. An example is presented below:
  Today's action, organized by Good Jobs Nation, comes a year after it filed a complaint with the Department of Labor that accused food franchises at federal buildings of violating minimum-wage and overtime laws. They want Obama to sign an executive order requiring federal agencies to contract only with companies that engage in collective bargaining.
  The union leaders pulling the strings behind Good Jobs Nation are the same people who got us into this mess in the first place. Most contract jobs used to be full-time union jobs, and the unions did nothing to stop the bosses from eliminating them. Now the unions are trying to rebuild their ranks among low-wage workers who replaced their former members. We need to abolish wage slavery with communist revolution. And the struggle between reform and revolution must be waged within struggles like this one.
- **News Corpus**: This corpus will serve as a reference point. It contains political news and opinion sections from various American news sources (CNN, FOX news, CNBC, and local media outlets), in order to construct a balanced corpus for analysis, 100 web pages were extracted as well for the task of training a machine-learning algorithm.

3.2. Dataset preprocessing

After the working corpus had been prepared, we divided each sentence into tokens and conducted part of speech tagging (with the use of NLTK default POS tagger).

Sentence and word length have been calculated as well; hence, creating a database containing the following attributes of all sentences:

- **Sentence Class**: a variable determining, whether the sentence belongs to the informative or fringe corpus.

- **Traits describing textual quantity**: number of tokens in a sentence and average number of characters in words constituting a sentence.
- **Frequency of adjectives and adverbs**.

We also included the occurrence of other parts of speech in order to test if they have any impact on distinguishing fringe language from informative sources.

After the database was prepared, machine-learning algorithms were implemented. In order to evaluate the performance of said algorithms in both of the tasks, the following measures were used:

- Precision: Ratio of true positives to all positives.
- Recall: Ratio of true positives to the sum of true positives and false negatives
- Accuracy: the percentage of all true positives and true negatives produced by the machine-learning algorithm.
- AUC (Area under ROC Curve): the measure of True Positive to False Positive ratio for various thresholds provided by the machine-learning algorithm.

3.3. Machine learning:

The task given to the algorithms was to identify whether a sentence or article belongs to either the fringe or news corpus. A 5-fold cross-validation procedure was implemented in order to validate the performance of the algorithms.

3.4. Article prediction

As shown in Table 1 containing the standardized differences between the mean values of these attributes, the frequency of adjectives was higher in both fringe corpora in relation to the informative sources. In the case of adverbs, the effect was visible only for Nazi materials. What is also interesting is this: articles belonging to fringe ideologies contained fewer proper nouns than informative sources. This may be because ideological materials tend to be more vague and connected with more general processes and entities. However, such a hypothesis needs to be verified separately.

Table 1

Standardized differences between mean values for adjective, adverb, and proper noun frequency.

Attributes %	Nazi/News	Communist/News
Adjectives	0,24	0,39
Adverbs	0,2	−0,04
Proper Nouns	−0,38	−0,33

In this section, we analyze the performance of article prediction based on the frequency of parts of speech. Since our corpora are balanced, the baseline measuring performance is 50%, both for accuracy and AUC. Three algorithms turned out to be the most effective, these being Naive Bayes, K-NN (K = 7, cosine distance), and Neural Net.

Table 2

Algorithm performance – accuracy and ROC AUC.

Algorithms applied	**Acc. Nazi** [%]	**AUC Nazi** [%]	**Acc. Comm** [%]	**AUC Comm** [%]
Naive Bayes	70	80	75	81
K-NN	60	64	59	61
Neural Net	69	72	77	82

Table 3

Algorithm performance – precision and recall.

Algorithms	**P Nazi/News** [%/%]	**R Nazi/News** [%/%]	**P Com/News** [%/%]	**R Com/News** [%/%]
Naive Bayes	81/65	52/88	89/69	57/93
K-NN	60/60	62/58	58/61	65/53
Neural Net	68/75	77/64	81/74	71/83

As shown in Table 2, the best scores were produced by the Naive Bayes and Neural Net Algorithms. The accuracy and AUC values for these were within the range of 70 to 80 percent, which indicates a moderately-strong performance. It can also be pointed out that higher scores were achieved for the communist corpus. This may be, in part, due to the fact that communist articles came mainly from two large sources. Still, both types of fringe ideologies achieved similar results. Both precision and recall scores as well as accuracy and AUC painted a similar picture (see Table 3).

3.5. Sentence prediction

As shown in Tables 4 and 5, the results tend to be between the threshold of 60 and 70%. No larger differences have been observed when analyzing the Nazi and communist corpora. It is also worth noting that classification conducted at the sentence level provided weaker scores than the one done for the articles. This may be due to the fact that sentences are shorter, therefore providing less data that the machine-learning algorithms can use.

Table 4

Algorithm performance – accuracy and ROC AUC.

Algorithms	**Acc. nazi** [%]	**AUC nazi** [%]	**Acc. communist** [%]	**AUC communist** [%]
Naive Bayes	61	65	61	65
k-NN	62	69	64	64
Neural Net	65	72	63	69

Similar to the article prediction task, this is also complemented by attribute-extraction scores from chi-square method. According to chi-squared method, the

most important attributes include: sentence length, adjectives, adverbs, and proper noun frequency. The relative importance of the attributes used is shown in Table 6, containing standardized differences between the mean values. Only the attributes with the highest differences are presented in this table. Positive values indicate that the mean value of a given attribute was higher for the fringe corpus. Negatives indicate that the value was higher for the news corpus.

Table 5

Algorithm performance – accuracy and ROC AUC.

Algorithms	**P Nazi/News** [%/%]	**R Nazi/News** [%/%]	**P Com/News** [%/%]	**R Com/News** [%/%]
Naive Bayes	58/66	64/60	60/64	64/58
k-NN	62/63	63/63	63/66	63/66
Neural Net	64/67	59/72	66/62	52/75

Table 6

Standardized differences between mean values for adjective, adverb, noun, and proper noun frequency.

Attributes		**Nazi/News**	**Communist/News**
Sentence length		0,12	−0,14
Adjectives	[%]	0,3	0,3
Adverbs	[%]	0,2	0,04
Nouns	[%]	−0,18	0,36
Proper Nouns	[%]	−0,32	−0,4

The table shows a pattern similar to that observed when classifying articles. Fringe ideology sentences, both communist and Nazi, contained a higher number of adjectives and adverbs. They also had fewer proper nouns than informative sources. What sets sentence classification apart from article classification is the relatively high difference in the frequency of nouns between communist material and informative sources, as well as the fact that both corpora varied in regards to average sentence length. Nazi sources tend to contain longer sentences, while the communist collection has shorter ones. In both cases, the differences are not as pronounced as in the other attributes.

3.6. Sentence-based article prediction

For this task, we decided to use the labels applied to sentences in the previous subsection in order to predict whether the entire article belonged to the fringe or informative class. The frequency of fringe sentences, as provided by the Neural Net classifier, has been calculated for each article from the corpora used in our research. Afterwards, we calculated the optimal threshold for classification. For Nazi materials, the threshold value was 79% of fringe sentences per article, and for communists, this was 74%.

The performance scores for these corpora were calculated on a separate collection of news and fringe material that was not previously used in training. The results are shown in Tables 7 and 8.

Table 7

Performance – Accuracy and RoC AUC.

Positive class	Accuracy	AUC
Nazi	78%	82%
Communist	80%	83%

Table 8

Performance – precision and recall.

Positive class	Precision	Recall
Nazi	75%/85%	88%/70%
Communist	76%/86%	89%/71%

When compared to the previous scores of machine-learning methods, the one used in this subsection provides higher accuracy with a relatively similar level of AUC. It is also worth noting that the performance of this method is high even though the scores for sentence prediction rarely exceeded the 70% threshold for either accuracy or AUC. This observation indicates that, even though prediction on a sentence level can be faulty, it can provide a strong signal when aggregated, thus allowing for the more-successful identification of fringe sources. What is more, this method proves to be slightly more efficient on the article level than using part of speech information. Further work on the possibilities and limits of this approach will be pursued at a later date.

4. Conclusions and observations

Our research has led us to the following conclusions and observations:

- Article classification based on part of speech tagging has provided robust scores, indicating that the chosen attributes can be used for identifying fringe ideological sources. With a baseline of 50% for both accuracy and AUC, the machine-learning algorithms achieved scores exceeding 70% for accuracy, and 80% for AUC. The Neural Net and K Nearest Neighbors algorithms produced the best performance.
- The same task conducted at the sentence level provided weaker scores, mostly within the 60%–70% range both for accuracy and AUC. However, when the information about the frequency of fringe sentences was applied to articles, performance increased to 80% for accuracy and AUC. Consequently, these score provide a more-robust classification than the one based on part of speech frequency at the article level. This may indicate that sentence-level classification is burdened with high noise, which may be countered by taking into account the fringe-sentence frequency in the article.

- In view of the collected-data hypothesis, 2 cannot be considered true. Sentence length proved not to be an important attribute for determining whether the sentence belongs to a fringe or informative source. The texts from the Nazi corpus contained longer sentences on average. Conversely, communist sources where constructed with shorter ones. However, neither produced differences large enough to affect the machine-learning algorithms performance.
- Analysis on both the article and sentence levels shows that adjective and adverb frequency plays an important role in identifying fringe sources. Moreover, the collected data allows us to consider hypotheses 1 and 3 as verified.
- Analysis has also shown that fringe materials contain fewer named entities than informative sources. This may suggest that fringe text tends to be more vague than that which is solely informative.

In summary, our research lends credence to the notion that such basic linguistic cues as the occurrence of parts of speech can be used for determining whether a text belongs to the language of information or fringe ideology.

5. Future work

The proposed system for detecting hate language in public discourse could be implemented in a centralized manner. However, for many applications such as Twitter or public fora, a Peer-to-Peer system would be more scalable [1, 20].

The results we came up with show that there is room for further research. The use of basic stylistic cues for identifying specific narrations may be extended to include language of political, religious, or scientific discourse.

We also plan to include other elements of propaganda in our model, such as repetition and vagueness [7, 15].

We plan to conduct a more-detailed analysis of the two observations we made when researching the role of linguistic features: the differences in named entity frequency in fringe and informative sources, and the frequency of positive class sentences as a method of article prediction. Devising a computational model may lead to the development of tools dedicated to automatic detection of specific forms of languages. So far, we worked with modern documents written in English that are focused mostly on one ideology. This is why we aim to extend our focus to historical texts written in different languages in our future work. This is also important in order to identify linguistic cues that can be applied only to fringe ideologies. To this end, we plan to apply the developed model for other fringe ideologies and pseudoscience.

In summary, our current work serves as preliminary research into a field of computational analysis of discourse. Therefore, we can obtain an effective model that can be extrapolated to various topical domains. This will prove useful as a way of constructing a more-general theory of computational models of recognizing language of fringe ideologies (as well as other forms of written language, such as language of religion, science, marketing, or various social classes). This will be especially important

for a classification method based on fringe-sentence frequency in the article. Focusing on structural aspects would also make it more resistant to deceptive strategies on the part of content producers.

Acknowledgements

This work was financially supported by the European Community from the European Social Fund within the INTERKADRA project.

It was also supported by the grant Reconcile: Robust Online Credibility Evaluation of Web Content from Switzerland through the Swiss Contribution to the enlarged European Union.

References

[1] Barkai D.: *Peer-to-Peer Computing: technologies for sharing and collaborating on the net*. Intel Press, Santa Clara, USA, 2001.

[2] Bird S., Klein E., Loper E.: *Natural language processing with Python*. O'Reilly Media, Baijing [etc.], 2009.

[3] Chandy R.: Searching Wikiganda: Identifying Propaganda Through Text Analysis. *Caltech Undergraduate Research Journal*, vol. 9(1), pp. 10–15, 2008.

[4] Feng S., Banerjee R., Choi Y.: Syntactic stylometry for deception detection. In: *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers, vol. 2*, pp. 171–175. Association for Computational Linguistics, 2012.

[5] Fornaciari T., Poesio M.: Lexical vs. surface features in deceptive language analysis. In: *Proceedings of the ICAIL 2011 Workshop: Applying Human Language Technology to the Law*, pp. 2–8, 2011.

[6] Gîfu D., Cristea D.: Towards an Automated Semiotic Analysis of the Romanian Political Discourse. *Computer Science Journal of Moldova*, vol. 21(1), pp. 36–64, 2013.

[7] Gîfu D., Dima I. C.: An operational approach of communicational propaganda. *International Letters of Social and Humanistic Sciences*, vol. 23, pp. 29–38, 2014.

[8] Humpherys S. L., Moffitt K. C., Burns M. B., Burgoon J. K., Felix W. F.: Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems*, vol. 50(3), pp. 585–594, 2011.

[9] Lee A. M., Lee E. B.: *The fine art of propaganda*. Octagon Press, Limited, London, UK, 1972.

[10] Metaxas P.T.: Web spam, social propaganda and the evolution of search engine rankings. In: *Web Information Systems and Technologies*, pp. 170–182. Springer, 2010.

[11] Mihalcea R., Strapparava C.: The lie detector: Explorations in the automatic recognition of deceptive language. In: *Proceedings of the ACL-IJCNLP 2009*

Conference Short Papers, pp. 309–312. Association for Computational Linguistics, 2009.

[12] Ott M., Choi Y., Cardie C., Hancock J. T.: Finding deceptive opinion spam by any stretch of the imagination. In: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1*, pp. 309–319. Association for Computational Linguistics, 2011.

[13] Paik W., Yilmazel S., Brown E., Poulin M., Dubon S., Amice C.: Applying natural language processing (nlp) based metadata extraction to automatically acquire user preferences. In: *Proceedings of the 1st international conference on Knowledge capture*, pp. 116–122, ACM, 2001.

[14] Pennebaker J. W., Francis M. E., Booth R. J.: Linguistic Inquiry and Word Count: LIWC 2001. Erlaub Publishers, Mahwah, NJ, 2001.

[15] Philippe L. L., Nancy J. L.: *Le mythe nazi*. Editions de l'Aube, La Tour d'Aigues, France.

[16] Santini M., Power R., Evan R.: Implementing a characterization of genre for automatic genre identification of web pages. In: *Proceedings of the COLING/ACL on Main conference poster sessions*, pp. 699–706. Association for Computational Linguistics, 2006.

[17] Sharoff S.: Classifying Web corpora into domain and genre using automatic feature identification. In: *Proceedings of the 3rd Web as Corpus Workshop*, pp. 83–94. 2007.

[18] Turek P., Wierzbicki A., Nielek R., Datta A.: WikiTeams: How Do They Achieve Success? In: *Potentials IEEE*, vol. 30(5), pp. 15–20, 2011.

[19] Turek P., Wierzbicki A., Nielek R., Hupa A., Datta A.: Learning about the quality of teamwork from wikiteams. In: *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pp. 17–24, IEEE, 2010.

[20] Wierzbicki A., Szczepaniak R., Buszka M.: Application layer multicast for efficient peer-to-peer applications. In: *Internet Applications. WIAPP 2003. Proceedings. The Third IEEE Workshop on*, pp. 126–130, IEEE, 2003.

Affiliations

Bartłomiej Balcerzak
Polish-Japanese Institute of Information Technology, Warsaw, Poland,
bartlomiej.balcerzak@pjwstk.edu.pl

Wojciech Jaworski
Polish-Japanese Institute of Information Technology, Warsaw, Poland;
Institute of Informatics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland
wjaworski@mimuw.edu.pl

Received: 19.01.2015
Revised: 19.03.2015
Accepted: 20.03.2015