

La Clasificación de Emociones y su Aplicación a la Detección del Lenguaje Ofensivo

Emotion Classification and its Application to Offensive Language Detection

Flor Miriam Plaza-del-Arco

Departamento de Informática, Escuela Politécnica Superior de Jaén
Universidad de Jaén, Campus Las Lagunillas, 23071, Jaén, Spain
fmplaza@ujaen.es

Resumen: La clasificación de emociones consiste en la identificación de diferentes categorías emocionales en el texto. Se considera una tarea relevante dentro del Procesamiento del Lenguaje Natural ya que ofrece beneficios en una amplia gama de aplicaciones, incluyendo la minería de datos, el aprendizaje online, la interacción persona-ordenador y la psicología. Debido a su complejidad y a la escasa disponibilidad de recursos léxicos anotados, se encuentra en una primera etapa de investigación. Además, la mayoría de los estudios y recursos existentes están disponibles para el inglés, pero la presencia en Internet de otras lenguas, como el español, es cada vez mayor. Por esta razón, en este trabajo, se describe un proyecto de tesis cuyo objetivo es el desarrollo de sistemas orientados a la clasificación de emociones en textos en español. Además, se pretende utilizar dichos sistemas para resolver la detección del lenguaje ofensivo en las redes sociales.

Palabras clave: Clasificación de emociones, procesamiento del lenguaje natural, recursos léxicos afectivos, lenguaje ofensivo

Abstract: Emotion classification consists of identifying discrete emotions in textual information. It is an important natural language processing task whose solution can benefit several applications in different fields, including data mining, e-learning, human-computer interaction, and psychology. However, due to its complexity and the limited availability of annotated lexical resources, it is still in its infancy. In addition, most of the work and resources have been focus on English texts, but the presence of other languages, such as Spanish, is growing everyday on the Web. Therefore, in this work, we describe a thesis project that will focus on the development of emotion classification systems in Spanish texts. Moreover, we aim to use these systems to solve the task of offensive language detection in social media.

Keywords: Emotion classification, natural language processing, affective lexicons, offensive language

1 *Justificación de la investigación propuesta*

Los seres humanos experimentamos diferentes emociones en nuestro día a día, a menudo estas dirigen nuestras conductas y nos ayudan a relacionarnos. Estos estados afectivos, desde el punto de vista psicológico, han sido estudiadas a lo largo de la historia, dando lugar a diferentes teorías que han intentando clasificar las emociones básicas del ser humano.

Actualmente, la mayoría de las personas utilizan cada vez más algunas de las herramientas facilitadas por los ordenadores para realizar sus tareas diarias, lo que ha incre-

mentado la necesidad de mejorar las interacciones entre el ser humano y las máquinas. La insuficiencia de conocimientos de sentido común hace que la emoción sea difícil de reconocer por un programa. Por esta razón, cada vez son más los investigadores que estudian el reconocimiento de emociones, que se divide principalmente en tres categorías: reconocimiento de emociones a través de las expresiones faciales, reconocimiento de emociones a través del habla y reconocimiento de emociones en el texto.

El reconocimiento de emociones en el texto es una tarea que se enmarca dentro del Procesamiento del Lenguaje Natural (PLN),

y en concreto en una de sus principales áreas, el Análisis de Sentimientos (AS). Esta tarea consiste en asignar automáticamente una emoción a un texto en base a unas categorías emocionales previamente definidas, tales como la tristeza, la alegría, el enfado o el miedo. Tiene el potencial de humanizar las interacciones digitales y ofrecer beneficios en una gama casi ilimitada de aplicaciones. Por ejemplo, en una situación de aprendizaje online, se podría detectar cuándo un alumno se siente frustrado y ofrecer explicaciones ampliadas o información adicional. En la telemedicina, puede ayudar a los médicos a comprender rápidamente el estado de ánimo de un paciente o buscar signos de depresión. En las redes sociales, se podría identificar usuarios que están sufriendo ciberbullying o incluso usuarios que piensan en suicidarse. Otras aplicaciones comerciales que se están explorando actualmente incluyen la gestión de relaciones con los clientes, la gestión de recursos humanos, el marketing y el entretenimiento.

Por otra parte, las emociones están estrechamente relacionadas con la detección del lenguaje ofensivo. El odio, la ira y la agresividad son emociones que cada vez se propagan más en las redes sociales. La libertad de expresión no siempre se utiliza de forma respetuosa, ya que a veces encontramos usuarios que utilizan lenguaje ofensivo en sus comentarios provocando así daño emocional a sus destinatarios y contribuyendo a la estigmatización y deshumanización de ciertos colectivos. Un mensaje se clasifica como ofensivo cuando contiene alguna forma de lenguaje no aceptable (blasfemias o palabras malsonantes) o una ofensa dirigida, que puede ser directa o indirecta. Esta categoría incluye insultos, amenazas y mensajes que contienen lenguaje profano o palabrotas. Según un informe sobre la evolución de los incidentes relacionados con los delitos motivados por el odio en España realizado por el Ministerio del Interior en 2018¹, Internet y las redes sociales aparecen como los medios más utilizados para difundir la incitación al odio, con un 45,2 % y un 25,9 %, respectivamente. Unos de los principales problemas es que los propietarios de las redes sociales a menudo dependen de los usuarios y de los moderadores de contenido para identificar este tipo de comen-

tarios. Sin embargo, debido a la gran cantidad de contenido que se publica diariamente se requieren sistemas automáticos basados en PLN que ayuden a identificarlos.

Por otra lado, la mayor parte de las investigaciones realizadas hasta el momento sobre el análisis de emociones y sus aplicaciones se centran en inglés. No obstante, hay otros idiomas, como el español, cuya presencia en Internet es cada día mayor. Por esta razón, esta investigación se va a centrar mayoritariamente en textos en español lo que supondrá un gran progreso en esta área. En particular, este proyecto de tesis se centra en el reconocimiento de las emociones en textos escritos en español y su aplicación en varios campos, principalmente la identificación del lenguaje ofensivo aplicando técnicas basadas en el PLN y aprendizaje automático.

El resto del artículo está organizado de la siguiente forma: en primer lugar, en la Sección 2 se mencionará el origen y trabajo relacionado con el proyecto de tesis. En la Sección 3 se describe la investigación propuesta. La Sección 4 expone la metodología y los experimentos que se van a desarrollar y por último, se presentan los elementos de investigación propuestos para su discusión en la Sección 5.

2 Origen y trabajo relacionado

Uno de los primeros estudios relacionados con la computación afectiva es el de Picard (Picard, 1997). Ella propuso la idea de entrenar a los sistemas para identificar las emociones humanas. La construcción de sistemas afectivos requiere un procesamiento multimodal, ya que un ser humano puede expresar emociones a partir de una amplia gama de señales de comportamiento. Los investigadores realizan el análisis a través de diferentes fuentes de información, como los gestos, el habla, los movimientos, la expresión facial o las señales fisiológicas. El reconocimiento de la emoción en un texto es considerado una de las ramas más recientes de la computación afectiva. De hecho, las redes sociales representan una fuente enorme de expresividad emocional textual mayor que cualquier otra. Esta es una de las razones por las que muchos investigadores de áreas como el Procesamiento del Lenguaje Natural (PNL), la Inteligencia Artificial (IA) o la psicología están interesados en este campo.

Los estudios científicos sobre la clasifica-

¹<http://www.interior.gob.es/documents/642012/3479677/informe+2018/ab86b6d9-090b-465b-bd14-cfcafccdfebc>

ción de las emociones humanas datan de la década de 1960. Muchos teóricos han propuesto conjuntos de emociones que tienden a ser básicos con características innatas y universales (Tomkins, 1962; Izard, 1992). Si bien los psicólogos no están de acuerdo sobre qué modelo describe con mayor precisión el conjunto de emociones básicas, el más utilizado en la investigación informática es el propuesto por Ekman (1992) con 6 emociones (enfado, repulsión, miedo, alegría, tristeza y sorpresa) (Gholipour Shahraki, 2015).

Uno de los pilares fundamentales en la investigación relacionada con la minería de emociones se centra en los recursos lingüísticos disponibles. Los recursos léxicos son indispensables y existen varios disponibles para el idioma inglés, como WordnetAffect (Strapparava y Valitutti, 2004), Emolex (Mohammad y Turney, 2013) NRC Affect Intensity Lexicon (Mohammad y Kiritchenko, 2018) y LIWC (Pennebaker, Francis, y Booth, 2001). Sin embargo, con respecto a la disponibilidad de recursos para otros idiomas, nos encontramos con que el número es bastante más reducido (Yadollahi, Shahraki, y Zaiane, 2017). Concretamente, para el español podemos citar el recurso Spanish Emotion Lexicon (SEL) de Díaz Rangel, Sidorov, y Suárez Guerra (2014).

Los algoritmos basados en el reconocimiento de emociones en el texto pueden clasificarse en cinco categorías: enfoques basados en lexicón, en reglas, en algoritmos tradicionales de aprendizaje automático, en redes neuronales e híbridos (Alswaidan y Menai, 2020). El primero trata sobre el uso de recursos léxicos u ontologías clasificados por emoción (Mohammad, 2012) y en la asignación de una etiqueta de emoción basada en las palabras del texto detectadas en dichos recursos. El segundo trabaja mediante la aplicación de reglas, comparación de resultados y aplicación de las nuevas reglas basadas en situación modificada (Lee, Chen, y Huang, 2010). El tercero aplica algoritmos estadísticos sobre características lingüísticas, los cuales pueden ser supervisados o no supervisados (Chaffar y Inkpen, 2011). El cuarto es una rama del aprendizaje automático en la que los programas aprenden de la experiencia imitando el funcionamiento del cerebro de un humano, entendiendo el mundo en términos de una jerarquía de conceptos, donde cada concepto se define en términos de su relación con

conceptos más simples. Este enfoque permite a un programa aprender conceptos complicados construyéndolos en base a otros más simples (Wang et al., 2016). El último enfoque se basa en la combinación de diferentes enfoques como los explicados previamente (Seol, Kim, y Kim, 2008). Los últimos trabajos se centran en la aplicación de algoritmos basados en *Transfer Learning* para la clasificación de emociones, los cuales han obtenido resultados muy prometedores en el área, superando el estado de la cuestión para diferentes tareas dentro del PLN. El *Transfer Learning* es una forma de mejorar el aprendizaje en una nueva tarea a través de la transferencia de conocimiento de una tarea relacionada que ha sido entrenada previamente (Torrey y Shavlik, 2010). Se utiliza principalmente con el fin de desarrollar modelos para resolver problemas en los que la disponibilidad de datos es limitada. Algunos de los modelos con más éxito hasta el momento son BERT (Devlin et al., 2018) y XLM (Conneau y Lample, 2019).

Por otra parte, la detección del lenguaje ofensivo y el análisis de emociones están estrechamente relacionados, ya que generalmente las emociones negativas aparecen en comentarios maliciosos. En los últimos años, el interés por desarrollar sistemas para combatir contenido malicioso en las redes sociales apoyándose en técnicas basadas en el reconocimiento de emociones ha incrementado, tanto que, cada vez son más los trabajos que proliferan en el ámbito del PLN. Algunos trabajos realizados hasta el momento siguen un enfoque en el que se aplica en primer lugar un clasificador para detectar comentarios negativos antes de que el clasificador final verifique específicamente si hay evidencia de odio (Dinakar et al., 2012; Sood, Churchill, y Antin, 2012; Gitari et al., 2015). Los trabajos más recientes se centran en incorporar características emocionales del texto en arquitecturas basadas en redes neuronales. Safi Samghabadi et al. (2019) desarrollan un mecanismo de atención que utiliza las emociones subyacentes en el texto para encontrar las palabras más importantes. Este módulo lo incorporan en una arquitectura de red neuronal obteniendo buenos resultados en la detección de la ofensividad. Rajamanickam et al. (2020) presentan el primer modelo conjunto de detección de emociones y lenguaje ofensivo, experimentando en un marco de aprendizaje multitarea que permite que una tarea

influya en la otra compartiendo conocimiento. Los resultados obtenidos demuestran que la incorporación de características afectivas conduce a mejoras significativas en la precisión de la detección del lenguaje ofensivo.

3 Descripción de la investigación propuesta

Este proyecto de tesis se propone con la finalidad de desarrollar un sistema automático de reconocimiento de emociones en español con el objetivo de aplicarlo a tareas reales, como por ejemplo, la detección del lenguaje ofensivo en las redes sociales.

En primer lugar, se estudiaron en detalle los trabajos que tratan el reconocimiento de emociones en inglés y en español. Este estudio fue fundamental para el inicio de la investigación en español, ya que se permitió conocer los enfoques más utilizados en inglés y reproducirlos con objeto de conocer y comparar su funcionamiento en español. También se estudiaron en detalle los recursos léxicos disponibles para inglés con el objetivo de adaptarlos y utilizarlos en nuestros sistemas.

El estudio de los recursos léxicos estudiados hasta el momento en inglés, permitió realizar una adaptación de los recursos más utilizados y con más éxito en el análisis de emociones al español. Uno de nuestros primeros trabajos fue el desarrollo de tres sistemas multilingües con motivo de nuestra participación en tres subáreas (EI-oc, EI-reg, E-c) de la Tarea 1 de SemEval: Affect in Tweets (Mohammad et al., 2018). Son tareas relacionadas con la identificación de la intensidad de la emoción y con la clasificación de emociones en tweets. Nuestra principal contribución fue la implementación de un sistema para adaptar WordNet-Affect al español (Plaza-del-Arco et al., 2018a; Plaza-del-Arco et al., 2018b) utilizando diferentes recursos como BabelNet (Navigli y Ponzetto, 2012) o Babelfy (Moro, Raganato, y Navigli, 2014). Posteriormente, se integraron las características derivadas de este sistema junto con las obtenidas de otros lexicones de emociones en un sistema de clasificación supervisado tradicional. Otro de los trabajos relacionados con la adaptación de recursos léxicos al español, fue la adaptación del lexicon NRC Affect Intensity (Mohammad, 2018) y su evaluación sobre el conjunto de datos liberado en la tarea 1 de la competición de SemEval 2018 (Plaza-del-Arco et al., 2018). Más tarde, se desarrolló

un sistema para predecir la emoción expresada en una palabra excluida en el texto integrando características emocionales derivadas de diferentes lexicones en algoritmos basados en redes neuronales (Plaza-del-Arco et al., 2018a).

Tras la adaptación de algunos recursos léxicos al español, decidimos analizar la importancia de disponer de recursos emocionales para un idioma en concreto, comprobando si una simple traducción entre recursos es suficiente o si por el contrario, es necesario disponer de recursos para cada idioma con el objetivo de mejorar la precisión de los sistemas. En particular, experimentamos con dos recursos léxicos de emociones: Emolex, traducido al español y SEL, creado específicamente para el español. Nuestros experimentos mostraron que los mejores resultados se obtienen utilizando las características léxicas derivadas de SEL. Por tanto, pudimos observar que el uso de otros léxicos adaptado en español con una simple traducción no siempre mejora el sistema de clasificación y la importancia de crear recursos en un idioma en concreto para la mejora de los sistemas (Plaza-del-Arco et al., 2019).

El análisis de los conjuntos de datos etiquetados con emociones estudiados hasta el momento en inglés y español, permitió detectar la escasez de corpus disponibles para la evaluación de sistemas de clasificación de emociones. Por ello, otra de las aportaciones de esta tesis es un corpus de tweets multilingüe etiquetados con emociones y basados en diferentes eventos que tuvieron lugar en abril de 2019. El modelo de emociones que se utilizó para etiquetar el corpus es el de Ekman (Ekman, 1992), incluyendo la categoría “otros”. El etiquetado se realizó manualmente por tres anotadores. Finalmente, se obtuvo un total de 8.409 tweets en español y 7.303 tweets en inglés. Además, cada tweet también fue etiquetado como ofensivo o no ofensivo. Con el objetivo de observar la diferencia entre los angloparlantes y los hispano parlantes cuando expresan emociones relacionadas con los mismos eventos, se realizó el análisis de algunas estadísticas lingüísticas sobre el conjunto de datos. Además, para validar su efectividad, utilizamos un algoritmo de aprendizaje automático.

Debido a la multitud de aplicaciones que ofrece el ámbito del reconocimiento de emociones, optamos por centrarnos en una de

ellas, la detección del lenguaje ofensivo en las redes sociales, aplicando sistemas automáticos basados en PLN. Para ello, se ha participado en diferentes competiciones que tratan esta tarea. En primer lugar, participamos una vez más en SemEval, en concreto, en las tareas HatEval (2019) y OffensEval (2019 y 2020). En la primera tarea, implementamos un sistema multilingüe para la detección del odio en redes sociales dirigido a dos objetivos específicos: inmigrantes y mujeres, realizando un sistema que combinaba tres sistemas de aprendizaje supervisado tradicionales. Esta participación, permitió detectar la escasez de la disponibilidad de recursos léxicos en español para su integración en los sistemas de detección del odio. Es por ello que otra de las contribuciones de esta tesis es la creación de recursos lingüísticos apropiados para la detección del discurso de odio en español. Para su evaluación, se utilizó un enfoque basado en lexicón y se comparó su funcionamiento con otro enfoque utilizado en la clasificación de textos: enfoque basado en aprendizaje supervisado utilizando algoritmos tradicionales y redes neuronales (Plaza-Del-Arco et al., 2020).

Finalmente, los objetivos concretos que se pretenden alcanzar con este proyecto de tesis son los siguientes:

- Extraer información subjetiva de las diferentes plataformas (blogs, redes sociales, foros, etc) que dispongan de emociones.
- Generar y adaptar distintos recursos para el reconocimiento de emociones en español, tanto corpus como lexicones.
- Procesar dicha información para desarrollar sistemas que sean capaces de identificar las diferentes categorías emocionales.
- Aplicar los sistemas desarrollados a aplicaciones reales para solucionar problemas actuales como la detección del lenguaje ofensivo en las redes sociales.

4 Metodología y experimentos propuestos

La metodología que se propone para la consecución de esta tesis se presenta a continuación:

1. Estudio y revisión del estado del arte. Se realizará un estudio de la bibliografía

existente sobre la minería de emociones en inglés y en español.

2. Creación, adaptación e integración de recursos existentes para poder realizar un análisis de los métodos propuestos. Se intentará crear recursos lingüísticos además de adaptar ciertos recursos ya disponibles en inglés.
3. Desarrollo de un prototipo. Se tratará de implementar un sistema de detección de emociones para el español y se aplicará a determinadas tareas tales como la detección de la incitación al odio.
 - Diseño de una arquitectura modular que permita integrar nuevas funcionalidades a medida que se vaya avanzando en la investigación.
 - Construcción de la arquitectura modular diseñada.
 - Prueba del correcto funcionamiento del prototipo.
4. Experimentación y evaluación. Se utilizarán los recursos generados para llevar a cabo la experimentación y posteriormente se procederá a la evaluación del prototipo, llevando a cabo una comparación de los resultados obtenidos con los ya existentes. Los resultados obtenidos se pondrán a disposición de la comunidad científica.

5 Elementos de investigación específicos propuestos para discusión

Las principales cuestiones de investigación a las que se pretende responder con este proyecto de tesis son las siguientes:

- Estudios psicológicos muestran que las emociones del ser humano van ligadas a su cultura e idioma. Por tanto, ¿es necesario crear recursos emocionales teniendo en cuenta el idioma? o ¿una simple traducción entre recursos es suficiente?
- ¿Qué características se deben tener en cuenta en el proceso del análisis de emociones? ¿Cómo se pueden utilizar estas características para mejorar los sistemas de reconocimiento de emociones?

- ¿Qué algoritmos son los que nos proporcionan una mayor exactitud para reconocer las diferentes categorías emocionales en un texto?
- Dado que los usuarios en las redes sociales es donde más suelen expresar sus emociones, ¿es útil esta información para la creación de recursos léxicos?
- ¿Es útil incorporar conocimiento afectivo para detectar problemas como la detección del lenguaje ofensivo en las redes sociales?. En caso afirmativo, ¿qué características afectivas aportan más valor?

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el Fondo Europeo de Desarrollo Regional (FEDER) y el proyecto LIVING-LANG (RTI2018-094653-B-C2) del Gobierno de España.

Bibliografía

- Alswaidan, N. y M. E. B. Menai. 2020. A survey of state-of-the-art approaches for emotion recognition in text. *Knowledge and Information Systems*, páginas 1–51.
- Chaffar, S. y D. Inkpen. 2011. Using a heterogeneous dataset for emotion analysis in text. En *Canadian Conference on Artificial Intelligence*, páginas 62–67. Springer.
- Conneau, A. y G. Lample. 2019. Cross-lingual language model pretraining. En *Advances in Neural Information Processing Systems*, páginas 7057–7067.
- Devlin, J., M.-W. Chang, K. Lee, y K. Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.
- Díaz Rangel, I., G. Sidorov, y S. Suárez Guerra. 2014. Creación y evaluación de un diccionario marcado con emociones y ponderado para el español. *Onomazein*, 1(29).
- Dinakar, K., B. Jones, C. Havasi, H. Lieberman, y R. Picard. 2012. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(3):18.
- Ekman, P. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- Gholipour Shahraki, A. 2015. Emotion mining from text.
- Gitari, N. D., Z. Zuping, H. Damien, y J. Long. 2015. A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4):215–230.
- Izard, C. E. 1992. Basic emotions, relations among emotions, and emotion-cognition relations.
- Lee, S. Y. M., Y. Chen, y C.-R. Huang. 2010. A text-driven rule-based system for emotion cause detection. En *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, páginas 45–53.
- Mohammad, S. M. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741.
- Mohammad, S. M. 2018. Word affect intensities. En *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018)*, Miyazaki, Japan.
- Mohammad, S. M., F. Bravo-Marquez, M. Salameh, y S. Kiritchenko. 2018. Semeval-2018 Task 1: Affect in tweets. En *Proceedings of International Workshop on Semantic Evaluation (SemEval-2018)*, New Orleans, LA, USA.
- Mohammad, S. M. y S. Kiritchenko. 2018. Understanding emotions: A dataset of tweets to study interactions between affect categories. En *Proceedings of the 11th Edition of the Language Resources and Evaluation Conference, Miyazaki, Japan*.
- Mohammad, S. M. y P. D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29(3):436–465.
- Moro, A., A. Raganato, y R. Navigli. 2014. Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics*, 2:231–244.
- Navigli, R. y S. P. Ponzetto. 2012. Babelnet: The automatic construction, evaluation and application of a wide-coverage

- multilingual semantic network. *Artificial Intelligence*, 193:217–250.
- Pennebaker, J. W., M. E. Francis, y R. J. Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.
- Picard, R. W. 1997. Affective computing. 1997.
- Plaza-del-Arco, F. M., S. M. Jiménez-Zafra, M. Martín, y L. A. Ureña-Lopez. 2018a. Sinai at semeval-2018 task 1: Emotion recognition in tweets. En *Proceedings of the 12th International Workshop on Semantic Evaluation*, páginas 128–132.
- Plaza-del-Arco, F., S. M. Jiménez-Zafra, M.-T. Martín-Valdivia, y L. A. Ureña-López. 2018b. Using facebook reactions to recognize emotion in political domain.
- Plaza-del Arco, F. M., M. T. Martín-Valdivia, L. A. Ureña-López, y R. Mitkov. 2019. Improved emotion recognition in spanish social media through incorporation of lexical knowledge. *Future Generation Computer Systems*.
- Plaza-Del-Arco, F.-M., M. D. Molina-González, L. A. Ureña-López, y M. T. Martín-Valdivia. 2020. Detecting misogyny and xenophobia in spanish tweets using language technologies. *ACM Transactions on Internet Technology (TOIT)*, 20(2):1–19.
- Plaza-del-Arco, F. M., M. D. Molina-González, S. M. Jiménez-Zafra, y M. T. Martín-Valdivia. 2018. Lexicon adaptation for spanish emotion mining. *Procesamiento del Lenguaje Natural*, 61:117–124.
- Rajamanickam, S., P. Mishra, H. Yannakoudakis, y E. Shutova. 2020. Joint modeling of emotion and abusive language detection. *arXiv preprint arXiv:2005.14028*.
- Safi Samghabadi, N., A. Hatami, M. Shafaei, S. Kar, y T. Solorio. 2019. Attending the emotions to detect online abusive language. *arXiv*, páginas arXiv–1909.
- Seol, Y.-S., D.-J. Kim, y H.-W. Kim. 2008. Emotion recognition from text using knowledge-based ann. En *ITC-CSCC: International Technical Conference on Circuits Systems, Computers and Communications*, páginas 1569–1572.
- Sood, S. O., E. F. Churchill, y J. Antin. 2012. Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, 63(2):270–285.
- Strapparava, C. y A. Valitutti. 2004. Wordnet affect: an affective extension of wordnet. En *Language Resources and Evaluation Conference (LREC)*, volumen 4, páginas 1083–1086.
- Tomkins, S. 1962. *Affect imagery consciousness: Volume I: The positive affects*. Springer publishing company.
- Torrey, L. y J. Shavlik. 2010. Transfer learning. En *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. IGI Global, páginas 242–264.
- Wang, Y., S. Feng, D. Wang, G. Yu, y Y. Zhang. 2016. Multi-label chinese microblog emotion classification via convolutional neural network. En *Asia-Pacific Web Conference*, páginas 567–580. Springer.
- Yadollahi, A., A. G. Shahraki, y O. R. Zaiane. 2017. Current state of text sentiment analysis from opinion to emotion mining. *ACM Comput. Surv.*, 50(2):25:1–25:33, Mayo.