

Análisis de emociones en textos en español

Emotions analysis in Spanish texts

Flor Miriam Plaza del Arco

Departamento de Informática, Escuela Politécnica Superior de Jaén
Universidad de Jaén, E-23071 - Jaén
fimplaza@ujaen.es

Resumen: La minería de emociones es una tarea relativamente reciente que trata de identificar diferentes categorías emocionales en el texto. Sin embargo, debido a su complejidad y a la escasa disponibilidad de recursos léxicos anotados, todavía se encuentra en una primera etapa de investigación. Además, la mayoría de los trabajos y recursos existentes se han realizado para textos en inglés, pero la presencia en Internet de otras lenguas, como el español, es cada vez mayor. Por esta razón, se describe un proyecto de tesis que se va a centrar en el análisis de emociones en textos en español. Los sistemas de minería de emociones son de gran relevancia ya que ofrecen una amplia gama de aplicaciones, desde la salud y el bienestar hasta la creación de perfiles de usuario, la educación y el marketing, entre otras.

Palabras clave: Minería de emociones, procesamiento del lenguaje natural, recursos léxicos afectivos, análisis de sentimientos

Abstract: Emotion mining is a relatively recent task that attempts to identify different emotional categories in text. However, due to its complexity and the limited availability of annotated lexical resources, it is still in the early stages of research. In addition, most of the existing work and resources have been done for English texts, but the presence of other languages, such as Spanish, is growing in the Internet. Therefore, in this work, we describe a thesis project that will focus on the analysis of emotions in Spanish texts. Emotions mining systems are of great relevance as they offer a wide range of applications, from health and wellness to user profiling, education and marketing, among others.

Keywords: Emotion mining, natural language processing, affect lexical resources, sentiment analysis

1 *Justificación de la investigación propuesta*

Las emociones son una pieza clave en el ser humano pues nos ayudan a relacionarnos y adaptarnos al mundo que nos rodea. Además, dirigen la mayor parte de nuestras conductas.

A lo largo de los años, han sido muchas las áreas tales como la psicología, la neurociencia o la filosofía que se han dedicado al estudio y el análisis de las emociones en el ser humano. Recientemente, ha atraído la atención en el campo de la Computación Afectiva (CA).

Uno de los medios más utilizado actualmente por el ser humano para expresar las emociones son las redes sociales. El acceso en tiempo real a la gran cantidad de contenido

generado por los usuarios puede proporcionar herramientas a los investigadores y a los ciudadanos en general para monitorear el pulso de la sociedad hacia temas de interés específicos, una tarea que tradicionalmente solo se realizaba a través de encuestas de opinión, que son costosas y lleva mucho tiempo y que, normalmente se limitan a tamaños de muestra pequeños. Dado el interés de analizar esta información subjetiva, surge un área dentro del Procesamiento del Lenguaje Natural (PLN) dedicada a ello, el Análisis de Sentimientos (AS).

La minería de la emoción se enmarca dentro del AS y de la CA. Esta disciplina trata de identificar diferentes categorías emocionales en un texto, 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 electrónico, un programa de CA podría detectar cuándo un alumno se siente frustrado y ofrecer explicaciones ampliadas o información adicional. En la telemedicina, la programación de CA 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.

La mayor parte de las investigaciones realizadas hasta el momento sobre el análisis de emociones se centran sobre un idioma en concreto que generalmente es el inglés. No obstante, hay otros idiomas, como el español, cuya presencia en Internet es cada día mayor. Además, si se desea desarrollar sistemas comerciales, estos deben adaptarse al entorno en el que serán utilizados. Por esta razón, esta investigación se va a centrar mayoritariamente en textos en español lo que supondrá un gran progreso en ésta área.

Este proyecto de tesis se centra en el análisis de las emociones en textos escritos en español. Analizar y comprender las emociones de los usuarios a través del texto que escriben proporciona múltiples beneficios en distintas áreas.

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 discusión en la Sección 5.

2 Origen y trabajo relacionado

La minería de emociones es cada vez más importante. Algunas de las conferencias principales que se ocupan de la minería y evaluación de datos actualmente incluyen talleres y comparten tareas relacionadas. Estos incluyen Evaluación Semántica (SemE-

val) (Mohammad et al., 2018), Enfoques computacionales de subjetividad y Análisis de Sentimiento (WASSA)¹ y talleres sobre modelado computacional de las opiniones de las personas, personalidad y emociones en las redes sociales (PEOPLE)².

Los estudios científicos sobre la clasificación de las emociones humanas datan de la década de 1960. Desde una perspectiva psicológica, una emoción es básica solo si no contiene otra emoción; es decir, si representa una construcción psicológica irreducible atómica (Celeghin et al., 2017). Además, los humanos pueden reconocer las emociones en todo el mundo, independientemente de su raza, cultura e idioma. Muchos teóricos han propuesto conjuntos de emociones que tienden a ser básicos. 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 (Ekman, 1992), con 6 emociones (enojo, disgusto, miedo, alegría, tristeza y sorpresa) (Gholipour Shahraki, 2015).

En los últimos años, las emociones han atraído la atención de los investigadores en las ciencias de la computación, especialmente en el campo de la interacción con la computadora humana, donde se han llevado a cabo estudios sobre expresiones faciales (Ekman, 1977) o sobre el reconocimiento de emociones a través de una variedad de sensores (Picard, 1997).

A pesar de que la minería de emociones es un área relativamente reciente y no proliferan los artículos que combinan con PLN, el interés es creciente y cada vez son más los trabajos relacionados con esta tarea. Uno de los pilares fundamentales en la investigación relacionada con la minería de la emoción 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 en un idioma distinto al inglés, nos encontramos con que el número es bastante más reducido (Yadollahi, Shahraki, y

¹<https://wt-public.emm4u.eu/wassa2018/>

²<https://peopleswsh.github.io/index.html>

Zaiane, 2017). Concretamente para español podemos citar el recurso Spanish Emotion Lexicon (SEL) de Díaz Rangel, Sidorov, y Suárez Guerra (2014).

Muchos de los estudios relacionados con el reconocimiento de emociones y la creación de recursos léxicos se centran en la evaluación de la información de las redes sociales ya que son lugares apropiados para compartir las emociones de manera fácil y rápida. Las dos principales redes sociales en las que se focalizan dichos estudios son Twitter y Facebook.

Con respecto a Twitter, encontramos algunos trabajos como el de Purver y Battersby (2012) que utilizan tanto los emoticones como los hashtags para el reconocimiento de emociones. Mohammad y Kiritchenko (2015) usan hashtags para reconocer las categorías emocionales de los tweets. Bollen, Mao, y Zeng (2011) analizan las emociones de todos los tweets en un marco de tiempo específico utilizando una prueba psicométrica. En cuanto a Facebook, hay pocos artículos centrados en el reconocimiento de emociones ya que hasta 2016 no se incorporó la nueva característica que permite a los usuarios reaccionar a una publicación o comentario. Algunos trabajos relacionados con esta red social se mencionan a continuación: Krebs et al. (2017) recopiló publicaciones y sus reacciones de páginas de Facebook y construyó un conjunto de datos, para predecir la distribución de las reacciones en las publicaciones, combinaron las técnicas de análisis de sentimiento y de emoción con las arquitecturas de red neuronal. Pool y Nissim (2016) aprovechan la función de reacción de Facebook de forma distante y supervisada para entrenar un clasificador SVM para la detección de emociones, usando varias combinaciones de características y combinando diferentes páginas de Facebook.

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.

En primer lugar, se están estudiando en detalle los trabajos que tratan el reconocimiento de emociones en inglés y en español. Este estudio es fundamental para obtener conocimiento de los enfoques más utilizados en inglés y reproducirlos con objeto de conocer

y comparar su funcionamiento en español.

Elegimos realizar el trabajo en español, ya que son escasos los recursos disponibles actualmente en nuestra lengua, a pesar de ser la segunda lengua más hablada en el mundo y la tercera lengua más usada en la Web.

Uno de nuestros primeros trabajos ha sido el desarrollo de tres sistemas multilingües con motivo de nuestra participación en tres subtareas (EI-oc, EI-reg, E-c) de la Tarea 1 de SemEval: Affect in Tweets (Mohammad et al., 2018), tanto en inglés como en español. Son tareas relacionadas con la identificación de la intensidad de la emoción y con la clasificación de emociones en tweets. Nuestra principal contribución ha sido la implementación de un sistema para adaptar WordNet-Affect al español (Plaza-del-Arco et al., 2018a) utilizando diferentes recursos como BabelNet (Navigli y Ponzetto, 2012) o Babelify (Moro, Raganato, y Navigli, 2014). Otro de los trabajos ha sido la adaptación del lexicon NRC Affect Intensity construyendo un nuevo lexicón para el español que ha sido probado sobre el conjunto de datos liberado en la tarea 1 de la competición SemEval 2018. Además, se ha comparado con el único lexicón de intensidades existente en español, el lexicón SEL, y se ha demostrado la dificultad de la tarea y la importancia de continuar trabajando en el desarrollo de recursos. (Plaza-del-Arco et al., 2018c). Por otra parte, se ha realizado un sistema de clasificación de emociones para la Tarea 4 de la competición TASS 2018 (Plaza-del-Arco et al., 2018b). Nuestro sistema se basa en un método supervisado con SVM utilizando características emocionales. Para ello, hacemos uso de distintos lexicones emocionales y realizamos la adaptación de las distintas emociones a la polaridad de las mismas (positiva o negativa).

Por último, los sistemas se desarrollarán en un principio de manera general, pero en una segunda fase se adaptarán a un dominio específico, puesto que consideramos que es posible extraer de redes sociales diversos corpus interesantes que expresen emociones. Podremos utilizar estos recursos para entrenar nuestros sistemas así como para validar el modelo propuesto.

Los objetivos concretos que se pretenden alcanzar con este proyecto son los siguientes:

- Extraer información subjetiva de las diferentes plataformas (blogs, redes socia-

les, foros, etc) que dispongan de emociones.

- Generar distintos recursos para reconocimiento de emociones en español, tanto corpus como lexicones.
- Procesar dicha información para desarrollar sistemas que sean capaces de identificar las diferentes categorías emocionales.

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. Adaptación e integración de recursos existentes para poder realizar un análisis de los métodos propuestos. Se intentará adaptar ciertos recursos disponibles en inglés al español, como, por ejemplo, WordNetAffect o Emolex.
3. Desarrollo de un prototipo. Se tratará de implementar un sistema de detección de emociones para el español.
 - 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 pro-

yecto 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?

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el Fondo Europeo de Desarrollo Regional (FEDER) y el proyecto REDES (TIN2015-65136-C2-1-R) del Gobierno de España.

Bibliografía

- Bollen, J., H. Mao, y X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.
- Celeghin, A., M. Diano, A. Bagnis, M. Viola, y M. Tamietto. 2017. Basic emotions in human neuroscience: neuroimaging and beyond. *Frontiers in Psychology*, 8:1432.
- 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).
- Ekman, P. 1977. Biological and cultural contributions to body and facial movement. 1977, páginas 34–84.
- Ekman, P. 1992. An argument for basic emotions. *Cognition & emotion*, 6(3-4):169–200.
- Gholipour Shahraki, A. 2015. Emotion mining from text.

- Krebs, F., B. Lubascher, T. Moers, P. Schaap, y G. Spanakis. 2017. Social emotion mining techniques for facebook posts reaction prediction. *arXiv preprint arXiv:1712.03249*.
- 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. 2015. Using hashtags to capture fine emotion categories from tweets. *Computational Intelligence*, 31(2):301–326.
- 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. M., E. Martínez-Cámara, M. T. M. Valdivia, y L. A. U. López. 2018b. SINAI en TASS 2018: Inserción de conocimiento emocional externo a un clasificador lineal de emociones (SINAI at TASS 2018: Lineal classification system with emotional external knowledge). En *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN, TASS@SEPLN 2018, co-located with 34th SEPLN Conference (SEPLN 2018), Sevilla, Spain, September 18th, 2018*, páginas 125–130.
- Plaza-del-Arco, F. M., M. D. Molina-González, S. M. Jiménez-Zafra, y M. T. Martín-Valdivia. 2018c. Lexicon adaptation for spanish emotion mining. *Procesamiento del Lenguaje Natural*, 61:117–124.
- Pool, C. y M. Nissim. 2016. Distant supervision for emotion detection using facebook reactions. *arXiv preprint arXiv:1611.02988*.
- Purver, M. y S. Battersby. 2012. Experimenting with distant supervision for emotion classification. En *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, páginas 482–491. Association for Computational Linguistics.
- 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.
- 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.