Proppy: A System to Unmask Propaganda in Online News

Alberto Barrón-Cedeño,[1] **Giovanni Da San Martino,**[1] **Israa Jaradat,**[2] **Preslav Nakov**[1]

[1]Qatar Computing Research Institute, HBKU, Qatar
[2]University of Texas Arlington, USA
{albarron, gmartino, pnakov}@hbku.edu.qa,
israa.jaradat@mavs.uta.edu

Abstract

We present `proppy`, the first publicly available real-world, real-time propaganda detection system for online news, which aims at raising awareness, thus potentially limiting the impact of propaganda and helping fight disinformation. The system constantly monitors a number of news sources, deduplicates and clusters the news into events, and organizes the articles about an event on the basis of the likelihood that they contain propagandistic content. The system is trained on known propaganda sources using a variety of stylistic features. The evaluation results on a standard dataset show state-of-the-art results for propaganda detection.

Introduction

Propaganda is the expression of an opinion or an action by individuals or groups deliberately designed to influence the opinions or the actions of other individuals or groups with reference to predetermined ends (Institute for Propaganda Analysis 1938). We are interested in propaganda from a journalistic point of view: how news management lacking neutrality shapes information by emphasizing positive or negative aspects purposefully (Jowett and O'Donnell 2012, p. 1). Propaganda uses psychological and rhetorical techniques that are intended to go unnoticed to achieve maximum effect. As a result, malicious propaganda news outlets have proven to be able to achieve large-scale impact. In particular, the power of disinformation and propaganda was arguably demonstrated during recent events, such as Brexit and the 2016 U.S. Presidential campaign.[1]

With the rise of the Web, a combination of freedom of expression and ease of publishing contents online has nurtured a number of news outlets that produce or distribute propagandistic content. Social media further amplified the problem by making it possible to reach millions of users almost instantaneously. Thus, with the aim of helping fight the rise of propaganda, here we introduce `proppy`, a system to unmask articles with propagandistic content, which can (*i*) help investigative journalists to study propaganda online and (*ii*) raise awareness that a news article, or a news outlet in general, might be trying to influence people's mindset.
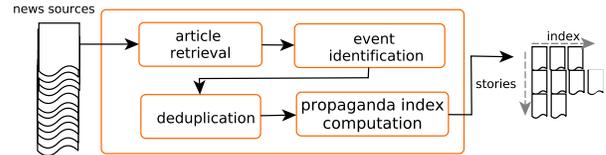
[1]https://www.justice.gov/file/1035477/

Figure 1: The architecture of `proppy`.

To the best of our knowledge, `proppy`[2] is the first publicly available real-world, real-time monitoring and propaganda detection system for online news, which aims at raising awareness about propaganda.

Architecture of the System

Figure 1 shows the architecture of `proppy`. We describe its four modules next.

1. Article retrieval. `Proppy` regularly monitors a variety of news outlets and extracts the content of the latest news articles from their websites. We use GDELT[3] to obtain links to news articles and the Newspaper3k Python library[4] to extract their content. `Proppy` then analyzes the articles in batches every 24 hours, and performs the remaining three steps.

2. Event identification. We use the DBSCAN clustering algorithm (Ester et al. 1996) for event identification, as it does not require information related to the expected number of events. We use *doc2vec* embeddings (Le and Mikolov 2014) for article representation, pre-trained on articles from Associated Press. We compute the pairwise distances for DBSCAN as 1 minus the cosine similarity. DBSCAN has two hyper-parameters: the minimum number of members in a cluster and the maximum distance between two members of the same cluster, ϵ. We set the former parameter to 2, thus discarding singletons. We estimate the parameter $\epsilon = 0.55$ on the METER corpus (Clough, Gaizauskas, and Piao 2002).

[2]Visit the `proppy` project at http://proppy.qcri.org
[3]http://gdeltproject.org
[4]https://newspaper.readthedocs.io

3. Deduplication. Next, we discard near-duplicates using a standard text re-use technique: comparison of word n-grams (Lyon, Barret, and Malcolm 2004) after standard pre-processing (case-folding, tokenization, and stopword removal). We compute the similarity between all pairs of documents in a cluster using the Jaccard coefficient. Once again, we use the METER corpus to optimize the value of n and the threshold to consider two documents as near-duplicates. At run time, we discard all near-duplicates but one.

4. Propaganda Index Computation. We train a maximum entropy classifier with L2 regularization to discriminate propagandistic vs non-propagandistic news articles. We use the confidence of the classifier, a value in the range $[0, 1]$, to group articles into bins. We call this value the *propaganda index*, since it reflects the probability for an article to have a propagandistic intent. We use four families of features:

Word n-gram features We use *tf.idf*-weighted word $[1, 3]$-grams (Rashkin et al. 2017).

Lexicon features. We try to capture the *typical vocabulary* of propaganda by considering representations reflecting the frequency of specific words from a number of lexicons coming from the Wiktionary, Linguistic Inquiry and Word Count (LIWC), Wilson's subjectives, Hyland hedges, and Hooper's assertives. Rashkin et al. (2017) showed that the words in some of these lexicons appear more frequently in propagandistic than in trustworthy articles.

Style, vocabulary richness, and readability. Our writing style representation consists of *tf.idf*-weighted character 3-grams. This representation captures different style markers, such as prefixes, suffixes, and punctuation marks. We further consider the *type-token ratio* (TTR) as well as the number of tokens appearing exactly once or twice in the document: *hapax legomena* and *dislegomena*. Moreover, we combine types, tokens, and hapax legomenæ to compute Honore's R and Yule's characteristic K. We also use three readability features originally designed to estimate the level of complexity of a text: *Flesch–Kincaid grade level*, *Flesch reading ease* and the *Gunning fog index*.

NELA. We also integrate the NEws LAndscape (NELA) features (Horne, Khedr, and Adalı 2018): 130 content-based features collected from the existing literature that measure different aspects of a news article (e.g., sentiment, bias, morality, complexity).

We evaluated proppy on data from Rashkin et al. (2017) in a binary setup of distinguishing propaganda vs non-propaganda. Their n-grams system yielded an F_1 of 88.21, whereas proppy achieved 96.72 (+8.51), a statistically significant improvement (measured with the McNemar test).

Online Interface

Figure 2 shows a screenshot of proppy. The architecture follows a *push* publishing model: it updates automatically the material that it presents to the user without her taking any action but exploring the available events. The left panel shows events from the last 24 hours. When the user clicks on an event, its articles are shown in the right panel, organized into five bins according to their propaganda index.
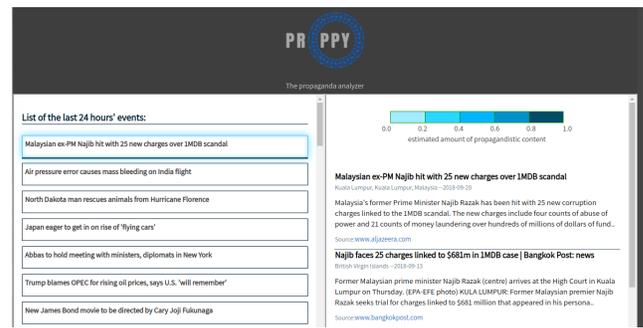

Figure 2: A screenshot of the online proppy.

The articles in bins 1 and 2 are considered nearly non-propagandistic, whereas those in the two right bins are propagandistic. In this way, the user can easily observe how different media cover related events on the propaganda dimension and may guide her further exploration and judgment.

Conclusion and Future Work

We have presented proppy, a publicly available real-world, real-time propaganda detection system for online news, which aims at raising awareness with the objective of limiting the impact of propaganda and helping fight disinformation. In future work, we plan to add support for multiple languages, and a *pull* mode where users will be able to submit any article and get its propaganda index.

References

Clough, P.; Gaizauskas, R.; and Piao, S. 2002. Building and annotating a corpus for the study of journalistic text reuse. In *LREC*, 1678–1691.

Ester, M.; Kriegel, H.-P.; Sander, J.; and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proc. of KDD*, 226–231.

Horne, B.; Khedr, S.; and Adalı, S. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Proc. of ICWSM*, 518–527.

Institute for Propaganda Analysis. 1938. How to Detect Propaganda. In *Propaganda Analysis. Vol. I of the Publications of the Institute for Propaganda Analysis*. chapter 2.

Jowett, G., and O'Donnell, V. 2012. *Propaganda and Persuasion*. Los Angeles, CA: SAGE, 5th edition.

Le, Q., and Mikolov, T. 2014. Distributed Representations of Sentences and Documents. In *Proc. of ICML*, 1188–1196.

Lyon, C.; Barret, R.; and Malcolm, J. 2004. A Theoretical Basis to the Automated Detection of Copying Between Texts, and its Practical Implementation in the Ferret Plagiarism and Collusion Detector. In *Plagiarism: Prevention, Practice and Policies Conference*. Plagiarism Advice.

Rashkin, H.; Choi, E.; Jang, J. Y.; Volkova, S.; and Choi, Y. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proc. of EMNLP*, 2931–2937.