Classificação de opinião no Twitter em português utilizando o Multilingual Universal Sentence Encoder para apoiar pesquisas sobre filter bubble

Jônatas Castro dos Santos, Sean Wolfgand Matsui Siqueira

Dep. de Informática Aplicada – Universidade Federal do Estado do Rio de Janeiro
Rio de Janeiro – RJ - Brasil

{jonatas.santos,sean}@uniriotec.br

Abstract. *Research on filter bubbles needs mechanisms to capture and classify opinion polarity of documents in crowdsourced environments. We present a preliminary machine learning based model to classify the opinion of tweets in Portuguese about Social Security Reform on Brazilian political context. Our approach uses the Multilingual Universal Sentence Encoder for Semantic Retrieval, a pre-trained model recently released by Google researchers to generate semantically rich vector representations. We train and classify our dataset in a deep feedfoward neural network. Our preliminary model presented an average accuracy of 82%. This task is part of ongoing work to support research on filter bubbles.*

Resumo. *Pesquisas sobre filter bubbles necessitam de mecanismos para capturar e classificar a polaridade de opinião de documentos em ambientes crowdsourced. Apresentamos um modelo preliminar baseado em aprendizado de máquina para classificar a opinião de tweets em português sobre o tema da Reforma de Previdência no contexto político brasileiro. Nossa abordagem utiliza o Multilingual Universal Sentence Encoder for Semantic Retrieval, um modelo pré-treinado recém-lançado por pesquisadores do Google para gerar representações vetoriais semanticamente ricas. Treinamos e classificamos nosso dataset em uma rede neural profunda feedfoward. Nosso modelo preliminar apresentou uma acurácia média de 82%. Esta tarefa é parte de um trabalho em andamento que visa apoiar a realização de pesquisas sobre filter bubbles.*

1. Introdução

A Ciência da Web é uma área que se preocupa em estudar fenômenos que ocorrem na World Wide Web (Hendler, Shadbolt, Hall, Berners-Lee, & Weitzner, 2008). O ambiente de constante produção de dados, volatilidade e mudanças rápidas na WWW elevam a complexidade de estudar esses fenômenos. Entre eles estão as *filter bubbles* que ganharam notoriedade e interesse da academia após a publicação de Eli Pariser (Pariser, 2011).

A hipótese sobre as *filter bubbles* ou *echo chambers* é que os algoritmos de recomendação e personalização criam uma barreira invisível que impede que usuários tenham acesso conteúdos e opiniões contrárias. Isso ocorreria sem a ciência do usuário.

Ao longo da década de 2010, há diversos estudos que se propõe a identificar (Dillahunt, Brooks, & Gulati, 2015), detectar (Tran & Herder, 2015), mensurar (Le et al., 2019) e combater (Bozdag & van den Hoven, 2015) esse fenômeno.

A consequência direta das *filter bubbles* seria o incentivo a polarização política ou de opiniões. A existência de inclinação política em plataformas de conteúdo, sejam em portais de notícias ou em mídias sociais seria algo prejudicial a democracia (Bozdag & van den Hoven, 2015; Nechushtai & Lewis, 2019). Nesse sentido, alguns estudos se preocupam em verificar a existência de inclinação política em canais de comunicação (Kulshrestha et al., 2019). O Twitter, por exemplo, tem sido amplamente utilizado como fonte colaborativa de dados para estudos sobre inclinação política (Ming, Wong, Tan, Sen, & Chiang, 2016).

Entre as tarefas necessárias para verificar a existência de polarização em um conjunto de documentos é a classificação automática da inclinação política de textos. Este é um desafio da disciplina de Processamento de Linguagem Natural (NLP).

Diversas técnicas de aprendizado de máquina têm sido empregadas para o desempenho de classificação de textos em NLP. Dentre as técnicas utilizadas, a codificação de sentenças em representações vetoriais ganhou destaque após a publicação do *Universal Sentence Encoder* (USE) (Cer et al., 2018). Trata-se de um modelo pré-treinado que permite gerar representações vetoriais baseados em transformadores. O modelo mostrou bons resultados no desempenho de diversas tarefas de NLP que necessitam transferir conteúdo semântico.

O USE já foi utilizado em estudos de classificação de inclinação política (Saligrama, 2019). Porém, esse modelo só era disponível para conteúdo em inglês tornando desafiador a aplicação do modelo em outros idiomas.

Em 2019, a mesma equipe lançou o *Multilingual Universal Sentence Encoder for Semantic Retrieval* (MUSE) (Yang et al., 2019) que embarca conteúdo pré-treinado 16 idiomas, incluindo o português, em um único espaço semântico.

Nós utilizamos o MUSE em conjunto com um estimador baseado em redes neurais profundas para classificar a opinião em um dataset capturado do Twitter. Construímos uma ferramenta para realizar a captura desses *tweets*. Utilizamos o contexto político das discussões sobre a Reforma de Previdência para classificar *tweets* que são contra ou a favor da Reforma. Obtemos uma acurácia de 83% nesse modelo preliminar.

Essa tarefa é parte de uma pesquisa que está em andamento e tem como o objetivo criar um modelo que apoie a realização de experimentos em pesquisas sobre *filter bubbles*.

2. Classificando opinião de Tweets

Apresentamos as tarefas que foram executadas na pesquisa até o momento. Definimos um modelo inicial preliminar para uma classificação binária (“a favor” ou “contra”) que servirá como base para os próximos experimentos (Figura 1).

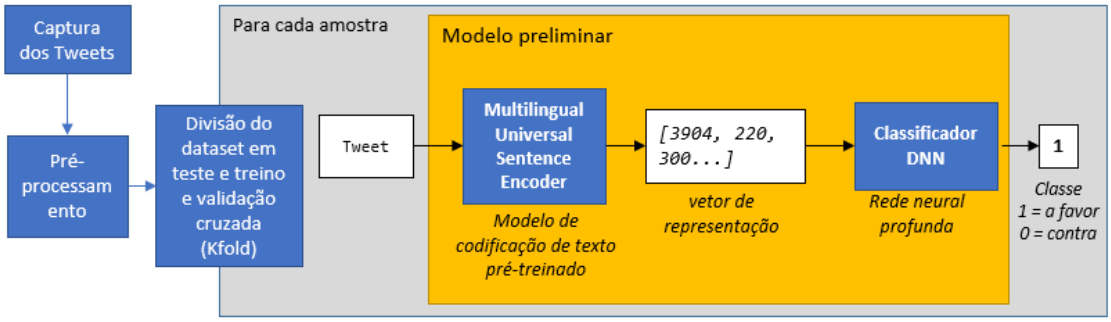

Figura 1 - Modelo preliminar

2.1. Contexto e captura do dataset

No contexto político brasileiro, o tema da Reforma da Previdência tem sido alvo de intensas discussões entre os cidadãos. Nos dias 22 e 23 de Março de 2019, o embate ficou evidente no Twitter quando as hashtags **`#EuApoioANovaPrevidencia`** e **`#LutePelaSuaAposentadoria`** subiram nos *trending topics* (Época Negócios, 2019; IG, 2019).

Portanto, capturamos esses conjuntos de tweets rotulando **`#EuApoioANovaPrevidencia`** como a favor da reforma da reforma e **`#LutePelaSuaAposentadoria`** como contra. Construímos uma ferramenta[1] para efetuar a captura dos tweets através da automação de pesquisa por hashtag no site oficial do Twitter. Foram capturados 12.960 tweets a favor e 15.252 tweets contra, totalizando 28.212 documentos rotulados.

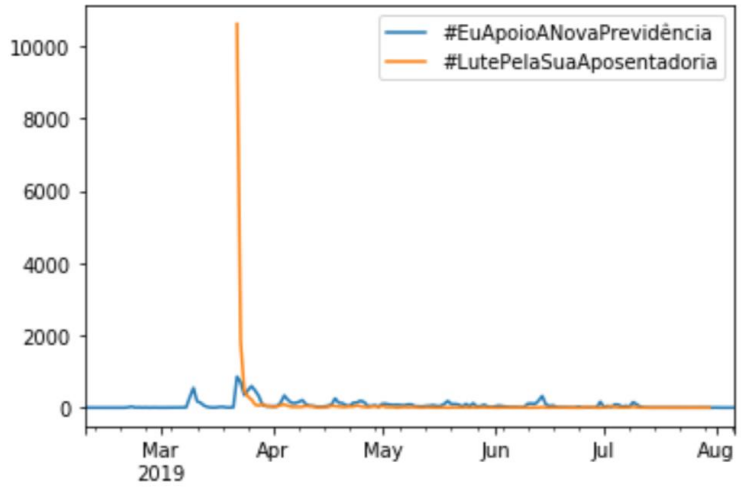

Figura 2 – Frequência dos tweets do dataset capturado

2.2. Pré-processamento

Realizamos uma etapa de pré-processamento no dataset removendo documentos duplicados, possíveis erros do processo de captura (nulos), adicionamos a coluna

[1] Puppetter Searcher - https://github.com/jonatascastro12/puppeter-searcher

correspondente a classe (0 = contra reforma; 1 = a favor da reforma) e removemos as hashtags **#EuApoioANovaPrevidencia / #LutePelaSuaAposentadoria** dos textos dos tweets. Optamos por manter outras hashtags presentes no texto dos tweets pois havia um número considerável de tweets composto por hashtgs. Ao final do pré-processamento nosso dataset era composto de 10.433 tweets classificados a favor da reforma e 11.212 tweets contra. Apresentamos, na Tabela 1, uma amostra do dataset.

Conteúdo	Classe
`#LulaLivre`	0
`Sempre lutaremos Paulo Paim`	0
`#AvanteNovaPrevidencia #Transparência`	1
`@estudioi Não falem por nós, povo brasieiro....`	1
`Meu pai aprosentou hj, que vitória! Já eu, sem...`	0

Tabela 1 - Amostra do dataset pré-processado

2.3. Modelo preliminar

Nosso modelo preliminar é composto por dois componentes principais. O primeiro é o modelo de codificação de texto pré-treinado MUSE[2]. Ele é um modelo multiuso responsável por converter sentenças em vetores de representação. Estes vetores de 512 dimensões capturam informações semanticamente ricas que podem ser utilizados para treinar diferentes tipos classificadores. O MUSE é uma extensão do USE e é baseado na técnica "Multi-task Dual Encoder" (Yang et al., 2018), com o principal diferencial de permitir a utilização do modelo em outras línguas, além do inglês.

O vetor de representação gerado pelo MUSE é a entrada de uma rede neural pré-configurada, disponibilizada pela biblioteca TensorFlow[3]. O *DNNClassifier*[4] abstrai uma rede neural profunda *feedfoward*. Instanciamos uma rede com duas camadas invisíveis.

A partir desses dois componentes fazemos o processo de validação cruzada com cinco iterações (k=5). Para cada iteração, o dataset é divido em uma amostra de treino e teste. Optamos por realizar utilizar o `sklearn.model_selection.StratifiedKFold`[5] por garantir que a divisão aleatória das amostras preserve a percentagem de amostras de cada classe. O modelo é, portanto, treinado e validado gerando um resultado de acurácia em cada iteração.

2.4. Resultados

Apresentamos os resultados de cada iteração na Tabela 2. O modelo preliminar apresentou uma acurácia média de 82,33%.

---

[2] https://tfhub.dev/google/universal-sentence-encoder-multilingual-large/1

[3] https://www.tensorflow.org

[4] https://www.tensorflow.org/api_docs/python/tf/estimator/DNNClassifier

[5] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html

k	Perda	Precisão	Recall	Acurácia
1	47,6432	0,83459	0,7954	0,825404
2	48,1628	0,818268	0,824149	0,827021
3	49,7152	0,813024	0,78965	0,811042
4	46,4234	0,816682	0,830777	0,828558
5	46,6882	0,815589	0,822627	0,824861

Tabela 2 – Resultados preliminares

3. Conclusão

Esse trabalho representa um mínimo produto viável para a tarefa de classificar opinião no Twitter utilizando técnicas de classificação de texto baseadas em aprendizado de máquina.

Na impossibilidade de obter um dataset histórico de tweets pela busca de hashtag via API pública do Twitter, desenvolvemos uma ferramenta para captura dos tweets. Esta ferramenta pode ser utilizada para geração de datasets em pesquisas sobre verificação de filter bubbles em redes sociais e sistemas de recuperação da informação.

Aplicamos um modelo de classificação de tweets em um dataset em português. Não encontramos trabalhos que explorem técnicas de classificação de textos mais recentes em datasets de tweets em português.

Com relação ao modelo apresentado, pretendemos otimizá-los utilizando técnicas de hyperparmeter tuning, explorar inferências com relação as hashtags do dataset e ainda experimentar outras técnicas de pré-processamento antes de injetar os documentos no modelo. Também, pretende-se comparar diferentes abordagens de redes neurais profundas.

A partir desse modelo, almejamos classificar opinião de notícias com objetivo de verificar polaridade de opinião de portais. Este é um passo para formação de um modelo para apoiar pesquisas de *filter bubbles.*

Referências

Bozdag, E., & van den Hoven, J. (2015). Breaking the filter bubble: democracy and design. *Ethics and Information Technology*, *17*(4), 249–265. https://doi.org/10.1007/s10676-015-9380-y

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., John, R. St., … Kurzweil, R. (2018). Universal Sentence Encoder. Retrieved from http://arxiv.org/abs/1803.11175

Dillahunt, T. R., Brooks, C. A., & Gulati, S. (2015). Detecting and Visualizing Filter Bubbles in Google and Bing, 1851–1856. https://doi.org/10.1145/2702613.2732850

Época Negócios. (2019). Reforma da Previdência cria "guerra" de hashtags no Twitter. Retrieved August 1, 2019, from https://epocanegocios.globo.com/Brasil/noticia/2019/03/reforma-da-previdencia-cria-guerra-de-hashtags-no-twitter.html

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., & Weitzner, D. (2008). Web science: an interdisciplinary approach to understanding the web. *Communications of the ACM*, *51*(7), 60. https://doi.org/10.1145/1364782.1364798

IG. (2019). Em dia de atos contra a reforma, nova Previdência domina as redes. Retrieved from https://economia.ig.com.br/2019-03-22/manifestacoes-previdencia.html

Kulshrestha, J., Eslami, M., Messias, J., Zafar, M. B., Ghosh, S., Gummadi, K. P., & Karahalios, K. (2019). Search bias quantification: investigating political bias in social media and web search. *Information Retrieval Journal*, *22*(1–2), 188–227. https://doi.org/10.1007/s10791-018-9341-2

Le, H., Maragh, R., Ekdale, B., High, A., Havens, T., & Shafiq, Z. (2019). Measuring Political Personalization of Google News Search, 2957–2963. https://doi.org/10.1145/3308558.3313682

Ming, F., Wong, F., Tan, C. W., Sen, S., & Chiang, M. (2016). Quantifying Political Leaning from Tweets, Retweets, and Retweeters. *IEEE Transactions on Knowledge and Data Engineering*, *28*(8), 2158–2172. https://doi.org/10.1109/TKDE.2016.2553667

Nechushtai, E., & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in Human Behavior*, *90*(June 2018), 298–307. https://doi.org/10.1016/j.chb.2018.07.043

Pariser, E. (2011). *The Filter Bubble: What The Internet Is Hiding From You*.

Saligrama, A. (2019). KnowBias: A Novel AI Method to Detect Polarity in Online Content. Retrieved from http://arxiv.org/abs/1905.00724

Tran, G., & Herder, E. (2015). Detecting Filter Bubbles in Ongoing News Stories. *Extended Proc. UMAP 2015*. Retrieved from http://www.l3s.de/~herder/research/papers/2015/umap2015-eumssi-lbr.pdf

Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., … Kurzweil, R. (2019). Multilingual Universal Sentence Encoder for Semantic Retrieval. Retrieved from http://arxiv.org/abs/1907.04307

Yang, Y., Cer, D., Yuan, S., Sung, Y., Strope, B., & Kurzweil, R. (2018). Learning Cross-Lingual Sentence Representations via a Multi-task Dual-Encoder Model. In *4th Workshop on Representation Learning for NLP*.