Automatic Identification of Propaganda in Text

Alberto Barrón-Cedeño
DIT, Università di Bologna, Forlì, Italy
a.barron@unibo.it

Posting contents online is at everybody's fingertips. As a result any person can distribute information to influence in the people's perception. The so-called *fake news* and, in more general terms, misinformation have become trendy terms as they have influenced in events as important as the elections in EEUU, Brasil, and other countries. This kind of information is transmitted to audiences through propaganda [1]. We have developed supervised models to identify whether a news article contains high levels of propaganda [2]. Nowadays, we focus on the development of models to identify specific fragments containing propaganda techniques.

In this poster I will introduce "proppy", our document-level system. Proppy is a logistic regressor trained upon style representations of the texts. The hypothesis behind such representations is that there is an inherent "style" attached to propaganda, which is independent of topic or bias (e.g., left or right). I will also discuss Proppy v2.0, which intends to identify up to 18 propaganda techniques at the fragment level (even if the output quality remains far from perfect). Proppy v2.0 is based on a recurrent neuronal network with BERT representations [3].

Proppy has been integrated in www.tanbih.org, a news aggregator that pays special attention to propaganda and bias in diverse English- and Arabic-speaking media.

References

[1] How to Detect Propaganda, in: Institute for Propaganda Analysis (Ed.), Propaganda Analysis. Volume I of the Publications of the Institute for Propaganda Analysis, New York, NY, 1938, Ch. 2, pp. 210–218.

[2] Barrón-Cedeño, Da San Martino, Jaradat, and Nakov, "Proppy: Organizing News Coverage on the Basis of Their Propagandistic Content," Information Processing and Management, 2019

[3] Devlin, Chang, Lee, Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805