Aplicación de modelos de lenguaje para la identificación de emociones presentes en twitter durante el periodo de elecciones presidenciales en Colombia 2022

Tesis presentada para optar por el titulo de

Magister en Explotación de Datos y Descubrimiento del Conocimiento

por

Juan Jose Iguaran Fernandez

Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Departamento de Ciencias de la Computación

[Insert Month and Year]

AGRADECIMIENTOS

Agradezco

Resumen

El presente trabajo .....

Palabras Clave: [aquí van]

Índice general

Índice de figuras

Índice de cuadros

Capítulo 1

Introducción

1.1. Motivación

Este trabajo es importante por que

1.2. Marco Teórico

En [Ekman, 1993] Ekman habla sobre las seis emociones básicas que sirven de base para el estudio.

En el libro [Picard, 2000] Picard da un vistazo general sobre el uso de computadoras para detectar emociones.

En [Ortony et al., 1987] se hace una relerencia al relacion que existe entre los estados emocionales y el lexico utilizado.

En [Hatzivassiloglou and McKeown, 1997], [Strapparava et al., 2004] y [Esuli and Sebastiani, 2006] se expande este concepto para elaborar un léxico robusto asociado a emociones.

En [Wiebe, 1994] se plantea que el analisis de sentimiento es un caso particular de analisis de subjetividad.

En [Yu and Hatzivassiloglou, 2003] se plantea un metodo para separar opiniones de hechos.

En [Wilson et al., 2009] se muestra como el contexto de na frase puede camiar el sentimiento de una palabra en particular

En [Pang et al., 2002], [Pang and Lee, 2004], [Dave et al., 2003], [Wilson et al., 2005], [Turney, 2002], [Nasukawa and Yi, 2003], [Kim and Hovy, 2004] y se analiza la detección de sentimiento en el texto

En [Wiebe et al., 2005][Strapparava and Mihalcea, 2008], [Strapparava and Mihalcea, 2007], [Alm et al., 2005],[Aman and Szpakowicz, 2007], [Liu et al., 2003] se puede apreciar como el texto puede ser utilizado para detectar emociones.

Luego, en [Pang et al., 2008] se muestra como los foros de Internet son una fuente de información de l cual se puede extraer valiosa información, entre esos detectar emociones.

En [Read, 2005] se utilizan emoticones en los blogs de internet para detectar sentimiento.

En [Pak and Paroubek, 2010], [Kouloumpis et al., 2011] y en [Go et al., 2009], [Barbosa and Feng, 2010] se aprecia como twitter puede ser usado como fuente para identificar sentimientos positivos, negativos y neutros.

En [O'Connor et al., 2010] se muestra como los sentimientos encontrados en twitter corresponden con resultados de encuestas de opinión.

En [Davidov et al., 2010] se utilizan los hashtags y los emoticones para la clasificacion

En [Hasan et al., 2014], , [Wang et al., 2012]y en [Roberts et al., 2012] se plantea la clasificación mediante distintos algoritmos de las emociones en los tweets.

En [Mohammad, 2012] se hace uso de los hashtags para identificar emociones y entrenar los modelos.

En [Bollen et al., 2011] se observa la relación entre los eventos sociales, políticos y económicos y las emociones detectadas.

En [Tumasjan et al., 2010] se realiza un análisis de sentimientos durante una campaña política.

En [Hochreiter and Schmidhuber, 1997] se desarrollan las redes LSTM que son un tipo de RNN.

En [Chung et al., 2014] se utilizan las GRU que son otro tipo de RNN y superan a las LSTM

En [Vaswani et al., 2017] se desarrollan la tecnica de trasnformers que combina LSTM y GRU, de las cuales bert es un ejemplo

En [Devlin et al., 2018] se desarrolla BERT

En [Acheampong et al., 2021] se hace un recuento de el uso de transformers para detectar emociones

EN [Canete et al., 2020] se propone una aplicación de BERT para español

En [Gonzalez et al., 2021], [Huang et al., 2019] se utiliza bert en twitter para detectar emociones

En [Plaza-del Arco et al., 2020], [Gil et al., 2013] se hace una clasificación de emociones en español.

En [Sidorov et al., 2012] propone un léxico de palabras en español asociadas a emociones ‘

Capítulo 2

Metodología

Capítulo 3

Datos

Bibliografía

[Acheampong et al., 2021] Acheampong, F. A., Nunoo-Mensah, H., and Chen, W. (2021). Transformer models for text-based emotion detection: a review of bert-based approaches. *Artificial Intelligence Review*, 54(8):5789–5829.

[Alm et al., 2005] Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 579–586.

[Aman and Szpakowicz, 2007] Aman, S. and Szpakowicz, S. (2007). Identifying expressions of emotion in text. In *International Conference on Text, Speech and Dialogue*, pages 196–205. Springer.

[Barbosa and Feng, 2010] Barbosa, L. and Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Coling 2010: Posters*, pages 36–44.

[Bollen et al., 2011] Bollen, J., Mao, H., and Pepe, A. (2011). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings*

of the international AAAI conference on web and social media, volume 5, pages 450–453.

[Canete et al., 2020] Canete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. *Pml4dc at iclr*, 2020:1–10.

[Chung et al., 2014] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.

[Dave et al., 2003] Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, pages 519–528.

[Davidov et al., 2010] Davidov, D., Tsur, O., and Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010: Posters*, pages 241–249.

[Devlin et al., 2018] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[Ekman, 1993] Ekman, P. (1993). Facial expression and emotion. *American psychologist*, 48(4):384.

[Esuli and Sebastiani, 2006] Esuli, A. and Sebastiani, F. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

[Gil et al., 2013] Gil, G. B., Jesús, A. B. d., and Lopéz, J. M. M. (2013). Combining machine learning techniques and natural language processing to infer emotions using spanish twitter corpus. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 149–157. Springer.

[Go et al., 2009] Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, 1(12):2009.

[Gonzalez et al., 2021] Gonzalez, J. A., Hurtado, L.-F., and Pla, F. (2021). Twilbert: Pre-trained deep bidirectional transformers for spanish twitter. *Neurocomputing*, 426:58–69.

[Hasan et al., 2014] Hasan, M., Rundensteiner, E., and Agu, E. (2014). Emotex: Detecting emotions in twitter messages.

[Hatzivassiloglou and McKeown, 1997] Hatzivassiloglou, V. and McKeown, K. (1997). Predicting the semantic orientation of adjectives. In *35th annual meeting of the association for computational linguistics and 8th conference of the european chapter of the association for computational linguistics*, pages 174–181.

[Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

[Huang et al., 2019] Huang, C., Trabelsi, A., and Zaïane, O. R. (2019). Ana at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and bert. *arXiv preprint arXiv:1904.00132*.

[Kim and Hovy, 2004] Kim, S.-M. and Hovy, E. (2004). Determining the sentiment of opinions. In *COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics*, pages 1367–1373.

[Kouloumpis et al., 2011] Kouloumpis, E., Wilson, T., and Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg! In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 538–541.

[Liu et al., 2003] Liu, H., Lieberman, H., and Selker, T. (2003). A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th international conference on Intelligent user interfaces*, pages 125–132.

[Mohammad, 2012] Mohammad, S. (2012). # emotional tweets. In ** SEM 2012: The First Joint Conference on Lexical and Computational Semantics–Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 246–255.

[Nasukawa and Yi, 2003] Nasukawa, T. and Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 70–77.

[O'Connor et al., 2010] O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Fourth international AAAI conference on weblogs and social media*.

[Ortony et al., 1987] Ortony, A., Clore, G. L., and Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive science*, 11(3):341–364.

[Pak and Paroubek, 2010] Pak, A. and Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

[Pang and Lee, 2004] Pang, B. and Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *arXiv preprint cs/0409058*.

[Pang et al., 2008] Pang, B., Lee, L., et al. (2008). Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval*, 2(1–2):1–135.

[Pang et al., 2002] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.

[Picard, 2000] Picard, R. W. (2000). *Affective computing*. MIT press.

[Plaza-del Arco et al., 2020] Plaza-del Arco, F. M., Martín-Valdivia, M. T., Ureña-López, L. A., and Mitkov, R. (2020). Improved emotion recognition in spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*, 110:1000–1008.

[Read, 2005] Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL student research workshop*, pages 43–48.

[Roberts et al., 2012] Roberts, K., Roach, M. A., Johnson, J., Guthrie, J., and Harabagiu, S. (2012). Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 3806–3813.

[Sidorov et al., 2012] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Trevino, A., and Gordon, J. (2012). Empirical study of machine learning based

approach for opinion mining in tweets. In *Mexican international conference on Artificial intelligence*, pages 1–14. Springer.

[Strapparava and Mihalcea, 2007] Strapparava, C. and Mihalcea, R. (2007). Semeval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74.

[Strapparava and Mihalcea, 2008] Strapparava, C. and Mihalcea, R. (2008). Learning to identify emotions in text. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1556–1560.

[Strapparava et al., 2004] Strapparava, C., Valitutti, A., et al. (2004). Wordnet affect: an affective extension of wordnet. In *Lrec*, volume 4, page 40. Lisbon, Portugal.

[Tumasjan et al., 2010] Tumasjan, A., Sprenger, T., Sandner, P., and Welpe, I. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 4, pages 178–185.

[Turney, 2002] Turney, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

[Wang et al., 2012] Wang, W., Chen, L., Thirunarayan, K., and Sheth, A. P. (2012). Harnessing twitter"big data"for automatic emotion identification. In *2012 Inter-*

national Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, pages 587–592. IEEE.

[Wiebe et al., 2005] Wiebe, J., Wilson, T., and Cardie, C. (2005). Annotating expressions of opinions and emotions in language. *Language resources and evaluation*, 39(2):165–210.

[Wiebe, 1994] Wiebe, J. M. (1994). Tracking point of view in narrative. *arXiv preprint cmp-lg/9407019*.

[Wilson et al., 2005] Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*, pages 347–354.

[Wilson et al., 2009] Wilson, T., Wiebe, J., and Hoffmann, P. (2009). Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational linguistics*, 35(3):399–433.

[Yu and Hatzivassiloglou, 2003] Yu, H. and Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the 2003 conference on Empirical methods in natural language processing*, pages 129–136.