Cuantificando la organización social a través del procesamiento del lenguaje natural

Franco Demarco[1], Juan Manuel Ortiz de Zarate[1,2] and Esteban Feuerstein[1,2]

[1] Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Argentina
[2] Instituto en Ciencas de la Computación

Abstract. El debate sobre la integración y fragmentación social en las plataformas de redes sociales online sigue en curso. El desplazamiento de los usuarios hacia extremos ideológicos y agrupamiento en "cámaras de eco" [1, 2] homogéneas son preocupantes. Waller et al. [3] recientemente desarrollaron un método para cuantificar el posicionamiento de las comunidades en Reddit a lo largo de las dimensiones sociales en base a la concurrencia de usuarios en distintas comunidades. Utilizaron embeddings de comunidades para proyectarlas en direcciones unidimensionales que representan "dimensiones ideológicas", obteniendo puntajes o scores que posicionan a cada comunidad en el espectro político-ideológico.
Proponemos desarrollar una técnica análoga pero utilizando el texto de los posteos y comentarios de los subreddits en lugar de las interacciones. La hipótesis es que las jergas, tópicos y formas discursivas de cada comunidad permiten cuantificar muchos de sus aspectos ideológicos de forma similar a sus interacciones. Utilizamos Fasttext [4] y LLMs [5, 6] para estimar diferentes tipos de embeddings de texto y RBO [7] para comparar los resultados obtenidos. Los resultados preliminares sugieren que existe una relación estadísticamente significativa entre los scores obtenidos y los reportados en el trabajo de Waller et al., lo que podría señalar la existencia de jergas propias de las comunidades que permiten cuantificar su posicionamiento ideológico.

Keywords: PLN - LLM - Redes Sociales - Comunidades

1 Corpus y método

Para este trabajo utilizamos un subconjunto de los datos usados en Waller et al.: todos los posteos hechos en Reddit durante 2012 y 2018. En [3], se basaron en el mismo período pero usaron también los comentarios de cada posteo.

El método desarrollado por nosotros se diferencia de [3] sólo en la creación de los embeddings. Mientras que ellos utilizan las interacciones de los usuarios con las comunidades para crear los vectores de cada comunidad nosotros utilizamos solamente el texto de los posteos de cada comunidad. Con ese texto probamos dos variantes de embeddings: Fasttext entrenado sobre el corpus y utilizando pretrained-word vectors y Cohere[3], un LLM basado en Attention similar a GPT-3 [8].

2 Resultados preliminares

En el siguiente gráfico puede observarse el ranking reportado por [3] versus el obtenido usando el texto de las discusiones de 2018. A la izquierda están las posiciones de Waller y al derecha las del LLM. Si la línea que une izquierda y derecha es derecha (sin inclinación) quiere decir que ambas comunidades fueron ubicadas en la misma posición del ranking por ambos métodos. En cambio si la recta tiene una pendiente positiva, quiere decir que en nuestro método la comunidad fue rankeada en una posición mayor y viceversa si la pendiente de la recta es negativa. Además del lado derecho puede verse el nombre de cada comunidad. Puede observarse por ejemplo que las comunidades "democrats" y "Conservative" tuvieron el mismo lugar en el ranking en ambos métodos, mientras que la comunidad "Enough_Sanders_Spam" quedó en segunda lugar en nuestro ranking mientras que en el de Waller estaba en la 6ta posición.

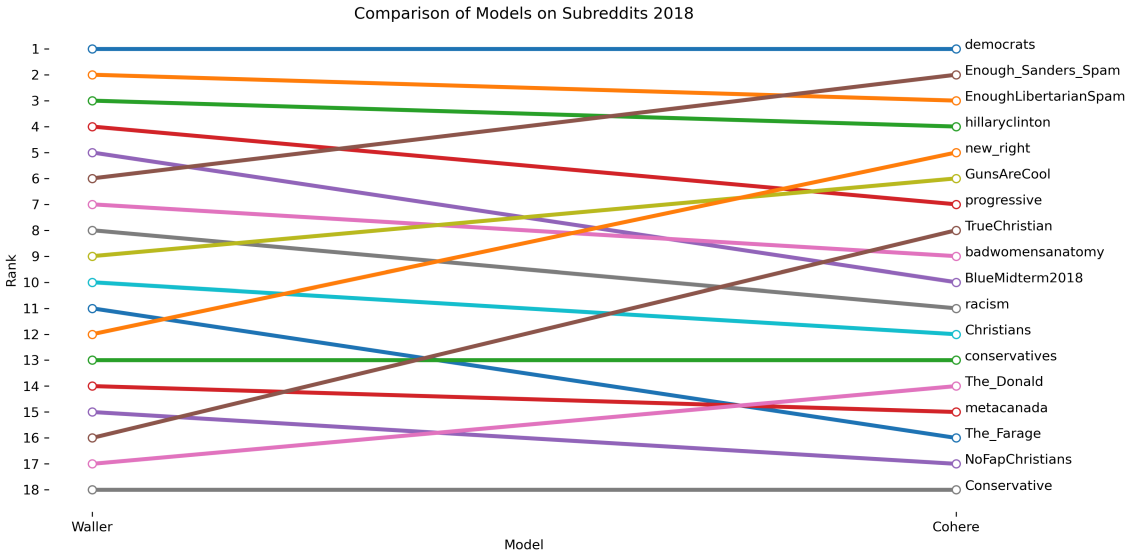

Fig. 1. Ranking de comunidades reportado por Waller et al. vs el obtenido usando LLM

[3] https://cohere.com/

Puede observarse que si bien varias líneas tienen inclinación, la misma es muy leve y los rankings parecen ser bastante parecidos. Para cuantificar mas concretamente esto medimos los parecidos de los rankings de cada año mediante RBO[7].

En el gráfico de barras a continuación podemos ver el score de RBO obtenido por Cohere y Fasttext en cada año."Raw" hace referencia a Fasttext entrenado sobre el corpus de ese año sin pre-trained word vectors, "pretrained" a la versión utilizándolos y "truncated 10k" a lo mismo pero usando sólo 10mil palabras de los posteos mas populares (con mayor cantidad de likes).

Fig. 2. RBO score por año y por tipo de embedding utilizado

Vemos que Cohere tiende a tener el maoyr ranking en todos los casos, teniendo además en muchos de ellos valores cercanos a uno (coincidencias perfecta entre ambos rankings).

References

1. Bail, C. A. et al. Exposure to opposing views on social media can increase political polarization. Proc. Natl Acad. Sci. USA 115, 9216–9221 (2018)
2. Farrell, H. The consequences of the internet for politics. Ann. Rev. Pol. Sci. 15, 35–52 (2012).
3. Waller, I., & Anderson, A. (2021). Quantifying social organization and political polarization in online platforms. Nature, 600(7888), 264-268.
4. Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2016). Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759.
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
6. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942.
7. WEBBER, William; MOFFAT, Alistair; ZOBEL, Justin. A similarity measure for indefinite rankings. ACM Transactions on Information Systems (TOIS), 2010, vol. 28, no 4, p. 1-38.
8. BROWN, Tom, et al. Language models are few-shot learners. Advances in neural information processing systems, 2020, vol. 33, p. 1877-1901.