

# ELiRF-UPV en TASS 2018: Análisis de Sentimientos en Twitter basado en Aprendizaje Profundo

## *ELiRF-UPV at TASS 2018: Sentiment Analysis in Twitter based on Deep Learning*

José-Ángel González, Lluís-F. Hurtado, Ferran Pla

Universitat Politècnica de València

Camí de Vera s/n

46022 València

{jogonba2, lhurtado, fpla}@dsic.upv.es

**Resumen:** En este trabajo se describe la participación del grupo de investigación ELiRF de la Universitat Politècnica de València en el Taller TASS2018, enmarcado dentro de la XXXIV edición del Congreso Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural. Presentamos las aproximaciones utilizadas para las tareas “Sentiment Analysis at Tweet level” y “Aspect-based Sentiment Analysis” del taller, los resultados obtenidos y una discusión de los mismos. Nuestra participación se ha centrado principalmente en explorar diferentes aproximaciones basadas en *Deep Learning*, consiguiendo resultados competitivos en las tareas abordadas.

**Palabras clave:** Twitter, Análisis de Sentimientos, Deep Learning.

**Abstract:** This paper describes the participation of the ELiRF research group of the Universitat Politècnica de València at TASS2018 Workshop which is a satellite event of the XXXIV edition of the International Conference of the Spanish Society for Natural Language Processing. We describe the approaches used for “Sentiment Analysis at Tweet level” and “Aspect-based Sentiment Analysis” tasks, the results obtained and a discussion of these results. Our participation has focused primarily on exploring different approaches of *Deep Learning* and we have achieved competitive results in the addressed tasks.

**Keywords:** Twitter, Sentiment Analysis, Deep Learning.

### 1 Introducción

El Taller de Análisis de Sentimientos (TASS) ha venido planteando una serie de tareas relacionadas con el análisis de sentimientos en Twitter con el fin de comparar y evaluar las diferentes aproximaciones presentadas por los participantes. Además, desarrolla recursos de libre acceso, básicamente, corpora anotados con polaridad, temática, tendencia política, aspectos, que son de gran utilidad para la comparación de diferentes aproximaciones a las tareas propuestas.

En esta séptima edición del TASS (Martínez-Cámara et al., 2018) se proponen diversas tareas en el contexto del análisis de sentimiento (subtareas 1 y 2), el descubrimiento de conocimiento en documentos médicos (subtarea 3) y la categorización emocional de noticias (subtarea 4). Respecto a las tareas de análisis de sentimiento, los organizadores proponen dos tareas diferentes: 1)

Determinación de la polaridad de los tweets a nivel global y 2) Determinación de la polaridad a nivel de aspecto. Así, para la primera tarea, a diferencia de la edición anterior (Martínez-Cámara et al., 2017), se ha propuesto como nuevo reto la multilingüedad por lo que se proporciona a los participantes tres corpus diferentes de tweets escritos en variedades del español (España, Costa Rica y Perú). Para la segunda tarea se utilizaron los corpus Social.TV, compuesto por tweets publicados durante la final de la Copa del Rey 2014 y STOMPOL, que consta de un conjunto de tweets sobre diferentes aspectos pertenecientes al dominio de la política.

El presente artículo resume la participación del equipo ELiRF-UPV de la Universitat Politècnica de València en ambas tareas, que tratan el análisis de sentimientos tanto a nivel global como de aspectos.

El resto del artículo se estructura como

sigue: para cada tarea, primero se describen las aproximaciones y recursos utilizados en la tarea. A continuación, se presenta la evaluación experimental realizada y los resultados obtenidos. Finalmente se muestran las conclusiones y posibles trabajos futuros.

## 2 *Análisis de sentimientos a nivel de tweet*

Esta tarea consiste en asignar una polaridad global a tweets en una escala de cuatro niveles de intensidad (**N**, **NEU**, **NONE** y **P**).

Así, las principales dificultades radican en aspectos relacionados con el dominio de la tarea. Por un lado, la falta de contexto debido a la limitada longitud de los tweets. Por otro lado, el lenguaje informal es una característica común a muchas redes sociales como Twitter, lo que lleva a errores de ortografía y a la utilización de términos especiales como emoticonos. Además, en esta edición, la organización ha propuesto como nuevo reto la multilingüedad por lo que es necesario conseguir que los sistemas sean capaces de generalizar entre diversas variedades del español.

### 2.1 Corpora

La organización ha definido tres subtarefas considerando tres corpus diferentes con variedades del español. En primer lugar, el corpus InterTASS-ES (España) compuesto por una partición de entrenamiento de 1008 muestras, una de validación de 506 muestras y otra de test formada por 1920 muestras. En segundo lugar, InterTASS-CR (Costa Rica) compuesto por 800 muestras de entrenamiento, 300 para validación y 1233 para test. Por último, InterTASS-PE (Perú), formado por 1000 muestras de entrenamiento, 500 de validación y 1428 de test. Además, es posible emplear la partición de entrenamiento del *General Corpus* (GE), compuesta por 8226 muestras, para entrenar los sistemas.

La distribución de *tweets* según su polaridad en el conjunto de entrenamiento del corpus InterTASS se muestra en la Tabla 1.

La distribución de *tweets* según su polaridad en los conjuntos de entrenamiento de los corpus de InterTASS y *General Corpus* se muestra en la Tabla 1.

Como se puede observar en la Tabla 1, los corpus están desbalanceados predominando generalmente las clases **N** y **P**, excepto en el corpus InterTASS-PE, donde la clase más representada es **NONE**. Así, la clase

	ES	CR	PE	GE
<b>N</b>	418	311	242	2600
<b>NEU</b>	133	94	166	803
<b>NONE</b>	139	165	361	1621
<b>P</b>	318	230	231	3202
$\Sigma$	1008	800	1000	8226

Tabla 1: Distribución de tweets en los conjuntos de entrenamiento de InterTASS y *General Corpus* según su polaridad.

**NEU** siempre es la menos poblada aunque en InterTASS-ES tiene un número de muestras similar a **NONE**.

### 2.2 Descripción de los sistemas

Los sistemas que estudiamos en TASS 2018 continúan con el enfoque utilizado por nuestro equipo en la pasada edición (Hurtado, Pla, y González., 2017). Para la presente edición, hemos definido dos baselines basados en Support Vector Machines (Cortes y Vapnik, 1995) que emplean representaciones *bag-of-words* (BOW) de los tweets, tanto a nivel de palabra como de caracteres (BOC). Tomando como referencia los resultados obtenidos por dichos baselines, se han explorado diversas arquitecturas *Deep Learning*.

La tokenización de los tweets consiste en la adaptación para el castellano del tokenizador de tweets *Tweetmotif* (O’Connor, Krieger, y Ahn, 2010). El preproceso ha consistido en sustituir todos los números, direcciones web, hashtags y menciones de usuario por un token común por clase, por ejemplo #1octL6  $\rightarrow$  #hashtag. Además, se han eliminado los acentos y convertido a minúsculas.

En este trabajo se han explorado diferentes arquitecturas de redes neuronales así como diferentes tipos de representaciones de los tweets. Entre estas arquitecturas, estudiamos el comportamiento de Convolutional Neural Network (CNN) (Kim, 2014), Attention Bidirectional Long Short Term Memory (Att-BLSTM) (Zhou et al., 2016) y Deep Averaging Networks (DAN) (Iyyer et al., 2015). Con respecto a las representaciones, en función de cada modelo, se han empleado diversos tipos como *bag-of-words*, *bag-of-chars*, word embeddings (WE) (Hurtado, Pla, y González., 2017) y lexicones de polaridad y emociones (LE) (Mohammad y Tur-

ney, 2013), (Urizar y Roncal, 2013), (L. Cruz et al., 2014), (Molina-González et al., 2013). Para llevar a cabo la experimentación con diversos sistemas y representaciones, hemos utilizado las librerías *Keras* (Chollet, 2015), *Scikit-Learn* (Buitinck et al., 2013) y *Gensim* (Řehůřek y Sojka, 2010).

Además, para el entrenamiento de las arquitecturas *Deep Learning*, se ha empleado una aproximación continua y derivable a la métrica de evaluación ( $macro-F_1$ ) como función de loss. Empleamos esta estrategia con el objetivo de evitar el impacto del desbalanceo entre clases durante el entrenamiento. Esto nos permite obtener modelos más robustos evaluados con  $macro-F_1$ .

Por último, con respecto al criterio de elección del mejor modelo, escogemos la arquitectura *Deep Learning* y la representación que mejor se comporta en validación sobre el corpus InterTASS-ES. Una vez determinada la representación y la arquitectura junto con sus hiperparámetros, este mismo sistema se emplea en las demás subtareas, i.e. no se ha ajustado ni estudiado modelos sobre InterTASS-CR e InterTASS-PE.

### 2.3 Fase de ajuste

Para estudiar el comportamiento de los diferentes modelos, se realizó un proceso de ajuste. De esta manera, evaluamos diversos modelos *Deep Learning* junto con los baselines sobre InterTASS-ES para escoger el mejor y emplearlo también en InterTASS-CR e InterTASS-PE. En la Tabla 2 se muestran los resultados obtenidos por cada sistema en los conjuntos de validación de cada subtarea. En dicha tabla, S hace referencia al sistema, R a la representación y C al conjunto de entrenamiento utilizado.

Respecto a la subtarea evaluada con el corpus InterTASS-ES, cuando se utiliza como medida de evaluación la Accuracy ( $Acc$ ), la mayoría de sistemas basados en *Deep Learning* mejoran los baselines propuestos. Con respecto a la  $macro-F_1$ , observamos una diferencia de 7 puntos entre los baselines SVM con BOW (1-2gramas) y BOC (1-9gramas), a su vez, todos los sistemas basados en *Deep Learning* obtienen valores más altos que ambos baselines. Entre los modelos de aprendizaje profundo observamos que se comportan mejor, tanto a nivel de  $Acc$  como de  $macro-F_1$ , aquellos con menor complejidad y que capturan características independientes del

orden de las palabras (DAN frente a CNN y Att-BLSTM). También es necesario destacar que la incorporación de información extraída de lexicones no aporta mejoras al sistema DAN.

Así, de entre todos los sistemas estudiados para InterTASS-ES, escogemos aquellos que maximizan  $Acc$  y  $macro-F_1$  de forma independiente, estos son DAN con WE entrenados con InterTASS-ES y con InterTASS-ES+GE respectivamente. Una vez se ha escogido DAN con WE como mejor sistema en InterTASS-ES para  $Acc$  (**es-run1**) y  $macro-F_1$  (**es-run2**), se utiliza en InterTASS-CR e InterTASS-PE entrenando con el conjunto de entrenamiento de cada tarea y la combinación de este con el *General Corpus*.

La Figura 1 muestra el sistema propuesto en este trabajo, donde  $x_i$  representa el embedding de la palabra  $i$ ,  $N$  representa el uso de Batch Normalization (Ioffe y Szegedy, 2015),  $F$  la no linealidad, en este caso ReLU,  $D$  se refiere al uso de Dropout (Srivastava et al., 2014) con  $p = 0,3$  y  $W_1 \in \mathbb{R}^{512*d_e}$  son los pesos de la única capa oculta y  $d_e$  la dimensionalidad de los embeddings. Como algoritmo de optimización se ha empleado Adagrad (Duchi, Hazan, y Singer, 2011).

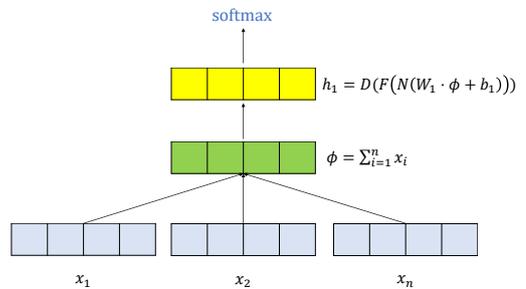


Figura 1: Deep Averaging Network que mejor se comporta en la partición de validación de InterTASS-ES.

### 2.4 Resultados experimentales

En la Tabla 3 se muestran los resultados obtenidos por nuestros sistemas en cada una de las subtareas utilizando los conjuntos de test. Se han elegido los dos sistemas que han obtenido mejores resultados en la fase de ajuste.

En el test de InterTASS-ES nuestro sistema **es-run1** obtiene los mejores resultados de la competición tanto a nivel de  $macro-F_1$  como de  $Acc$  (incluyendo la edición pasada sobre el mismo corpus). Sin embargo,

	S	R	C	Macro- $P$	Macro- $R$	Macro- $F_1$	$Acc$
ES	SVM	BOW	ES	42.69	40.20	40.34	53.16
	SVM	BOC	ES	51.09	46.89	47.48	59.88
	Att-BLSTM	WE	ES	51.96	49.67	49.81	58.10
	CNN	WE	ES	52.97	49.09	49.12	60.16
	DAN (es-run1)	WE	ES	51.54	49.49	49.99	<b>61.07</b>
	DAN	WE+LE	ES	48.20	47.02	47.52	58.30
	DAN (es-run2)	WE	ES+GE	52.97	51.46	<b>52.01</b>	60.67
CR	DAN (cr-run1)	WE	CR	54.98	52.28	<b>52.71</b>	<b>58.67</b>
	DAN (cr-run2)	WE	CR+GE	49.10	48.67	48.73	56.00
PE	DAN (pe-run1)	WE	PE	45.62	47.91	<b>45.79</b>	<b>49.60</b>
	DAN (pe-run2)	WE	PE+GE	46.49	50.77	44.53	44.80

Tabla 2: Resultados de los diversos sistemas sobre los conjuntos de validación.

		Macro- $P$	Macro- $R$	Macro- $F_1$	$Acc$
ES	run1	51.50	49.10	50.30	61.20
	run2	49.20	48.60	48.90	59.30
CR	run1	47.60	45.10	46.30	54.40
	run2	47.80	48.60	48.20	56.10
PE	run1	43.40	43.50	43.50	44.40
	run2	43.80	43.80	43.80	46.10

 Tabla 3: Resultados oficiales del equipo *ELiRF-UPV* en las tres subtareas de la Tarea 1 (los resultados de los demás participantes se muestran en (Martínez-Cámara et al., 2018)).

los mismos sistemas para InterTASS-CR e InterTASS-PE no se comportan igual de bien en comparación a los sistemas de otros participantes.

En el caso de InterTASS-CR, nuestro mejor sistema **cr-run2** se encuentra 2.2 puntos de  $macro-F_1$  por debajo del sistema LSTM de RETUYT y 2.1 puntos de  $Acc$  por debajo del sistema CR-LR de Atalaya (la diferencia de  $Acc$  no es significativa). Por otro lado, en InterTASS-PE, nuestro mejor sistema **pe-run2** se encuentra 3.4 puntos de  $macro-F_1$  por debajo del sistema CNN de RETUYT y 5.9 puntos de  $Acc$  por debajo del sistema MLP de Atalaya. En este caso, la diferencia a nivel de  $Acc$  sí es significativa.

Cabe destacar que el sistema con mejor comportamiento en el test de InterTASS-CR e InterTASS-PE, a diferencia de lo observado en la fase de ajuste, ha sido el basado en Deep Averaging Networks (DAN) con Embeddings (WE) entrenado utilizando el corpus de cada subtarea en combinación con el *General Corpus*. Por tanto, sobre el conjunto de test, la inclusión de datos de ediciones pasadas nos ha permitido mejorar al sistema entrenado

únicamente con el corpus de entrenamiento de esta edición (**cr-run1** y **pe-run1**).

En las Tablas 4, 5 y 6 se muestra la evaluación por clase de los mejores sistemas para las tres subtareas. En ellas se puede observar como, en todos los casos, las clases que mejor clasifican nuestros sistemas son **P** y **N**. Además, los resultados obtenidos con las clases **NEU** y **NONE** son siempre inferiores por un amplio margen en comparación a **P** y **N**, lo que muestra la dificultad de clasificar correctamente dichas clases. También, es necesario destacar que, mientras en InterTASS-ES e InterTASS-CR nuestros sistemas se comportan mejor con la clase **NO-NE** que con la clase **NEU**, en InterTASS-PE ocurre lo contrario y se equilibran los resultados de ambas clases.

Por otro lado, para estudiar el grado de confusión entre clases de nuestros mejores sistemas para las tres subtareas, en las Tablas 7, 8 y 9 se muestra la matriz de confusión para cada uno de ellos.

En InterTASS-ES, las mayores confusiones se dan entre las clases **N-P**, **P-N**, **NONE-N** y **NONE-P**. En InterTASS-CR,

	$P$	$R$	$F_1$
<b>N</b>	63.30	76.50	71.00
<b>NEU</b>	21.30	15.70	18.10
<b>NONE</b>	51.60	34.30	41.20
<b>P</b>	66.80	69.80	68.20

Tabla 4: Resultados de Precisión, Recall y  $F_1$  por clase para el sistema **es-run1** en InterTASS-ES

	$P$	$R$	$F_1$
<b>N</b>	65.10	67.20	66.10
<b>NEU</b>	22.80	11.00	14.80
<b>NONE</b>	42.30	51.80	46.60
<b>P</b>	61.10	64.40	62.70

Tabla 5: Resultados de Precisión, Recall y  $F_1$  por clase para el sistema **cr-run2** en InterTASS-CR

el sistema tiende a confundirse con mayor frecuencia en los casos **NEU-N**, **NEU-P**, **NONE-N** y **NONE-P**. Por último, en InterTASS-PE, el mejor sistema predice con mucha frecuencia la clase **NEU** lo que lleva a una gran confusión en los casos **N-NEU** y **P-NEU**, además, pasa algo similar con las predicciones **NONE**, destacando la confusión en los casos **N-NONE** y **NEU-NONE**.

### 3 Análisis de sentimientos a nivel de aspecto

Esta tarea consiste en asignar la polaridad a los aspectos que aparecen marcados en el corpus. Una de las dificultades de la tarea consiste en definir qué contexto se le asigna a cada aspecto para poder establecer su polaridad. Para un problema similar, detección de la polaridad a nivel de entidad, en la edición del TASS 2013, propusimos una segmentación de los tweets basada en un conjunto de heurísticas (Pla y Hurtado, 2013). Esta aproximación también se utilizó para la tarea de detección de la tendencia política de los usuarios de Twitter (Pla y Hurtado, 2014) y para este caso proporcionó buenos resultados. En este trabajo se emplea la aproximación utilizada en la edición del TASS 2017, que consiste en determinar el contexto de cada aspecto a través de una ventana fija definida a la izquierda y derecha de la instancia del aspecto. La longitud de la ventana óptima se ha determinado experimentalmente sobre el conjunto de entrenamiento mediante *holdout*.

La organización del TASS ha planteado dos subtarefas. La primera utiliza el corpus

	$P$	$R$	$F_1$
<b>N</b>	62.60	47.20	53.90
<b>NEU</b>	31.00	28.60	29.70
<b>NONE</b>	19.20	35.20	24.80
<b>P</b>	62.40	64.00	63.10

Tabla 6: Resultados de Precisión, Recall y  $F_1$  por clase para el sistema **pe-run2** en InterTASS-PE

	<b>N</b>	<b>NEU</b>	<b>NONE</b>	<b>P</b>
<b>N</b>	587	55	34	91
<b>NEU</b>	107	34	18	57
<b>NONE</b>	75	30	94	75
<b>P</b>	117	41	36	448

Tabla 7: Matriz de confusión del sistema **es-run1** en InterTASS-ES

Social\_TV y la segunda el corpus STOMPOL.

#### 3.1 Corpus Social\_TV

El corpus Social\_TV fue proporcionado por la organización y se compone de un conjunto de tweets recolectados durante la final de la Copa del Rey de fútbol de 2014. Está dividido en 1773 tweets de entrenamiento y 1000 tweets de test. El conjunto de entrenamiento está anotado con los aspectos y su correspondiente polaridad, utilizando en este caso sólo tres valores: P, N y NEU. El conjunto de test está anotado con los aspectos y se debe determinar la polaridad de éstos.

#### 3.2 Corpus STOMPOL

El corpus STOMPOL se compone de un conjunto de tweets relacionados con una serie de aspectos políticos, como economía, sanidad, etc. que están enmarcados en la campaña política de las elecciones andaluzas de 2015. Cada aspecto se relaciona con una o varias entidades que se corresponden con uno de los principales partidos políticos en España (PP, PSOE, IU, UPyD, Cs y Podemos). El corpus consta de 1.284 tweets, y ha sido dividido en un conjunto de entrenamiento (784 tweets) y un conjunto de evaluación (500 tweets).

#### 3.3 Aproximación y resultados

Los sistemas utilizados son iguales a los empleados con los corpora InterTASS de la primera tarea. Se han utilizado modelos basados en Deep Averaging Networks y Embeddings, tanto para Social\_TV como para STOMPOL. Además, no hemos realizado ningún ajuste en los hiperparámetros ni en la arquitectura,

	N	NEU	NONE	P
N	330	38	69	54
NEU	65	18	36	45
NONE	56	6	116	46
P	56	17	53	228

Tabla 8: Matriz de confusión del sistema **cr-run2** en InterTASS-CR

	N	NEU	NONE	P
N	223	126	82	41
NEU	64	105	117	81
NONE	32	27	56	44
P	37	81	37	275

Tabla 9: Matriz de confusión del sistema **pe-run2** en InterTASS-PE

únicamente se ha realizado una fase de ajuste para determinar la longitud óptima de la ventana para cada aspecto.

Con respecto a dicha longitud, en el corpus Social\_TV, observamos que cuanto menor es el tamaño de contexto, mejor se comportan los sistemas. Así, la mejor longitud para este caso consiste en un único token a izquierda y derecha del aspecto. En STOMPOL, por el contrario, cuanto mayor es el tamaño de contexto, mejores resultados obtienen los sistemas. En este caso, el mejor contexto está formado por 10 tokens a izquierda y derecha del aspecto, lo que equivale a considerar todo el tweet para cada aspecto.

En lo referente a los sistemas, los enviados para la tarea Social\_TV son los siguientes:

- **run1**: La primera alternativa es el **run1** de la Tarea 1, DAN con WE entrenado con el conjunto de entrenamiento de Social\_TV.
- **run2**: La segunda alternativa es el **run1** que el grupo envió a la edición pasada de Social\_TV. También se trata de DAN con WE pero con hiperparámetros y tamaño de contexto diferentes.
- **run3**: Para el último sistema utilizamos *transfer learning* desde el **es-run1** de la Tarea 1, cambiando la última capa del modelo y reentrenándolo con el corpus Social\_TV.

Por otro lado, para la tarea con STOMPOL, los sistemas considerados son idénticos a la anterior tarea, pero con un tamaño de contexto diferente y entrenados con el corpus STOMPOL.

		Acc	Macro- $F_1$
Social_TV	run-1	62.70	48.50
	run-2	62.50	47.60
	run-3	62.80	48.30
STOMPOL	run-1	63.30	52.60
	run-2	61.30	49.00
	run-3	57.60	44.70

Tabla 10: Resultados oficiales del equipo *ELiRF-UPV* en las dos subtareas de la tarea 2.

Por último, los resultados obtenidos de *Accuracy* y *macro- $F_1$*  con los distintos sistemas para cada subtarea se muestran en la Tabla 10.

Como se puede observar en la Tabla 10, en ambos casos mejoramos nuestros resultados de la edición anterior, haciendo uso del sistema DAN con WE. Es interesante ver cómo *transfer learning* nos permite mejorar el sistema en Social\_TV a nivel de *Acc*, sin embargo, en el caso de STOMPOL, el sistema inicializado con *transfer learning* se comporta peor que el resto tanto a nivel de *Acc* como de *macro- $F_1$* .

#### 4 Conclusiones y trabajos futuros

En este trabajo se ha presentado la participación del equipo ELiRF-UPV en las tareas 1 y 2 planteadas en TASS2018. Nuestro equipo ha utilizado modelos *Deep Learning*, consiguiendo los mejores resultados en InterTASS-ES. Sin embargo, la variabilidad lingüística afecta a la elección de la arquitectura y sus hiperparámetros, por lo que la aplicación del mismo sistema a las tareas InterTASS-CR e InterTASS-PE sin realizar ningún ajuste no nos ha permitido obtener resultados igual de competitivos que en InterTASS-ES.

Como trabajo futuro, nuestro grupo está interesado en seguir trabajando en la minería de textos en redes sociales y especialmente en la detección de *stance*, afecto y emociones, así como en la incorporación de recursos lingüísticos a sistemas de *Deep Learning*.

#### Agradecimientos

Este trabajo ha sido parcialmente subvencionado por MINECO y fondos FEDER bajo los proyectos ASLP-MULAN (TIN2014-54288-C4-3-R) y AMIC (TIN2017-85854-C4-2-R). El trabajo de José-Ángel González es también financiado por la Universidad Po-

litécnica de Valencia bajo la beca PAID-01-17.

### Bibliografía

- Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, y G. Varoquaux. 2013. API design for machine learning software: experiences from the scikit-learn project. En *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, páginas 108–122.
- Chollet, F. 2015. Keras. <https://github.com/fchollet/keras>.
- Cortes, C. y V. Vapnik. 1995. Support-vector networks. *Mach. Learn.*, 20(3):273–297, Septiembre.
- Duchi, J., E. Hazan, y Y. Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.*, 12:2121–2159, Julio.
- Hurtado, L.-F., F. Pla, y J.-A. González. 2017. Elirf-upv en tass 2017: Análisis de sentimientos en twitter basado en aprendizaje profundo. En J. Villena Román M. A. García Cumberras E. Martínez-Cámara M. C. Díaz Galiano, y M. García Vega, editores, *In Proceedings of TASS 2017: Workshop on Sentiment Analysis at SEPLN co-located with 33rd SEPLN Conference (SEPLN 2017)*, volumen 1896 de *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.
- Ioffe, S. y C. Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. En *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, páginas 448–456. JMLR.org.
- Iyyer, M., V. Manjunatha, J. Boyd-Graber, y H. Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. En *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, páginas 1681–1691. Association for Computational Linguistics.
- Kim, Y. 2014. Convolutional neural networks for sentence classification. En *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, páginas 1746–1751. Association for Computational Linguistics.
- L. Cruz, F., J. A. Troyano, B. Pontes, y F. J. Ortega. 2014. Building layered, multilingual sentiment lexicons at synset and lemma levels. 41:5984–5994, 10.
- Martínez-Cámara, E., Y. Almeida Cruz, M. C. Díaz-Galiano, S. Estévez Velarde, M. A. García-Cumberras, M. García-Vega, Y. Gutiérrez Vázquez, A. Montejo Ráez, A. Montoyo Guijarro, R. Muñoz Guillena, A. Piad Morffis, y J. Villena-Román. 2018. Overview of TASS 2018: Opinions, health and emotions. En E. Martínez-Cámara Y. Almeida Cruz M. C. Díaz-Galiano S. Estévez Velarde M. A. García-Cumberras M. García-Vega Y. Gutiérrez Vázquez A. Montejo Ráez A. Montoyo Guijarro R. Muñoz Guillena A. Piad Morffis, y J. Villena-Román, editores, *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018)*, volumen 2172 de *CEUR Workshop Proceedings*, Sevilla, Spain, September. CEUR-WS.
- Martínez-Cámara, E., M. C. Díaz-Galiano, M. A. García-Cumberras, M. García-Vega, y J. Villena-Román. 2017. Overview of TASS 2017. En J. Villena Román M. A. García Cumberras E. Martínez-Cámara M. C. Díaz Galiano, y M. García Vega, editores, *Proceedings of TASS 2017: Workshop on Semantic Analysis at SEPLN (TASS 2017)*, volumen 1896 de *CEUR Workshop Proceedings*, Murcia, Spain, September. CEUR-WS.
- Mohammad, S. M. y P. D. Turney. 2013. Crowdsourcing a Word-Emotion Association Lexicon. *Computational Intelligence*, 29(3):436–465.
- Molina-González, M. D., E. Martínez-Cámara, M.-T. Martín-Valdivia, y J. M. Perea-Ortega. 2013. Semantic orientation for polarity classification in spanish reviews. *Expert Systems with Applications*, 40(18):7250 – 7257.
- O'Connor, B., M. Krieger, y D. Ahn. 2010.

Tweetmotif: Exploratory search and topic summarization for twitter.

- Pla, F. y L.-F. Hurtado. 2013. Elirf-upv en tass-2013: Análisis de sentimientos en twitter. En *XXIX Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN 2013)*. TASS, páginas 220–227.
- Pla, F. y L.-F. Hurtado. 2014. Political tendency identification in twitter using sentiment analysis techniques. En *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, páginas 183–192, Dublin, Ireland, August. Dublin City University and Association for Computational Linguistics.
- Řehůřek, R. y P. Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. En *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, páginas 45–50, Valletta, Malta, Mayo. ELRA. <http://is.muni.cz/publication/884893/en>.
- Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, y R. Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, Enero.
- Urizar, X. S. y I. S. V. Roncal. 2013. Elhuyar at tass 2013. TASS.
- Zhou, P., W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, y B. Xu. 2016. Attention-based bidirectional long short-term memory networks for relation classification. En *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, páginas 207–212. Association for Computational Linguistics.