

hyperpartisian detection

A notebook

José Eleandro Custódio, Caio, Arthur, Leonardo, Rafael Sandroini, and Ivandré Paraboni

School of Arts, Sciences and Humanities (EACH)
University of São Paulo (USP)
São Paulo, Brazil
{eleandro, ivandre}@usp.br

Abstract We present an ensemble approach to cross-domain authorship attribution that combines predictions made by three independent classifiers, namely, standard char n-grams, char n-grams with non-diacritic obfuscation and word n-grams. Our proposal relies on variable-length n-gram models and multinomial logistic regression, and selects the prediction of highest probability among the three models as the output for the task. Results generally outperform the PAN-2018 baseline system that makes use of fixed-length char n-grams and linear SVM classification.

1 Introduction

O termo *Fake News* tornou-se amplamente conhecido após as eleições americanas de 2016 devido a alegações feitas pelo então candidato Donald Trump. Uma das principais alegações é de que diversos meios de comunicação publicam notícias tendenciosas, e algumas vezes falsas, de modo a atrair um número maior de visualizações e consequentemente mais rentabilidade para suas propagandas. Uma característica compartilhada por essas publicações é o nível de polarização política.

A detecção de tendência ideológica é uma tarefa difícil mesmo para humanos [4]

The rest of this paper is structured as follows. Section 3 describes our main approach. Section 4 presents our results and those provided by the relevant baseline method over the PAN-CLEF-2018 AA dataset. Finally, Section 5 discusses these results and suggests future work.

2 Related works

o trabalho em [6] estudou o engajamento político considerando classificando os textos em graus que variavam entre extrema esquerda até extrema direita. Foram coletados dados de usuários do twitter que receberam três dólares pela pesquisa. Estes usuários preencheram um questionário de mapeamento e traços psicológicos e sociais, informaram suas preferencias políticas e forneceram 3200 tweets cada. Além desses, foram extraídos tweets de políticos de diversos partidos cuja tendência partidária é conhecida. A extração de características foi feita usando Linguistic Inquiry and Word Count

(LIWC), tópicos do Word2Vec (cluster de word2vec), análise de sentimento, anotações linguísticas e entidade nomeadas. Cada um dos sete graus foram avaliados 2 a 2 e os resultados comentados. Os melhores resultados foram obtidos com clusteres word2vec.

O trabalho em [4] estudou a detecção de tendência política no nível sentença. Foram considerados conjuntos de dados com discursos do congresso americano e livros com conteúdo ideológico reconhecido. O experimento proposto analisou redes neurais recursivas (RecNN) usando inicialização aleatória e inicialização com word2vec. Os resultados foram comparados com métodos de baseline usando regressão logística e com bag-of-words. Os melhores resultados foram obtidos usando RecNN combinado com Word2Vec. Além disso, foi reportado também que Word2Vec obteve resultado superior ao Bag-of-Words.

O trabalho em [2] estudou a influência dos *bots* no referendo sobre o *brexit*. Esses bots apresentavam um comportamento hiperpartidário e atuavam no Facebook, Twitter e em diversos site de notícias. Foi observado que as contas associadas a tais robos foram excluídas após a votação e mais de 50% dos links publicados haviam sido removidos ou apontavam para endereços inexistentes.

O trabalho em [3] usou análise de sentimento para monitorar as preferências políticas no Twitter e tentar prever o resultado das eleições.

O trabalho em [9] usou modelo de tópicos para prever o tipo dos comentários obtidos por publicações políticas feitas em blogs.

O trabalho em [8] compilou um conjunto de dados para detecção de fake news compostos por 12,8 mil textos curtos extraídos do site WWW.POLITIFACT.COM. Este site compilou afirmações feitas por políticos norte-americanos em varios de comunicação, como TV, Facebook, Twitter e outros, e abrangendo os tópicos sobre saúde, imigração, educação e outros. Cada registro desse site foi anotado por especialistas com etiquetas variando entre "totalmente falso" a "certamente verdadeiro".

O conjunto de dados apresentado foi avaliado com métodos tradicionais e com um método proposto. Os métodos tradicionais considerados foram regressão logística com regularização L2, SVM e o baseline de classe majoritária. O método proposto avaliou uma combinação da rede neural CNN com Bi-LSTM. Foi utilizado Word Embeddings Google e meta-dados contendo o tópico, a profissão do autor, o autor, o partido e outros. O método usando apenas rede CNN obteve a melhor acurácia entre os métodos. A utilização do meta-dado contendo o autor do texto combinado com a CNN obteve o melhor resultado geral desse estudo. A rede Bi-LSTM apresentou a pior performance devido ao sobreajuste.

O trabalho em [1] estudou métodos para avaliar o nível de confiança de um notícias de modo a priorizar uma verificação de fatos. Foram utilizados artigos de diversos jornais e revistas, bem como classificação feitas por organizações não governamentais. Cada artigo foi classificado no intervalo de 1 a 9 de confiança e extraídos características como análise de sentimento, TD-IDF, Entidade Nomeadas, tempos verbais, word embeddings, análise de tópicos, características linguísticas, metadados contidos no texto e outros. Os embeddings obteve o melhor resultado individual seguido das características linguísticas. O pior resultado foi obtido através de análise de tópicos.

O trabalho em [5] analisou fake news em notícias de jornais e blogs e comparou TF-IDF de bigramas de palavras com gramáticas de livre de contexto probabilísticas (PCFG). Sendo o melhor resultado obtido com TF-IDF.

O trabalho em [7] utilizou extração de entidades nomeadas e de fatos e suas relações para detecção de

3 Method

4 Results

5 Final remarks

Acknowledgements. The second author received support by FAPESP grant nro. 2016/14223-0.

References

1. A context-aware approach for detecting worth-checking claims in political debates. International Conference Recent Advances in Natural Language Processing, RANLP **2017-Septe**, 267–276 (2017). <https://doi.org/10.26615/978-954-452-049-6-037>
2. Bastos, M.T., Mercea, D.: The Brexit Botnet and User-Generated Hyperpartisan News. Social Science Computer Review (2017). <https://doi.org/10.1177/0894439317734157>
3. Bermingham, A., Smeaton, A.F.: On Using Twitter to Monitor Political Sentiment and Predict Election Results. Proceedings of the Workshop on Sentiment Analysis Where AI Meets Psychology (SAAIP) 2011, Thailand pp. 2–10 (2011)
4. Fallis, A.: Political Ideology Detection Using Recursive Neural Networks. Journal of Chemical Information and Modeling **53**(9), 1689–1699 (2013). <https://doi.org/10.1017/CBO9781107415324.004>
5. Gilda, S.: Evaluating machine learning algorithms for fake news detection. In: IEEE Student Conference on Research and Development: Inspiring Technology for Humanity, SCORED 2017 - Proceedings. vol. 2018-January, pp. 110–115 (2018). <https://doi.org/10.1109/SCORED.2017.8305411>
6. Preoțiuc-Pietro, D., Liu, Y., Hopkins, D., Ungar, L.: Beyond Binary Labels: Political Ideology Prediction of Twitter Users. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) pp. 729–740 (2017). <https://doi.org/10.18653/v1/P17-1068>, <http://aclweb.org/anthology/P17-1068>
7. Sharonova, N., Doroshenko, A., Cherednichenko, O.: Issues of fact-based information analysis. In: Vysotska V. Lytvyn V., S.N.H.T.J.K.I.G.N.K.S.A. (ed.) CEUR Workshop Proceedings. vol. 2136, pp. 11–19 (2018)
8. Wang, W.Y.: "liar, liar pants on fire": A new benchmark dataset for fake news detection. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers. pp. 422–426 (2017). <https://doi.org/10.18653/v1/P17-2067>, <https://doi.org/10.18653/v1/P17-2067>
9. Yano, T., Cohen, W.W., Smith, N.A.: Predicting response to political blog posts with topic models (June), 477 (2009). <https://doi.org/10.3115/1620754.1620824>