

ELiRF-UPV en TASS 2015: Análisis de Sentimientos en Twitter

ELiRF-UPV at TASS 2015: Sentiment Analysis in Twitter

Lluís-F. Hurtado, Ferran Pla
Universitat Politècnica de València
Camí de Vera s/n
46022 València
{lhurtado, fpla}@dsic.upv.es

Davide Buscaldi
Université Paris 13
Sorbonne Paris Cité, LIPN
F-93430 Villetaneuse, France
davide.buscaldi@lipn.univ-paris13.fr

Resumen: En este trabajo se describe la participación del equipo del grupo de investigación ELiRF de la Universitat Politècnica de València en el Taller TASS2015. Este taller es un evento enmarcado dentro de la XXXI edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural. Este trabajo presenta las aproximaciones utilizadas para las todas las tareas del taller, los resultados obtenidos y una discusión de los mismos. Nuestra participación se ha centrado principalmente en explorar diferentes aproximaciones para combinar un conjunto de sistemas. Mediante esta técnica hemos conseguido mejorar las prestaciones de ediciones anteriores.

Palabras clave: Twitter, Análisis de Sentimientos.

Abstract: This paper describes the participation of the ELiRF research group of the Universitat Politècnica de València at TASS2015 Workshop. This workshop is a satellite event of the XXXI edition of the Annual Conference of the Spanish Society for Natural Language Processing. This work describes the approaches used for all the tasks of the workshop, the results obtained and a discussion of these results. Our participation has focused primarily on exploring different approaches for combining a set of systems. Using this technique we have improved the performance of previous editions.

Keywords: Twitter, Sentiment Analysis.

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 cuarta edición del TASS se proponen dos tareas de ediciones anteriores (Villena-Román y García-Morera, 2013): 1) Determinación de la polaridad en tweets, con diferentes grados de intensidad en la polaridad: 6 etiquetas y 4 etiquetas y 2) Determinación de la polaridad de los aspectos en el corpus Social.TV, compuesto por tweets pu-

blicados durante la final de la Copa del Rey 2014. En esta edición del TASS 2015 (Villena-Román et al., 2015), se propone una tarea similar a la 2) pero utilizando un nuevo corpus llamado STOMPOL. Este corpus 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 todas las tareas planteadas en este taller. Primero se describen las aproximaciones y recursos utilizados en cada tarea. A continuación se presenta la evaluación experimental realizada y los resultados obtenidos. Finalmente se muestran las conclusiones y posibles trabajos futuros.

2. *Descripción de los sistemas*

Los sistemas presentados en el TASS 2015, utilizan muchas de las características, desarrollos y recursos utilizados en las ediciones

en las que nuestro equipo ha participado (Pla y Hurtado, 2013) (Hurtado y Pla, 2014). El preproceso de los tweets utiliza la estrategia descrita en el trabajo del TASS 2013 (Pla y Hurtado, 2013). Esta consiste básicamente en la adaptación para el castellano del tokenizador de tweets *Tweetmotif* (Connor, Krieger, y Ahn, 2010)¹. También se ha usado *Freeling* (Padró y Stanilovsky, 2012)² como lematizador, detector de entidades nombradas y etiquetador morfosintáctico, con las correspondientes modificaciones para el dominio de Twitter. Usando esta aproximación, la tokenización ha consistido en agrupar todas las fechas, los signos de puntuación, los números y las direcciones web. Se han conservado los hashtags y las menciones de usuario. Se ha considerado y evaluado el uso de palabras y lemas como tokens así como la detección de entidades nombradas.

Todas las tareas se han abordado como un problema de clasificación. Se han utilizado Máquinas de Soporte Vectorial (SVM) por su capacidad para manejar con éxito grandes cantidades de características. En concreto usamos dos librerías (*LibSVM*³ y *LibLinear*⁴) que han demostrado ser eficientes implementaciones de SVM que igualan el estado del arte. El software se ha desarrollado en *Python* y para acceder a las librerías de SVM se ha utilizado el toolkit *scikit-learn*⁵. (Pedregosa et al., 2011).

En este trabajo se ha explotado la técnica de combinación de diferentes configuraciones de clasificadores para aprovechar su complementariedad. Se ha utilizado la técnica de votación simple utilizada en trabajos anteriores (Pla y Hurtado, 2013) (Pla y Hurtado, 2014b) pero en este caso extendiéndola a un número mayor de clasificadores, con diferentes parámetros y características (palabras, lemas, n-gramas de palabras y lemas) así como estrategias de combinación alternativas. Además, se ha incluido un nuevo recurso léxico, el diccionario *Afinn* (Hansen et al., 2011), que se ha traducido automáticamente del inglés al castellano y se ha adaptado para las tareas consideradas.

Cada tweet se ha representado como un vector que contiene los coeficientes tf-idf de

las características consideradas. En toda la experimentación realizada, las características y los parámetros de los clasificadores se han elegido mediante una validación cruzada de 10 iteraciones (10-fold cross-validation) sobre el conjunto de entrenamiento.

3. Tarea 1: Análisis de sentimientos en tweets

Esta tarea consiste en determinar la polaridad de los tweets y la organización ha definido dos subtareas. La primera distingue seis etiquetas de polaridad: N y N+ que expresan polaridad negativa con diferente intensidad, P y P+ para la polaridad positiva con diferente intensidad, NEU para la polaridad neutra y NONE para expresar ausencia de polaridad. La segunda sólo distinguen 4 etiquetas de polaridad: N, P, NEU y NONE.

El corpus proporcionado por la organización del TASS consta de un conjunto de entrenamiento, compuesto por 7219 tweets etiquetados con la polaridad usando seis etiquetas, y un conjunto de test, de 60798 tweets, al cual se le debe asignar la polaridad. La distribución de tweets según su polaridad en el conjunto de entrenamiento se muestra en la Tabla 1.

Polaridad	# tweets	%
N	1335	18.49
N+	847	11.73
NEU	670	9.28
NONE	1483	20.54
P	1232	17.07
P+	1652	22.88
TOTAL	7219	100

Tabla 1: Distribución de tweets en el conjunto de entrenamiento según su polaridad.

A partir de la tokenización propuesta se realizó un proceso de validación cruzada (10-fold cross validation) para determinar el mejor conjunto de características y los parámetros del modelo. Como características se probaron diferentes tamaños de n-gramas de palabras y de lemas. También se exploró la combinación de los modelos mediante diferentes técnicas de votación para aprovechar su complementariedad y mejorar las prestaciones finales. Algunas de éstas técnicas proporcionaron mejoras significativas sobre el mismo conjunto de datos, como se muestra en (Pla

¹<https://github.com/brendano/tweetmotif>.

²<http://nlp.lsi.upc.edu/freeling/>

³<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

⁴<http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

⁵<http://scikit-learn.org/stable/>

y Hurtado, 2014b). En todos los casos se han utilizado diccionarios de polaridad, tanto de lemas (Saralegi y San Vicente, 2013), como de palabras (Martínez-cámara et al., 2013). Además se ha incorporado el diccionario *Afinn* traducido automáticamente del inglés al castellano y adaptado a la tarea.

Se han considerado tres alternativas para abordar la tarea:

- **run1** La primera alternativa combina mediante votación simple los 3 sistemas presentados en la edición del TASS de 2014.
- **run2-run4** La segunda alternativa explora diferentes combinaciones de parámetros y características de un modelo SVM. Para ello se han tenido en cuenta 192 configuraciones. A partir de éstas, se ha aprendido un segundo modelo SVM que sirve para proporcionar la nueva salida combinada. La diferencia entre el run2 y el run4 es que en el primero no se ajustan los parámetros del modelo SVM mientras que en el segundo, se elige una parte del entrenamiento para ajustar los parámetros. El sistema que se considera para la competición bajo esta aproximación es el run4.
- **run3** La tercera alternativa combina mediante un sistema de votación de mayoría simple las 192 configuraciones contempladas.

Para la subtarea de 4 etiquetas no se ha construido ningún sistema específico. Los tres sistemas enviados utilizan las salidas de la subtarea de 6 etiquetas uniendo P y P+ como P y N y N+ como N.

En la Tabla 2 se muestran los valores de Accuracy obtenidos para las dos subtareas. Con los sistemas presentados se obtienen mejoras respecto a los resultados presentados en la edición anterior.

4. Tarea 2: Análisis de Polaridad de Aspectos en Twitter

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

	Run	Accuracy
6-ETIQUETAS	run1	0.648
	run3	0.658
	run4	0.673
4-ETIQUETAS	run1	0.712
	run3	0.721
	run4	0.725

Tabla 2: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea 1 de la competición TASS-2015 sobre el conjunto de test para 6 y 4 etiquetas.

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, 2014a) y para este caso proporcionó buenos resultados. En este trabajo se propone una aproximación más simple 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. Esta aproximación es similar a la que se utilizó en nuestro sistema del TASS 2014, pero para esta edición hemos considerado ventanas de diferente longitud. La longitud de la ventana óptima se ha determinado experimentalmente sobre el conjunto de entrenamiento mediante una validación cruzada. Para entrenar nuestro sistema, se ha considerado el conjunto de entrenamiento únicamente, se han determinado los segmentos para cada aspecto y se ha seguido una aproximación similar a la Tarea 1.

La organización del TASS ha planteado dos subtareas. La primera utiliza el corpus *Social TV* y la segunda el corpus *STOMPOL*.

4.1. 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.

4.2. 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 enmarcado 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).

4.3. Aproximación y resultados

A continuación presentamos una pequeña descripción de las características de nuestro sistema así como el proceso seguido en la fase de entrenamiento. El sistema utiliza un clasificador basado en SVM. Para aprender los modelos sólo se utiliza el conjunto de entrenamiento proporcionado para la tarea y los diccionarios de polaridad previamente descritos. Antes de abordar el entrenamiento se determinan los segmentos de tweet que constituyen el contexto de cada una de los aspectos presentes. Se ha tenido en cuenta tres tamaños de ventana de longitudes 5, 7 y 10 palabras a la izquierda y derecha del aspecto. Cada uno de los segmentos se tokeniza y se utiliza Freeling para determinar sus lemas y ciertas entidades. A continuación se aprenden diferentes modelos combinando tamaños de ventana, parámetros del modelo y diferentes características (palabras, lemas, NE, etc). Mediante validación cruzada se elige el mejor modelo. Para esta tarea sólo hemos presentado un modelo.

	Run	Accuracy
SocialTV	run1	0.655
STOMPOL	run1	0.633

Tabla 3: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea2 de la competición TASS-2015 para los corpus SocialTV y STOMPOL respectivamente.

En la tabla 3 se presentan los resultados obtenidos para la Tarea 2 sobre los dos corpus propuestos. Nuestra aproximación ha obtenido la primera posición en ambos corpus.

5. Conclusiones y trabajos futuros

En este trabajo se ha presentado la participación del equipo *ELiRF-UPV* en las 2 tareas planteadas en TASS 2015. Nuestro equipo ha utilizado técnicas de aprendizaje automático, en concreto, aproximaciones basadas en máquinas de soporte vectorial. Para ello hemos utilizado la librería para Python *scikit-learn* y las librerías externas *LibSVM* y *LibLinear*. Nuestra participación se ha centrado principalmente en explorar diferentes aproximaciones para combinar un conjunto de sistemas. Mediante esta técnica hemos conseguido mejorar las prestaciones de ediciones anteriores.

Nuestro grupo está interesado en seguir trabajando en la minería de textos en redes sociales y especialmente en incorporar nuevos recursos a los sistemas desarrollados y estudiar nuevas estrategias y métodos de aprendizaje automático.

Como trabajo futuro nos planteamos desarrollar nuevos métodos de combinación de sistemas. También estamos interesados en considerar diferentes paradigmas de clasificación más heterogéneos (distintos de los SVM) para aumentar la complementariedad de los sistemas combinados.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por los proyectos DIANA: DIScourse ANALysis for knowledge understanding (MEC TIN2012-38603-C02-01) y ASLP-MULAN: Audio, Speech and Language Processing for Multimedia Analytics (MEC TIN2014-54288-C4-3-R).

Bibliografía

- Connor, Brendan O, Michel Krieger, y David Ahn. 2010. Tweetmotif: Exploratory search and topic summarization for twitter. En William W. Cohen y Samuel Gosling, editores, *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*. The AAAI Press.
- Hansen, Lars Kai, Adam Arvidsson, Finn Årup Nielsen, Elanor Colleoni, y Michael Etter. 2011. Good friends, bad news-affect and virality in twitter. En

- Future information technology*. Springer, páginas 34–43.
- Hurtado, Lluís-F y Ferran Pla. 2014. Elirf-upv en tass 2014: Análisis de sentimientos, detección de tópicos y análisis de sentimientos de aspectos en twitter. En *TASS2014*.
- Martínez-cámara, E., M. T. Martín-valdivia, M. D. Molina-gonzález, y L. A. Ureña-lópez. 2013. Bilingual Experiments on an Opinion Comparable Corpus. En *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, página 87–93.
- Padró, Lluís y Evgeny Stanilovsky. 2012. Freeling 3.0: Towards wider multilinguality. En *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*, Istanbul, Turkey, May. ELRA.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, y E. Duchesnay. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pla, Ferran y Lluís-F Hurtado. 2013. Tass-2013: Análisis de sentimientos en twitter. En *Proceedings of the TASS workshop at SEPLN 2013*. IV Congreso Español de Informática.
- Pla, Ferran y Lluís-F. Hurtado. 2014a. 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.
- Pla, Ferran y Lluís-F. Hurtado. 2014b. Sentiment analysis in twitter for spanish. En Elisabeth Métais Mathieu Roche, y Maelonne Teisseire, editores, *Natural Language Processing and Information Systems*, volumen 8455 de *Lecture Notes in Computer Science*. Springer International Publishing, páginas 208–213.
- Saralegi, Xabier y Iñaki San Vicente. 2013. Elhuyar at tass 2013. En *Proceedings of the TASS workshop at SEPLN 2013*. IV Congreso Español de Informática.
- Villena-Román, Julio y Janine García-Morera. 2013. Workshop on sentiment analysis at sepln 2013: An over view. En *Proceedings of the TASS workshop at SEPLN 2013*. IV Congreso Español de Informática.
- Villena-Román, Julio, Janine García-Morera, Miguel A. García-Cumbreras, Eugenio Martínez-Cámara, M. Teresa Martín-Valdivia, y L. Alfonso Ureña-López. 2015. Overview of tass 2015.