

para los seres humanos. El lenguaje figurado es especialmente común en los textos que podemos encontrar en la web y en las redes sociales, especialmente en Twitter o Facebook. La limitación en la longitud en los textos de la red social Twitter, así como el uso de expresiones plagadas de argot y errores gramaticales, dificulta la comprensión del mensaje. En definitiva, el lenguaje figurado presenta un desafío para el rendimiento de los sistemas de análisis de sentimientos convencionales basados en la semántica léxica de las palabras ya que a menudo resultan insuficientes para detectar los significados indirectos. En el trabajo de Hernández et al., (2016) puede encontrarse un estudio detallado del impacto de la ironía y el sarcasmo en el análisis de sentimientos.

En este trabajo se presenta un estudio del impacto de las emociones en la detección de la polaridad de un tweet. Partimos de la hipótesis de que no todos los recursos disponibles favorecen la detección de la polaridad en igual medida, por lo tanto llevamos a cabo una serie de experimentos para evaluar cómo afectan diferentes recursos de emociones tanto en el lenguaje figurado como en el lenguaje literal. Nuestra metodología está compuesta por dos fases: en la primera de ellas estudiaremos el impacto de los recursos léxicos sobre el entrenamiento de clasificadores que predigan la polaridad del conjunto completo de tweets de la tarea 11 de SemEval2015¹ y a continuación evaluaremos en detalle el impacto de cada uno de los recursos para las diferentes tipologías del lenguaje figurado presentes en este corpus.

El resto del artículo está organizado de la siguiente manera: en primer lugar describiremos el estado de la cuestión en el apartado 2; la metodología se describe en detalle en el apartado 3; a continuación en el apartado 4 se presenta el conjunto de datos sobre el cual validaremos nuestra metodología, que como ya hemos introducido se trata de la tarea 11 de SemEval2015; en el apartado 5 se presentan los experimentos que se han llevado a cabo; a continuación en el apartado 6 evaluaremos los resultados obtenidos y, por último, extraeremos las debidas conclusiones en el apartado 7.

2 Estado de la cuestión

Como ya hemos introducido, la definición más extendida de la tarea de análisis de sentimientos se centra en clasificar los textos en tres categorías: textos positivos, negativos y neutros. Los trabajos pioneros (Pang, Lee, y Vaithyanathan, 2002) abordaron esta tarea como un problema de clasificación supervisada aunque en la literatura también podemos encontrar aproximaciones no supervisadas (Turney, 2002). En el trabajo de Pang y Lee (2008) se recoge un amplio estudio de las distintas técnicas que se han empleado para tratar de resolver la tarea del análisis de sentimientos sobre textos extraídos de Internet.

La detección del lenguaje figurado es una tarea en si misma, y distintas aproximaciones han intentado abordarla. Cuando se trata de analizar los textos, la información disponible en la web se puede utilizar como una fuente de conocimiento para generar características auxiliares. En el trabajo de Veale y Hao (2007) se describe una forma semiautomática de recopilar el conocimiento y la semántica de los estereotipos de la web atacando directamente a las construcciones del lenguaje. Los autores demostraron que alrededor del 20% de los símiles de la web eran irónicos. Sin embargo, su trabajo no se puede utilizar para detectar la ironía de forma general ya que utilizaba las propias estructuras del lenguaje. Tradicionalmente, el lenguaje figurado se ha intentado detectar explorando las características superficiales de los textos. Por una parte, existen estudios que intentan detectar el lenguaje figurado teniendo en cuenta el orden sintáctico, las propiedades léxicas o los elementos afectivos que componen el texto (Reyes, Rosso, y Buscaldi, 2012; Reyes, Rosso, y Veale, 2013). Por otra parte, otros trabajos se centran en investigar como los hashtags de Twitter se emplean para remarcar una intención figurativa en el mensaje transmitido, en especial para la expresión de la ironía o sarcasmo (Sulis et al., 2016). El interés que despierta la tarea de la detección de la polaridad así como el impacto que tiene sobre ésta el lenguaje figurado motivó en 2015 una tarea en la competición internacional para el la evaluación semántica (*Semantic Evaluation - (SemEval)*) (Ghosh et al., 2015).

Quince equipos participaron en la tarea 11 de SemEval 2015 que fue abordada siguiendo múltiples perspectivas. La mayor parte de los participantes plantearon soluciones super-

¹<http://alt.qcri.org/semeval2015/task11/>

visadas para intentar resolver la tarea, predominando dos modelos de aprendizaje automático: las Máquinas de Soporte Vectorial (MSV) y los Modelos Regresión. Dichos modelos se entrenaron utilizando un conjunto de características cuidadosamente seleccionadas para esta tarea como pueden ser: n-gramas de caracteres, n-gramas de palabras, valores extraídos de distintos léxicos, etc.².

Nuestro trabajo pretende extender la aproximación presentada por Hernández et al., (2015), en la cual se abordó la tarea incorporando recursos externos adicionales. Los autores proponen representar un tweet mediante un conjunto de valores de características extraídas de recursos léxicos externos que modelan tanto las emociones como la información psicolingüística contenida en un tweet. Asimismo, el trabajo de Sulis et al., (2016) presenta un análisis de la distribución y correlación de un conjunto de características psicolingüísticas y emocionales extraídas de recursos léxicos para realizar la clasificación de tweets irónicos y sarcásticos.

Sin embargo, a diferencia de los citados trabajos sobre el estudio de las emociones en el lenguaje figurado, en este artículo presentamos un estudio exhaustivo sobre la capacidad de diferentes recursos léxicos de emociones para predecir la polaridad del conjunto de datos de Twitter de la tarea 11 de SemEval2015 detallando cómo afectan estos recursos a los tweets que contienen lenguaje figurado y lenguaje literal.

3 Metodología

En este apartado describiremos la metodología que empleamos para el estudio del impacto de ciertos recursos léxicos sobre la detección de la polaridad.

Se ha trabajado con diferentes recursos: LIWC, EmoLex y Smilies, que se detallarán en el siguiente apartado. Los recursos que hemos estudiado almacenan distintos niveles de información respecto a las palabras. El nivel más básico es la información sobre si una palabra es “positiva” o “negativa” aunque también incluyen otras categorías que indican que emociones están vinculadas a las palabras. Primeramente, se ha realizado una serie de experimentos para determinar la polaridad de los tweets utilizando únicamente las

categorías de “positivo” y “negativo” de dichos recursos por separado y a continuación para todas las categorías de los recursos.

El procedimiento que hemos llevado a cabo para evaluar el impacto de cada recurso sobre la detección de la polaridad consistió en una vez tokenizados los datos de entrenamiento y test, desarrollar un estudio ablativo que evalúa cómo el uso de diferentes técnicas, como la bolsa de palabras (*Bag of words (BOW)*) o *Term frequency – Inverse document frequency (Tf-Idf)*, así como los recursos para representar un tweet, afectan a la calidad de la clasificación. Este proceso se explicará en detalle en el apartado 5.

Independientemente de la representación elegida, se ha entrenado un sistema de clasificación automático para inferir la polaridad utilizando la librería scikit-learn (Pedregosa et al., 2011). Teniendo en cuenta que el sistema debía predecir un valor de polaridad continuo, se ha empleado una Máquina de Soporte Vectorial adaptada para regresión (MSVR).

Este estudio se ha realizado tanto a nivel de todo el corpus, como a nivel de los distintos tipos de lenguaje figurado presentes en este corpus. Para dividir los tweets entre aquellos que contienen lenguaje figurado o lenguaje literal utilizamos los hashtags, asumiendo que el usuario etiqueta su propio tweet con el tipo de lenguaje empleado facilitando su comprensión, siguiendo la aproximación presentada por Sulis et al., (2016).

3.1 Recursos

Se han utilizado varios recursos para obtener los sentimientos y emociones de los tweets. Estos recursos son los siguientes:

NRC Word-Emotion Association Lexicon (EmoLex) (Mohammad y Turney, 2010): El recurso NRC Emotion Lexicon es una lista de palabras en inglés con sus correspondientes asociaciones con las ocho emociones básicas de Plutchik (Plutchik, 1980): ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y el disgusto (*anger, fear, anticipation, trust, surprise, sadness, joy y disgust*) y dos sentimientos: positivo y negativo (*negative y positive*). Este recurso fue manualmente anotado. Si la palabra pertenece a la categoría se indica con un 1, en caso contrario con un 0. En este recurso podemos encontrar 14,182 palabras etiquetadas. En la Tabla 1

²Para más información ver el artículo de Gosh et al., (2011).

se muestra un ejemplo de como se codifica la información en este recurso.

| Palabra | Categoría | Asociación |
|---------|--------------|------------|
| dark | anger | 0 |
| dark | anticipation | 0 |
| dark | disgust | 0 |
| dark | fear | 0 |
| dark | joy | 0 |
| dark | negative | 0 |
| dark | positive | 0 |
| dark | sadness | 1 |
| dark | surprise | 0 |
| dark | trust | 0 |

Tabla 1: EmoLex: representación de la palabra *dark*

Linguistic inquiry and word count (LIWC) (Pennebaker, Francis, y Booth, 2001): Este recurso le asocia a cada palabra una serie de categorías. En total hay un conjunto de 64 categorías diferentes y se muestra la asociación para un total de 4485 palabras.

Smilies (Suttles y Ide, 2013): También se utilizó este recurso de *smilies* que clasifica 176 *smilies* diferentes según la emoción asociada a los mismos. En este trabajo, en lugar de asociar un smile a una de las seis emociones básicas definidas en la teoría de Ekman (alegría, ira, miedo, asco, sorpresa, tristeza) (Ekman, 1972), los autores utilizan los ocho tipos de emociones avanzadas definidas en la teoría de Plutchik. A partir de estos ocho tipos de emociones y utilizando una lista con los hashtags emocionales más frecuentes, los autores de este recurso seleccionaron quince categorías para etiquetar los diferentes *smilies*: feliz, risueño, amoroso, enfadado, triste, llanto, disgustado, sorpresa, beso, guiño, lengua, escéptico, indeciso, avergonzado y maligno (*happy, laugh, love, annoyed, sad, cry, disgust, surprise, kiss, wink, tongue, skeptical, indecision, embarrassed y evil*). Se puede apreciar un ejemplo en la Tabla 2.

4 Descripción de la tarea

En la tarea 11 de SemEval 2015 se utilizó un corpus de lenguaje figurado extraído de la red social Twitter el cual presenta un gran número de ironías, sarcasmos o metáforas, sin embargo, no se puede garantizar que se mani-

| Emoticono | Emoción |
|-----------|---------|
| :D | LAUGH |
| :@ | SAD |
| ;-) | WINK |
| 3:-) | EVIL |

Tabla 2: Smilies: Clasificación de *smilies*

fieste cualquiera de estos fenómenos en cada uno de los tweets.

La ironía y el sarcasmo normalmente se utilizan para criticar o burlarse y, por lo tanto, sesgar la percepción del sentimiento hacia un valor negativo, por lo que no es suficiente para un sistema determinar simplemente si el sentimiento de un tweet dado es positivo o negativo atendiendo únicamente al lenguaje literal presente en el mismo.

Los organizadores de la tarea proporcionaron el corpus etiquetado siguiendo una escala de 11 puntos que oscilaban desde -5 (muy negativo, para tweets con significados muy críticos) a 5 (muy positivo, tweets con significados muy optimistas). El punto cero de esta escala se utiliza para determinar los tweets neutros.

Los sistemas se evaluaron utilizando dos métricas la distancia coseno y el Error Cuadrático Medio (ECM), ambas métricas apropiadas para problemas de regresión. Por simplicidad computacional se ha decidido evaluar los sistemas aquí presentados únicamente el Error Cuadrático Medio, que define el error cometido por el vector de predicciones $\hat{Y} \in \mathbb{R}^n$ con respecto al vector con los valores correctos para esas n muestras $Y \in \mathbb{R}^n$ siguiendo la siguiente fórmula:

$$ECM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (1)$$

Para más información acerca de los detalles de la tarea se pueden consultar las actas de la tarea (Ghosh et al., 2015).

El reto de nuestro sistema es determinar cómo un recurso o la combinación de estos, pueden influir en la capacidad del sistema de clasificación que desarrollemos para predecir el sentimiento presente en un tweet. Validaremos esta metodología propuesta sobre el conjunto de datos que describiremos en detalle en el siguiente apartado.

4.1 Corpus

El conjunto de datos empleado en la tarea 11 de SemEval fueron recolectados a través de la API Twitter4J, que soporta la recolección de tweets en tiempo real mediante la búsqueda consultas. Se utilizaron consultas de hashtags como *#sarcasm*, *#sarcastic* y *#irony* para obtenerlo.

Este conjunto de datos fue recogido durante 4 semanas, del 1 de junio al 30 de junio de 2014. Se eliminaron aquellos tweets que no cumplieran una serie de condiciones como por ejemplo, no contener al menos 30 caracteres sin incluir el hashtag. Asimismo, se filtró también únicamente aquellos tweets que estuvieran escritos en inglés, por lo tanto se trata de una tarea monolingüe.

Cada tweet fue etiquetado por siete anotadores, tres de los cuales eran hablantes nativos de inglés y el resto de los anotadores eran competentes en el idioma. A todos ellos se les pidió asignar una puntuación que oscilaba desde -5 a 5, donde 0 es el valor neutro para aquellos tweets que tienen el mismo valor negativo que positivo.

El sentimiento general de cada tweet se calculó como una media ponderada de las siete puntuaciones donde las puntuaciones de los nativos del inglés valían el doble.

El conjunto de tweets de entrenamiento y test está compuesto por 8000 y 4000 tweets respectivamente. Un ejemplo de varios tweets se muestra en la siguiente tabla:

5 Experimentación

Nuestra experimentación se ha llevado a cabo en varias fases. En la primera de ellas se ha estudiado el impacto de los recursos sobre el conjunto completo de tweets y a continuación se ha evaluado el grado de impacto para cada uno de los diferentes conjuntos de tweets con lenguaje figurado en el corpus.

Se ha tokenizado el corpus utilizando la librería NLTK (Bird, Klein, y Loper, 2009), se han eliminado las palabras que no aportan información discursiva (*stopwords*) y se ha convertido todo el texto que no fueran *smilies* a minúsculas. A continuación, se han obtenido las representaciones BOW y Tf-Idf de los tweets utilizando la librería scikit-learn (Pedregosa et al., 2011).

Para los recursos de EmoLex y LIWC se han creado diccionarios que representan eficientemente la información. Cada entrada del diccionario corresponde con una categoría

| Tweet | Polaridad |
|----------------------------------------------------------------------------------------------------------------------|-----------|
| “@erikaekengren: From 50 to 100 degrees in less than a week #kansas” #cantwait #sarcasm | -3 |
| Updated my router and it froze. Now I can’t access the internet to google a solution. #irony #thankfulforsmartphones | -3.48 |
| I’ve had a lot of wake up calls in my day, but I’ve always been good at hitting the snooze #metaphor #nailedit | 0.22 |

Tabla 3: Ejemplos extraídos del corpus de la tarea 11 de SemEval 2015. En la columna “Polaridad” se especifica la polaridad con la que se puntuó de media el tweet mostrado como ejemplo

emocional y en ella se almacenan las palabras que forman parte de dicha categoría. Para utilizar el recurso de *smilies* se ha tenido que crear un tokenizador ad hoc con cada una de las expresiones regulares necesarias para identificar todos los *smilies*.

Una vez se han obtenido los diccionarios de cada recurso, se han elaborado representaciones vectoriales de las muestras de entrenamiento y test. Cada uno de estos vectores indican para cada tweet, el número de veces que aparece una palabra de las categorías que se tienen en el diccionario. De esta manera se tiene un vector diferente para cada recurso que posteriormente se combinarán para realizar la experimentación. La combinación de estos vectores consiste simplemente en agregar al final del vector del primer recurso el vector del segundo.

Utilizando estas estructuras, se han realizado varios experimentos sobre todo el conjunto del corpus. Primero utilizando únicamente las categorías “positivo” y “negativo”, y, a continuación, utilizando todas las categorías disponibles de cada uno de los recursos.

En la última columna de la Tabla 4 se muestran los resultados alcanzados tras combinar diferentes recursos para entrenar una SVR y calculando el resultado utilizando el ECM. Además, para poder comparar los re-

| Recurso | #irony | #sarcasm | #not | otros | total |
|------------------------------|---------------|---------------|---------------|---------------|---------------|
| BOW + EmoLex p/n | 0,8466 | 0,5790 | 5,8184 | 6,8359 | 4,6025 |
| TF-IDF + EmoLex p/n | 0,8781 | 0,5794 | 5,9199 | 6,9498 | 4,6825 |
| BOW + EmoLex Todas | 0,8459 | 0,5787 | 5,8136 | 6,8269 | 4,5972 |
| TF-IDF + EmoLex Todas | 0,8770 | 0,5800 | 5,9101 | 6,9374 | 4,6744 |
| BOW + LIWC p/n | 0,8471 | 0,5817 | 5,8257 | 6,8421 | 4,6074 |
| TF-IDF + LIWC p/n | 0,8754 | 0,5821 | 5,9148 | 6,9451 | 4,6789 |
| BOW + LIWC Todas | 0,8331 | 0,5780 | 5,8076 | 6,8187 | 4,5897 |
| TF-IDF + LIWC Todas | 0,8583 | 0,5782 | 5,8972 | 6,9265 | 4,6628 |
| BOW + EmoLex + LIWC p/n | 0,8451 | 0,5788 | 5,8203 | 6,8421 | 4,6022 |
| TF-IDF + EmoLex + LIWC p/n | 0,8756 | 0,5796 | 5,9173 | 6,9460 | 4,6796 |
| BOW + EmoLex + LIWC Todas | 0,8338 | 0,5759 | 5,7883 | 6,7972 | 4,5756 |
| TF-IDF + EmoLex + LIWC Todas | 0,8586 | 0,5755 | 5,8778 | 6,9054 | 4,6487 |
| BOW + <i>Smilies</i> | 0,8484 | 0,5817 | 5,8227 | 6,8407 | 4,6360 |
| TF-IDF + <i>Smilies</i> | 0,8748 | 0,5813 | 5,9041 | 6,9355 | 4,6719 |
| <i>Baseline: Naïve Bayes</i> | - | - | - | - | 5,6720 |

Tabla 4: Resultados obtenidos empleando la métrica ECM evaluando cada uno de subconjuntos de tipos de lenguaje figurado que podemos encontrar en este corpus. Señalamos con la abreviatura “p/n” aquellas representaciones en las que únicamente se emplean las características “positiva” y “negativa” del recurso en cuestión, mientras que aquellos recursos en los que empleamos todas las categorías se señalan como “Todas”

sultados se incluye el ECM de un sistema de control (*Baseline*) Naïve Bayes entrenado con una aproximación de bolsa de palabras facilitado por la organización de la tarea.

Dada esta experimentación preliminar, la segunda fase ha consistido en determinar el grado de impacto de cada uno de los recursos sobre las diferentes expresiones de lenguaje figurado más habituales en este corpus. Para ello se han diferenciado un total de cuatro grupos de tweets, en función de la aparición de los siguientes hashtags: *#irony* (765 tweets), *#sarcasm* (536 tweets), *#not* (981 tweets) y *otros* (1718 tweets). Si un tweet tiene varios hashtags pertenecerá a ambos conjuntos.

En este último conjunto, *otros* se agrupan aquellos los tweets que no forman parte de los tres primeros grupos y que por lo tanto asumimos que se trata de lenguaje literal. Esta separación se ha hecho sobre los datos del test del corpus, aceptando que el usuario ha empleado el hashtag para auto-etiquetar el tipo de lenguaje que su tweet contenía. La organización de SemEval reportó resultados sobre un conjunto de test con metáforas, pe-

ro no aparece el hashtag *#metaphor* en el test y no hemos podido llevar a cabo una separación automática de este conjunto de tweets. Por lo que en el conjunto *otros* aparecerán metáforas, que según las actas de la tarea (Ghosh et al., 2015) es una de las formas de lenguaje figurado más difíciles de clasificar.

Una vez separados los tweets, se ha llevado a cabo la segunda fase de la experimentación en la que se han utilizado como datos de entrenamiento todos los tweets del conjunto de entrenamiento de la tarea, pero como datos de test se ha utilizado cada uno de los grupos acabamos de describir. Al igual que en la primera fase de experimentos, se ha probado cada recurso por separado así como la combinación de ellos. En la Tabla 4 se muestran para cada uno de los grupos los resultados obtenidos con los diferentes recursos.

6 Análisis de los Resultados

Como ha podido comprobarse en los resultados de la experimentación que hemos presentado en el apartado anterior, la inclusión de nuevos recursos nos conduce a mejorar significativamente el comportamiento de los mo-

delos que entrenemos. Sin embargo, el ECM varía considerablemente en función del subconjunto de lenguaje que estemos considerando. Todos los modelos que hemos presentados consiguen mejorar el modelo de control o *baseline*, lo cual nos indica que, efectivamente, los recursos léxicos que aportan información acerca de las emociones ayudan a mejorar la predicción del sentimiento comunicado en un tweet. Cabe destacar que, cuando incluimos la información respecto a todas las emociones disponibles en un recurso léxico, y no únicamente las categorías “positivas” y “negativas”, conseguimos mejorar el comportamiento del modelo entrenado independientemente de si estamos ante lenguaje figurado o literal. Además, a pesar de que estamos ante un corpus con una baja frecuencia de *smilies* y por lo tanto la cobertura del léxico *smilies* es escasa, este recurso también consigue mejorar el sistema. La aproximación que incluye la bolsa de palabras y los recursos afectivos EmoLex y LIWC con todas las emociones obtiene resultados satisfactorios aunque en el caso de los tweets en los que aparece el *hashtag* *#irony* la aproximación no utiliza el recurso EmoLex es la que presenta un mejor comportamiento mientras que en el caso de los tweets con el *hashtag* *#sarcasm* el mejor sistema utiliza una representación de palabras basada en Tf-Idf. Sin embargo, la diferencia entre estos sistemas no es significativa y podemos concluir que el mejor sistema es el que emplea las características “BOW + EmoLex + LIWC Todas”. El mejor sistema participante en la tarea, ClaC (Ozdemir y Bergler, 2015), obtuvo un EMC 2.117 para lo cual se desarrolló un complejo proceso para la extracción de la polaridad de las palabras en función del contexto en el que aparezcan. Nuestro sistema comparte características con el sistema denominado ValenTo (Farias et al., 2015), aunque en este trabajo se incluyen más recursos que pretendemos explorar en trabajos futuros.

El estudio de cómo se distribuye el error entre los distintos subgrupos de tweets ha arrojado resultados sorprendentes: la mayor parte del error se concentra en el subgrupo que no contenía ningún tipo de *hashtags* del conjunto de *hashtags* estudiando (*#irony*, *#sarcasm*, *#not*), lo cual puede explicarse porque no se puede analizar de modo independiente el impacto de la metáfora sobre el conjunto *otros*. No obstante, se requiere un

estudio pormenorizado de los tweets y la polaridad para explicar este fenómeno.

7 Conclusiones y Trabajo Futuro

En este trabajo, se ha presentado un estudio sobre la capacidad de distintos recursos léxicos de emociones para predecir la polaridad de un conjunto de datos extraídos de Twitter. Se ha visto el impacto de cada uno de ellos sobre las distintas formas de lenguaje figurado como la ironía y el sarcasmo y la importancia de desarrollar técnicas capaces de representar esa información para clasificar el sentimiento que el autor emitió en un texto. Se han obtenido unos resultados que apuntan a que la inclusión de información relativa a las emociones ayuda a clasificar correctamente la polaridad tanto a nivel global como a nivel del lenguaje figurado o literal.

Como trabajo futuro, se pretende extender el estudio a distintos algoritmos de aprendizaje automático para comprobar cómo afecta a su comportamiento la inclusión de información emocional recurrente o ruidosa, puesto que las MSV son capaces de descartar aquellas muestras no significativas para la clasificación y son más robustas respecto al ruido. Asimismo, se evaluará de forma sistemática la contribución de la representación de palabras y de nuevos recursos léxicos en la detección de la polaridad de un tweet. Igualmente, se estudiará cómo se puede aumentar la cobertura de los recursos léxicos, es decir el número de palabras que encontramos en el diccionario, utilizando técnicas como por ejemplo la corrección automática del texto, para eliminar, en la medida de lo posible los errores gramaticales presentes en Twitter.

Agradecimientos

Este trabajo se ha desarrollado en el marco del proyecto de investigación SomEMBED (TIN2015-71147-C2-1-P) del Ministerio de Economía y Sostenibilidad (MINECO). Asimismo, el trabajo de la segunda autora ha sido financiado a través del Programa de Ayudas de Investigación y Desarrollo de la Universitat Politècnica de València (PAID 2015).

Bibliografía

- Bird, S., E. Klein, y E. Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media, Inc.
- Ekman, P. 1972. Universals and cultural differences in facial expressions of emo-

- tions. *Nebraska Symposium on Motivation*, 19:207–283.
- Farias, D. I. H., E. Sulis, V. Patti, G. Ruffo, y C. Bosco. 2015. Valento: Sentiment analysis of figurative language tweets with irony and sarcasm. *SemEval-2015*, página 694.
- Ghosh, A., L. G., T. Veale, P. Rosso, E. Shutova, J. Barnden, y A. Reyes. 2015. SemEval-2015 Task 11: Sentiment Analysis of Figurative Language in Twitter. *Proc. 9th Int. Workshop on Semantic Evaluation (SemEval 2015)*, páginas 470–478.
- Hernández, I. y P. Rosso. 2016. Irony, sarcasm, and sentiment analysis. En F. Pozzi E. Fersini E. Messina, y B. Liu, editores, *Sentiment Analysis in Social Networks*. Morgan Kaufmann, capítulo 7, páginas 113–128.
- Kiritchenko, S., S. Mohammad, y M. Salameh. 2016. Semeval-2016 task 7: Determining sentiment intensity of english and arabic phrases. En *Proceedings of the International Workshop on Semantic Evaluation (SemEval)*, San Diego, California, June.
- Mohammad, S. M. y P. D. Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. En *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, páginas 26–34. Association for Computational Linguistics.
- Ozdemir, C. y S. Bergler. 2015. Clac-sentipipe: Semeval2015 subtasks 10 b, e, and task 11. En *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, páginas 479–485.
- Pang, B. y L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.
- Pang, B., L. Lee, y S. Vaithyanathan. 2002. Thumbs up?: Sentiment classification using machine learning techniques. En *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, páginas 79–86. Association for Computational Linguistics.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, y others. 2011. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pennebaker, J. W., M. E. Francis, y R. J. Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71:2001.
- Plutchik, R. 1980. Emotion: Theory, research, and experience. *Theories of Emotion*, 1.
- Reyes, A., P. Rosso, y D. Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.
- Reyes, A., P. Rosso, y T. Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, 47(1):239–268.
- Rosenthal, S., P. Nakov, S. Kiritchenko, S. Mohammad, A. Ritter, y V. Stoyanov. 2016. SemEval-2015 Task 10: Sentiment Analysis in Twitter. *Proc. 9th Int. Workshop on Semantic Evaluation (SemEval 2015)*.
- Sulis, E., I. Hernández, P. Rosso, V. Patti, y G. Ruffo. 2016. Figurative Messages and Affect in Twitter: Differences Between #irony, #sarcasm and #not. *Knowledge-Based Systems*, 108:132–143.
- Suttles, J. y N. Ide. 2013. Distant Supervision for Emotion Classification with Discrete Binary Values. *Computational Linguistics and Intelligent Text Processing*, páginas 121–136.
- Turney, P. D. 2002. Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews. En *Proceedings of the 40th annual meeting on association for computational linguistics*, páginas 417–424. Association for Computational Linguistics.
- Veale, T. y Y. Hao. 2007. Learning to understand figurative language: from similes to metaphors to irony. *Proceedings of CogSci 2007*.