Procesamiento del Lenguaje Natural, Revista nº 62, marzo de 2019, pp. 69-76 recibido 22-11-2018 revisado 08-01-2019 aceptado 06-02-2019

Análisis comparativo de las características computacionales en los sistemas modernos de análisis de sentimiento para el español

Comparative analysis of the computational characteristics in modern sentiment analysis systems for Spanish

Edgar Casasola,[1] **Alejandro Pimentel,**[2] **Gerardo Sierra,**[2]
Eugenio Martínez Cámara,[3] **Gabriela Marín**[1]
[1]Universidad de Costa Rica
[2]Universidad Nacional Autónoma de México
[3]Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI). Universidad de Granada, España.
edgar.casasola@ucr.ac.cr, {apimentela, gsierram}@iingen.unam.mx,
emcamara@decsai.ugr.es, gabriela.marin@ucr.ac.cr

Resumen: Existen múltiples sistemas para análisis de sentimiento con diseños heterogéneos y niveles variados de desempeño. En este artículo se presenta un modelo de generación de especificaciones computacionales de los sistemas para identificación de polaridad tendientes a facilitar comparaciones más profundas y detalladas de las técnicas que se utilizan. Buscamos crear conciencia entre los investigadores de la información necesaria para la construcción de especificaciones que permitan replicar sistemas. A su vez, discutimos las dificultades que se tiene al evaluar y al hacer comparaciones congruentes entre sistemas e ir más allá del resultado que se puede obtener para una tarea específica sobre un conjunto de datos particular. Estamos convencidos de que una estructuración completa y clara de todos los procesos y de los ajustes a que son sometidos los trabajos presentados en las competencias es crucial para enriquecer el conocimiento del uso de estrategias hacia la mejora general de los sistemas.
Palabras clave: Análisis de sentimiento, especificación de sistemas, revisión de literatura, corpus en español

Abstract: There are multiple systems for sentiment analysis with heterogeneous designs and varying levels of performance. In this paper, we propose a model of computational specifications of polarity identification systems. The model makes easier the comparison of the different techniques used. We seek to create awareness among researchers of the necessary information for the elaboration of specifications that allow the replication of systems. Additionally, we discuss the difficulties of evaluating and conducting consistent comparisons among systems, and going beyond the result that can be obtained for a specific task on a particular data set. We are convinced that a complete and clear framework that encompasses all the modules of the systems that participate in competitions is crucial to enrich the knowledge for improving the state-of-the-art in sentiment analysis.
Keywords: Sentiment analysis, systems specification, literature review, corpus in Spanish

1 Introducción

En la Web 2.0 la comunidad de usuarios conectados a nivel mundial a través de plataformas como *YouTube*, *Facebook*, y *Twitter* están produciendo información que posibilita la creación de nuevos tipos de aplicaciones de inteligencia colectiva (O'Reilly y Battelle, 2009). Una de estas aplicaciones es el análisis de sentimiento, y una tarea específica particular que ha recibido mucha atención ha sido la identificación automática de la polaridad de textos cortos publicados en estas redes sociales. Sin embargo, el desarrollo de estas tecnologías del lenguaje varía entre idio-

ISSN 1135-5948. DOI 10.26342/2019-62-8
© 2019 Sociedad Española para el Procesamiento del Lenguaje Natural

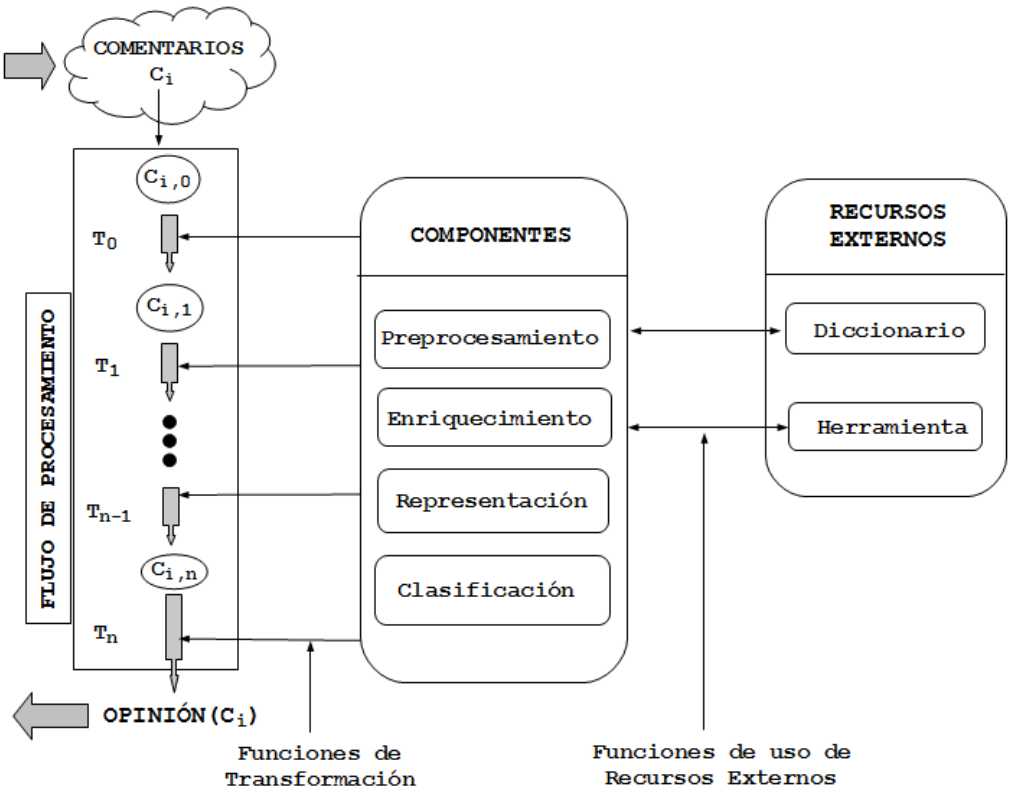

Figura 1: Diagrama general del modelo conceptual

mas y en el año 2012 se hizo evidente la existencia de una brecha entre el nivel alcanzado para idiomas como el inglés con respecto al español (Melero et al., 2012). Este trabajo se desarrolla orientado a este tipo de sistemas para el idioma español con el interés particular en disminuir esa brecha entre idiomas. En es misma línea y con el fin de promover la interacción entre grupos de investigación de diferentes países de habla hispana se desarrolló un modelo base que proporciona una especificación conceptual uniforme para discutir y poder comparar las implementaciones de sistemas de análisis semántico. Las particularidades de este modelo serán presentadas en la sección 2.

En su primera versión, el modelo se enfocó en los sistemas para análisis de sentimiento y particularmente para los que llevan cabo identificación de polaridad a partir de comentarios de texto en idioma español. El modelo consiste de una abstracción de referencia centrada en los componentes generales de los sistemas, una descripción conceptual mediante el uso de una ontología en lenguaje OWL de los términos utilizados y la relación entre ellos, e incluyen un conjunto de plantillas para la especificación de los componentes computacionales de los sistemas. En este trabajo mostramos el uso de esas plantillas para llevar a cabo una revisión de los sistemas presentados en el taller TASS 2017 (Díaz-Galiano et al., 2018) de la Sociedad Española para Procesamiento de Lenguaje Natural. TASS (Taller de Análisis de Semántico en la SEPLN) es un taller que se lleva a cabo desde el año 2012 con la finalidad de fomentar el desarrollo de técnicas para el tratamiento de opinión en español. Por lo tanto, se utilizaron las plantillas del modelo común para describir las especificaciones de cada uno de los sistemas.

2 Modelo base para la especificación de sistemas los sistemas

Para efectos de análisis se utiliza como base el modelo **SAM** que se muestra en la Figura 1 (Casasola-Murillo, 2018) este modelo conceptual permite especificar en términos computacionales las características implementadas en los sistemas para análisis de sentimiento.

En la Figura 1 se muestra la relación existente entre los elementos del modelo. En el lado izquierdo del diagrama se puede apreciar el *flujo de procesamiento* típico de aplicaciones de procesamiento de lenguaje natural. Sin embargo, el modelo incluye la especificación de los componentes computacionales involucrados en este procesamiento. En la parte central del diagrama se pueden observar los diferentes componentes del modelo según su rol dentro del sistema. Se incluyen: los componentes de *preprocesamiento del texto*, los componentes utilizados para *enriquecimiento* que agregan información extra al texto de cada comentario analizado, los componentes de *representación* que modifican la estructura de representación del comentario con fines computacionales, y finalmente los componentes utilizados para llevar a cabo la clasificación de los comentarios. Adicionalmente, a la derecha del diagrama, se muestran los *recursos externos*, tales como herramientas computacionales y diccionarios o lexicones. Este trabajo se centra en la especificación de estos componentes de preprocesamiento, enriquecimiento de texto, representación, clasificación y los recursos externos reportados en cada artículo de investigación analizado. Para el análisis de los componentes de los sistemas se utilizaron las plantillas para especificación de sistemas del modelo SAM. Uno de los aportes más relevantes del trabajo de Casasola-Murillo (2018) fue la construcción de una plantilla taxonómica para especificación de sistemas de análisis de sentimiento. El modelo **SAM** permite construir una especificación utilizando un grupo de plantillas para los diferentes componentes según su rol. Como parte del modelo se incluye un documento con la especificación detallada del modelo y está acompañado con una ontología en formato OWL con la descripción y organización jerárquica de los conceptos.

3 Metodología

En la edición de TASS 2017 se propusieron dos tareas: la clasificación de polaridad a nivel de documento; y la clasificación de polaridad a nivel de aspecto. En la primera tarea participaron 10 equipos con el corpus InterTASS (Díaz-Galiano et al., 2018). Esa tarea corresponde a la clasificación de polaridad en un *tweet*. Para dicha tarea se ofreció el recurso del InterTASS, un corpus de más de 2,000 *tweets* anotados con cuatro categorías de opinión: *positive*, *neutral*, *negative* y *none*. Este trabajo se limita al análisis de los sistemas que reportaron resultados para esta tarea particular.

ID	Sistema	M-F1	Acc.
2	ELiRF-UPV	0.493	0.607
9	RETUYT	0.471	0.596
7	ITAINNOVA	0.461	0.576
6	Jacerong	0.460	0.602
1	INGEOTEC	0.526	0.595
3	Tecnolengua	0.456	0.582
10	SINAI	0.442	0.575
5	LexFAR	0.432	0.541
4	OEG	0.367	0.386
8	GSI	0.327	0.558

Tabla 1: Resultados obtenidos por cada sistema en TASS 2017

En la Tabla 1 se presentan los 10 sistemas con los resultados obtenidos. De igual manera se presenta un identificador numérico que será utilizado como índice de referencia para cada trabajo en lo que resta del artículo. Si bien hubo sistemas que obtuvieron mejores resultados al ser evaluados con otro corpus conocido como el corpus general *General Corpus*, en la tabla comparativa se eligieron los sistemas que presentaron evaluaciones utilizando el corpus InterTASS ya que para este corpus hubo una mayor participación.

La especificación del modelo y las plantillas fueron entregadas a un grupo de evaluadores previo a la especificación de artículos. Posteriormente, cada artículo fue revisado por al menos tres evaluadores, quienes revisaron un máximo de dos artículos cada uno. El perfil de estos evaluadores era el de estudiantes de Maestría en Computación y Lingüística con conocimiento de Procesamiento de Lenguaje Natural. Cada evaluador generó una especificación del sistema descrito utilizando las plantillas suministradas. Las plantillas fueron llenadas en forma individual y posteriormente se unificaron por consenso entre los tres evaluadores y el investigador a cargo. Posteriormente el proceso de especificación para todos los trabajos fue replicado por un investigador internacional y mediante la comparación de ambos resultados se obtuvieron los resultados que aquí se reportan. Cabe mencionar que todas las especificaciones fueron generadas a partir de los textos de los artículos publicados.

Componentes de preprocesamiento	Artículos									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
acentos	✓				✓	✓				
recodificación						✓				
emoticons	✓		✓	✓		✓				
hashtag		✓	✓			✓			✓	
mayúsculas	✓		✓		✓		✓		✓	✓
mención	✓	✓	✓	✓	✓	✓	✓		✓	
negación	✓		✓	✓	✓	✓			✓	
numeral	✓	✓		✓						
diccionario			✓	✓	✓	✓			✓	
puntuación		✓	✓		✓					
repetición	✓			✓	✓	✓			✓	✓
stopwords	✓			✓	✓		✓		✓	
url	✓	✓	✓	✓	✓	✓	✓		✓	
fechas		✓								

Tabla 2: Componentes de preprocesamiento reportados en cada trabajo

4 Descripción de los componentes de cada sistema

Para cada grupo de componentes se presentan los resultados comparativos según la organización propuesta en el modelo.

4.1 Pre-procesamiento

En la Tabla 2 se muestran las diferentes prácticas que se utilizaron en los sistemas que participaron en el TASS 2017.

El preprocesamiento se refiere a las transformaciones que se efectúan sobre los datos crudos con el objetivo de obtener algo uniforme y sin elementos que se puedan considerar ruidosos para el mejor manejo del texto.

Como ejemplos podemos observar la eliminación de puntuación, caracteres y ciertas palabras que se consideran sin contenido ni beneficio en su consideración; la normalización de diferentes formas de escribir expresiones informales, la eliminación de números y ligas, entre otras.

La etapa de preparar los datos para un procesamiento más sencillo y efectivo no solo se ha vuelto una práctica básica, sino que ha llegado a tener un carácter tan cotidiano que muchas veces no se le da la importancia y especificación suficientes y necesarios para llevar a cabo experimentación empírica. En la Tabla 2 se hace evidente cómo se le da un peso considerable a elementos propios del área de análisis de sentimiento y que se han reportado con buenos resultados como la consideración de la negación. O datos característicos del corpus (*tweets*) como las menciones o los *hashtags*. Sin embargo, la mayoría de las consideraciones que se hacen para el pre-procesamiento se dejan sin detallar o simplemente no se mencionan aunque se hayan tomado en cuenta.

Cada *framework* tiene un conjunto de acciones de pre-procesmiento propias de las que se puede hacer uso. Esto ayuda a establecer una línea de comparación entre aquellos sistemas que hagan uso de dichos *frameworks*. No obstante, muchos de los trabajos anexan capas extra de pre-procesamiento, no utilizan los *frameworks* o no reportan con exactitud los parámetros que utilizan.

Antes no se tenía una herramienta para capturar los diferentes módulos y procesos que se llevaban a cabo en cada una de las etapas de un sistema de SA. Gracias al modelo propuesto es posible dar una estructuración a este tipo de características, tanto como una herramienta para la normalización de la información reportada como para la concientización de los autores para la retroalimentación de la información y la eficiente réplica experimental.

4.2 Enriquecimiento

El enriquecimiento se refiere al anexo de datos mediante diferentes técnicas que resultan en un aporte de información que se puede obtener de la misma información mediante sistemas que hacen uso de conocimiento de dominio. Por ejemplo, los etiquetadores sintácticos, lematizadores y las técnicas de ponderación de la importancia de las palabras como el *tf-idf* entre otros.

Con el modelo de comparación que estamos utilizando, podemos observar para este módulo ciertos elementos populares a través

Técnica de enriquecimiento	Artículos									
	1	2	3	4	5	6	7	8	9	10
categoría gramatical		✓	✓	✓	✓	✓				
parsing				✓						
entidad		✓								
frecuencia			✓		✓	✓			✓	
lema		✓		✓	✓		✓			
stemming	✓			✓						
stemming'	✓									
ponderación	✓	✓				✓				✓
polaridad				✓	✓	✓	✓	✓	✓	
embedding									✓	
aprendizaje a distancia	✓	✓								

Tabla 3: Técnicas de enriquecimiento reportadas en cada trabajo

de los diferentes sistemas como es el caso del etiquetado de partes de la oración (POS), la lematización y el enriquecimiento mediante el uso de lexicones de polaridad. El aprendizaje a distancia o *distance supervision* (Mintz et al., 2009) se menciona como técnica de enriquecimiento de los conjuntos de datos de entrenamiento.

El modelo comparativo que estamos utilizando es particularmente valioso para notar técnicas compartidas entre diferentes sistemas. En la Tabla 3 se puede observar que el enriquecimiento mediante lexicones de polaridad es una técnica bastante popular, y un elemento utilizado por varios de los sistemas más sobresalientes como el 9, el 7 y el 6. No obstante, el sistema 2, que es el más sobresaliente, no los utiliza, e incluso en su trabajo se puede leer (Hurtado, Pla, y González, 2017):

> "Tampoco se han utilizado lexicones de polaridad debido a que no mejoraban los resultados sobre validación..."

Misma conclusión a la que llegaron Rosá et al. (2017) del sistema 9, aunque ellos si lo utilizaron:

> "En particular, se observa que el aporte del léxico subjetivo es poco relevante."

Esto plantea nuevas implicaciones sobre la eficacia de las técnicas reportadas por trabajos sobresalientes del estado del arte. Los sistemas son una serie de herramientas en conjunto. La aportación que tiene cada una de ellas en el resultado que se obtiene es desconocido, e imposible de averiguar si no se tiene una estructuración de las diferentes opciones que se tienen para cada módulo.

En el estado del arte se reportaron buenos resultados con ayuda de los lexicones de polaridad (Moreno-Ortiz y Hernández, 2017). Pero esto no implicaba que fuera gracias a los lexicones de polaridad que se obtuvieran los resultados sobresalientes. No obstante, se siguen utilizando casi indiscriminadamente sin llegar a hacerse pruebas específicas de la ganancia que están produciendo. Es posible que haya técnicas junto con las que funcionan bien y otras con las que no. Es debido a consideraciones de esta naturaleza por lo que se vuelve evidente la necesidad de construir un mapa de eficacia para cada una de las herramientas que se pueden utilizar.

En contraposición con lo anterior, para el sistema 7, Montañés-Salas et al. (2017) defienden que sus resultados si se vieron beneficiados con la inclusión de información de polaridad:

> "Como se puede observar, la utilización del diccionario afectivo hace que tengamos mejores resultados, aunque el algoritmo también obtiene resultados bastante satisfactorios sin ellos."

En las dos secciones siguientes ahondaremos más acerca de los procesos que se usaron para incluir la información de polaridad y las posibles razones que pueden estar detrás de estas discrepancias en los resultados.

4.3 Representación

La representación se refiere a la configuración en la que se ingresarán los datos para su procesamiento y clasificación. Es decir, qué características en particular se toman en cuenta para definir los vectores de entrada de un sistema computacional.

Método de representación	Artículos									
	1	2	3	4	5	6	7	8	9	10
bolsa de palabras				✓	✓			✓	✓	
n-grama	✓					✓				
q-grama	✓					✓				
skip-grama	✓									
features		✓	✓	✓	✓	✓			✓	
embeddings		✓					✓		✓	✓

Tabla 4: Métodos de representación del texto utilizadas en cada trabajo

En la Tabla 4 se listan una serie de características que cada uno de los sistemas reporta como haber tomado en cuenta. Es posible

observar una tendencia hacia la agrupación en un vector de caraceteríticas (*features*).

Por otro lado, un detalle que resalta a la vista es que los tres sistemas más sobresalientes (2,7 y 9) emplean *word embeddings* para la representación de sus datos. Esto es de particular interés para la comparación de sistemas en general, pues en todos los módulos de estudio, no se encontró otra característica que identificara tan bien a estos tres sistemas.

En adición a lo anterior, es importante resaltar dos experimentos similares que llevaron a cabo los sistemas 2 y 7 referente a la representación de sus palabras en combinación con información de polaridad (Hurtado, Pla, y González, 2017; Montañés-Salas et al., 2017).

Como se mencionó en la sección 4.2, el sistema 2 (Hurtado, Pla, y González, 2017) descartó el uso de diccionarios de polaridad debido a la ausencia de mejoras en el desempeño de dicho enriquecimiento. No obstante, hacen experimentos con *sentiment specific word embeddings* (SS-WE), para los cuales aplican *distant supervision* con heurísticas de emoticones. Esto podría verse como un uso indirecto de información de polaridad para la creación de *embeddings*.

Por su parte, para el sistema 7, Montañés-Salas et al. (2017) obtiene sus propios SS-WE mediante una técnica distinta. Se basa en el uso de varios diccionarios afectivos para introducir sinónimos basados en emociones y con esto reduce el vocabulario del que obtiene *embeddings* especiales.

De sus experimentos al utilizar *word embeddings* entrenados especialmente tomando en cuenta sentimientos, Hurtado, Pla, y González (2017) no obtiene mejoras mientras que Montañés-Salas et al. (2017) sí. Lo que supone que la información que se obtiene de diccionarios afectivos necesita consideraciones especiales para poder ser aprovechada. A continuación, la sección 4.4 presenta el análisis de los componentes de clasificación.

4.4 Clasificación

La clasificación se refiere a los algoritmos que se utilizan para obtener el resultado final de la categoría a la que pertenece la polaridad de los comentarios. En la Tabla 5 se muestran los diferentes sistemas que reportan los participantes del TASS 2017

De manera similar a los componentes de representación, para este tipo de componentes destacan como característica común a los sistemas más sobresalientes (2, 9 y 7) el uso de redes neuronales.

Estrategia de clasificación	**Artículos**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Bayes				✓						
Ensamble	✓	✓				✓		✓	✓	
Red neuronal		✓					✓	✓	✓	✓
Logística			✓			✓				
SVM	✓				✓	✓			✓	✓
Genético	✓									

Tabla 5: Estrategias de clasificación reportadas en cada trabajos

Por otro lado, se puede notar que al igual que con los componentes de representación, se tiene la presencia del sistema 10, sin embargo, como se observa en las tablas, este sistema se encuentra pobremente especificado y no nos es posible analizar por qué, a pesar de tener estas cosas en común con los sistemas más sobresalientes, no tuvo un mejor desempeño.

Un elemento muy interesante dentro de este rubro que no puede ser capturado con el modelo actual es la presencia de indicaciones dentro de los artículos que demuestran una tendencia acerca del mejor funcionamiento de algunos sistemas sobre otros, como se mencionan Moreno-Ortiz y Hernández (2017):

> "We also tried a SVM classifier on the same feature sets, but we consistently obtained poorer results compared to the logistic regression classifier."

Lo anterior se puede traducir como:

> "También tratamos con clasificadores SVM sobre el mismo conjunto de características, pero obtuvimos de manera consistente peores resultados comparados con los de un clasificador logístico."

Este tipo de aportación puede resultar muy valioso para la réplica, la identificación de las características que pueden combinar bien con ciertos algoritmos de clasificación y el fomento en el cambio de estrategias. Lo anterior tiene el potencial de enriquecer los trabajos mediante una experimentación dirigida y, al mismo tiempo, facilitar la comparación entre sistemas inclusive si no comparten el resto de los módulos ni características.

4.5 Diccionarios

Recurso	Artículos									
	1	2	3	4	5	6	7	8	9	10
polarización	✓		✓	✓	✓	✓	✓	✓	✓	
emoticones				✓						
stoplist				✓			✓		✓	
lexicon				✓			✓			
embedding									✓	✓

Tabla 6: Recursos estáticos reportados como herramientas en cada trabajo

Los diccionarios se refieren a aquellas herramientas estáticas de las que hacen uso los participantes para añadir información a sus sistemas. Esto incluye lexicones de polaridad, diccionarios de emoticones, listas de palabras de paro, embeddings pre-entrenados entre otros. Todas aquellas herramientas que están listas para usarse como consulta.

En la Tabla 6 se muestra el uso que tuvieron los participantes del TASS 2017 de distintas herramientas estáticas. Como se puede observar, la mayoría de los sistemas utilizó lexicones de polarización. Como ya se había mencionado en la sección 4.2, Hurtado, Pla, y González (2017) y Rosá et al. (2017) llegan a la conclusión de que esta estrategia en particular no logra un mejor rendimiento en los sistemas. Mientras que Montañés-Salas et al. (2017) llega a la conclusión contraria.

En la sección 4.3 discutimos un cambio en la manera en la que se incorporó la información de polaridad entre los sistemas 2 y 7 como una posible explicación para la diferencia de los resultados. No obstante, no podemos descartar la posibilidad de que la diferencia, o parte de ella, provenga del recurso utilizado. El sistema 2 no utiliza diccionarios afectivos, pero el sistema 9 sí a pesar de que también reporta que no obtuvo mejoras con ello (Rosá et al., 2017).

Podemos notar una interesante diferencia entre los sistemas 7 y 9 en cuanto a los recursos de polaridad que utilizan. Rosá et al. (2017) (sistema 9) utilizan tres léxicos subjetivos disponibles para el español (Cruz et al., 2014; Saralegi y San Vicente, 2013; Brooke, Tofiloski, y Taboada, 2009). Mientras que Montañés-Salas et al. (2017) (sistema 7) utilizan el *Dictionary of Affect in Language* (Cynthia M. Whissell et al., 1986) y el *Affective Norms for English Words* (ANEW; Bradley et al., 1999). A pesar de que Montañés-Salas et al. (2017) utilizan diccionarios en inglés, obtienen mejoras en sus resultados mientras que Rosá et al. (2017) no las obtienen con diccionarios en español.

Como mencionamos en la sección 4.3, la diferencia en el rendimiento puede venir de la representación. Esto supone un espacio para experimentar, la representación por sinónimos y *embeddings* con diccionarios en español, o bien, la anexión de información de polaridad mediante el uso diccionarios en inglés. Esto es tan solo un ejemplo de los nichos que se pueden localizar mediante la estructuración de las características de los sistemas con ayuda de modelos como los que estamos utilizando. Finalmente, el último tipo de componentes por mencionar son los tipos de herramientas, y herramientas específicas que son utilizadas en cada uno de los trabajos.

4.6 Herramientas

Las herramientas se refieren a la paquetería computacional de la que se apoyan los participantes para llevar a cabo cada uno de los módulos antes mencionados, si es el caso.

Herramienta	Artículos									
	1	2	3	4	5	6	7	8	9	10
POS tagger		✓		✓	✓	✓	✓			
clasificador	✓			✓	✓		✓		✓	
ortografía						✓				
tokenizador		✓		✓						
embeddings		✓							✓	
ponderación	✓									
sentimientos			✓							
framework		✓	✓	✓						

Tabla 7: Herramientas computacionales reportadas en cada trabajo

En la Tabla 7 se muestran las diferentes tareas para las que los sistemas que participaron en el TASS 2017 reportaron utilizar paquetería especializada. Este rubro es particularmente útil para la especificación y estandarización de las librerías y sus respectivos parámetros de operación. Al identificar de manera precisa los recursos que se utilizan en cada etapa del proceso, se facilita en gran medida la réplica, la experimentación, el reporte de resultados y la construcción de paquetes de software que coordinen todo de manera automática.

5 Conclusiones

Es claro que la alta heterogeneidad que se presenta en los sistemas hace que no se pueda establecer una relación causa efecto entres

los componentes de cada sistema y su efecto sobre el rendimiento.

Muchos de los procesos que se aplican en los diferentes módulos se hacen debido a resultados exitosos reportados en otros trabajos. No obstante no se tiene una verdadera perspectiva de comparación para identificar en qué medida las técnicas replicadas son las que tuvieron un aporte en los sistemas reportados en el estado del arte.

Son escasos los trabajos que cuentan con la experimentación para determinar la efectividad de las herramientas periféricas que utilizan. De allí la importancia de generar conciencia tendiente a producir una adecuada declaración explícita de cada etapa. Además, resaltar la importancia de generar especificaciones para capturar las pruebas satisfactorias e insatisfactorias que se reporten. Si un sistema similar logra mejoras al variar algún elemento particular esto puede servir de base para replicación y observación del efecto de las variantes.

Un modelo comparativo de esta naturaleza permite agilizar la experimentación tendiente a determinar el efecto de los componentes más polémicos y sus combinaciones. Lo anterior, para encontrar aquellos casos en que funcionen, la manera en la que funcionan, o bien, descartar prácticas que en realidad no están aportando un beneficio. De esta forma, se puede determinar el panorama general asociado al desarrollo los estos sistemas con el fin de lograr mejoras en su desempeño.

Agradecimientos

Se agradece el apoyo de MICITT y CONICIT del Gobierno de Costa Rica y al proyecto IN403016 DGAPA-PAPIIT de la UNAM. Eugenio Martínez Cámara fue financiado por el programa Juan de la Cierva Formación (FJCI-2016-28353) del Gobierno de España.

Bibliografía

Casasola-Murillo, E. 2018. *Desarrollo de un modelo computacional para la especificación de sistemas de análisis de sentimiento con comentarios de redes sociales en español*. Ph.D. tesis, Escuela de Ciencias de la Computación, U. de Costa Rica.

Díaz-Galiano, M. C., E. Martínez-Cámara, M. García-Cumbreras, M. García-Vega, y Villena-Román. 2018. The democratization of deep learning in tass 2017. *Procesamiento del Lenguaje Natural*, 60:37–44.

Hurtado, L.-F., F. Pla, y J.-Á. González. 2017. Elirf-upv en tass 2017: Análisis de sentimientos en twitter basado en aprendizaje profundo. En *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, volumen 1896, páginas 29–34, Murcia, Spain, September. SEPLN, CEUR Workshop Proceedings. ISSN: 1613:0073.

Melero, M., A.-B. Cardús, A. Moreno, G. Rehm, K. de Smedt, y H. Uszkoreit. 2012. *The Spanish language in the digital age*. Springer.

Mintz, M., S. Bills, R. Snow, y D. Jurafsky. 2009. Distant supervision for relation extraction without labeled data. En *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, páginas 1003–1011. Association for Computational Linguistics.

Montañés-Salas, R.-M., d.-H.-M. Alonso-Rafael, J. Vea-Murguía, R. Aznar-Gimeno, y F. J. Lacueva-Pérez. 2017. Fasttext como alternativa a la utilización de deep learning en corpus pequeños. En *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, volumen 1896, páginas 65–69, Murcia, Spain, September. SEPLN, CEUR Workshop Proceedings. ISSN: 1613:0073.

Moreno-Ortiz, A. y C. P. Hernández. 2017. Tecnolengua lingmotif at tass 2017: Spanish twitter dataset classification combining wide-coverage lexical resources and text features. En *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, volumen 1896, páginas 35–42, Murcia, Spain, September. SEPLN, CEUR Workshop Proceedings. ISSN: 1613:0073.

O'Reilly, T. y J. Battelle. 2009. *Web squared: Web 2.0 five years on*. O'Reilly Media.

Rosá, A., L. Chiruzzo, M. Etcheverry, y S. Castro. 2017. Retuyt en tass 2017: Análisis de sentimientos de tweets en español utilizando svm y cnn. En *Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN*, volumen 1896, páginas 77–83, Murcia, Spain, September. SEPLN, CEUR Workshop Proceedings. ISSN: 1613:0073.