Fake news: an algorithmic approach

Martijn Schouten

April 24, 2019

1 Personal details

My email `mailto:martijn.schouten@student.uva.nl`

My supervisors email `mailto:maartenmarx@uva.nl`

The wiki on my GitHub account `https://github.com/MeMartijn/BachelorThesis/wiki`

2 Research question

The following research question is defined: *how well can state-of-the-art natural language processing techniques in combination with machine learning algorithms classify fake news?*

For this research question, the following subquestions will be answered:

- RQ1: How can fake news be defined and characterized?

- RQ2: What new ways of word- or sentence embeddings can be used for encoding plain text?

- RQ3: What is the performance of combinations of these novel embedding techniques with machine learning algorithms?

- RQ4: To what extent can performance of fake news classifiers be improved with increased amounts of raw data?

3 Related Work

3.1 RQ1

Fake news as a term only caught public attention starting from the end of 2016, during the Presidential Elections of the United States [7].

3.2 RQ2

In the last couple of years, using transfer learning for natural language processing has given promisable results. The following sentence embeddings will be used to detect fake news:

- Bag of Words as a baseline for performance of non-pretrained embeddings;

- Facebook's InferSent [1];

- ELMo from the Allen Institute for Artificial Intelligence [5];

- OpenAI's GPT-2 [6];

- Transformer-XL [2];

- Microsoft's MT-DNN [4];

- and Google's BERT [3].

3.3   RQ3

Aligned with the original research on this dataset by Wang [8], the following machine learning algorithms will be used to test the applicability of the abovementioned embedding techniques:

- SVMs;

- Logistic regression;

- Bi-LSTMs;

- CNNs.

References

[1] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.

[2] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime G. Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. Transformer-xl: Attentive language models beyond a fixed-length context. *CoRR*, abs/1901.02860, 2019.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. Multi-task deep neural networks for natural language understanding. *arXiv preprint arXiv:1901.11504*, 2019.

[5] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *CoRR*, abs/1802.05365, 2018.

[6] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[7] Google Trends. Explore. `https://trends.google.nl/trends/explore?date=today%205-y&q=fake%20news`, 2019. Retrieved on 16th of April, 2019.

[8] William Yang Wang. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *CoRR*, abs/1705.00648, 2017.