

ELiRF-UPV en TASS 2016: Análisis de Sentimientos en Twitter

ELiRF-UPV at TASS 2016: Sentiment Analysis in Twitter

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

Resumen: En este trabajo se describe la participación del equipo del grupo de investigación ELiRF de la Universitat Politècnica de València en el Taller TASS2016. Este taller es un evento enmarcado dentro de la XXXII edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural. Este trabajo presenta las aproximaciones utilizadas para las dos tareas planteadas en el 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 con lo que se ha obtenido los mejores resultados en ambas tareas.

Palabras clave: Twitter, Análisis de Sentimientos.

Abstract: This paper describes the participation of the ELiRF research group of the Universitat Politècnica de València at TASS2016 Workshop. This workshop is a satellite event of the XXXII edition of the Annual Conference of the Spanish Society for Natural Language Processing. This work describes the approaches used for the two 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 these approaches we have achieved the best results in both tasks.

Keywords: Twitter, Sentiment Analysis.

1. Introducción

El Taller de Análisis de Sentimientos (TASS) en sus cinco ediciones ha venido planteando tareas relacionadas con el análisis de sentimientos en Twitter. El objetivo principal es el de comparar y evaluar diferentes aproximaciones a estas tareas. 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 quinta edición del TASS se proponen dos tareas de ediciones anteriores (García-Cumbreras et al., 2016): 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 STOMPOL. Este corpus consta de un con-

junto 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 2016 se basan en el sistema desarrollado en la edición anterior del TASS 2015 (Hurtado, Pla, y Buscaldi, 2015). Muchas de las características y recursos de este sistema fueron 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). También se ha usado *Freeling* (Padró y Stanilovsky, 2012)¹ 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*² y *LibLinear*³) que han demostrado ser eficientes implementaciones de SVM que igualan el estado del arte. El software está desarrollado en *Python* y para acceder a las librerías de SVM se ha utilizado el toolkit *scikit-learn*⁴. (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.

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 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) y el diccionario *Afinn* (Hansen et al., 2011) traducido automáticamente del inglés al castellano.

¹<http://nlp.lsi.upc.edu/freeling/>

²<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

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

⁴<http://scikit-learn.org/stable/>

Se han considerado dos alternativas para abordar la tarea:

- **run1** La primera alternativa combina mediante un sistema de votación ponderada la salida de 192 clasificadores basados en el uso de SVM. La diferencia entre los clasificadores radica en el pre-procesado y la tokenización utilizada, las características seleccionadas y los valores de los parámetros del propio modelo SVM.

En concreto se realizaron todas las combinaciones posibles entre 8 tokenizaciones (lemas o palabras, detectar NE o no, detectar menciones a usuarios y hashtags, ...); 4 conjuntos distintos de características (palabras o bigramas con y sin diccionarios de polaridad) y 6 valores distintos del parámetro c del modelo SVM con kernel lineal.

La clase asignada a cada tweet t viene determinada por la siguiente fórmula.

$$\hat{c} = \underset{c \in \mathcal{C}}{\operatorname{argmax}}(N_t(c) \cdot P(c)) \quad (1)$$

Donde \mathcal{C} es el conjunto de todas las clases, $N_t(c)$ es el número de clasificadores que asignan la clase c al tweet t , y $P(c)$ es la probabilidad a priori de la clase c calculada utilizando el corpus de entrenamiento.

- **run2** La segunda alternativa explora la combinación de modelos mediante el aprendizaje de un metaclasificador. Utilizando las salidas de los mismos 192 clasificadores que en el run anterior, se ha aprendido un segundo modelo SVM que sirve para proporcionar la nueva salida combinada. Se ha destinado una parte del corpus de entrenamiento para ajustar los parámetros del metamodelo. Esta aproximación es la misma que la utilizada en la edición del TASS 2015.

Para la subtarea de 4 etiquetas el **run1** se ha aprendido utilizando el corpus de aprendizaje con 4 etiquetas mientras que el **run2**, dada la complejidad del ajuste de parámetros del metamodelo se ha optado por adaptar el resultado de la subtarea de 6 etiquetas uniendo P y P+ como P y N+ como N.

En la Tabla 2 se muestran los valores de Accuracy obtenidos para las dos subtareas.

Los sistemas presentados han obtenido las dos primeras posiciones en las dos subtareas consideradas.

	Run	Accuracy
6-ETIQUETAS	run1	0.662
	run2	0.673
4-ETIQUETAS	run1	0.707
	run2	0.721

Tabla 2: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea 1 de la competición TASS-2016 sobre el conjunto de test para 6 y 4 etiquetas.

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 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 la que se utilizó en nuestro sistema del TASS 2015 la cual utiliza 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.

El corpus de la tarea, corpus *STOMPOL*, se compone de un conjunto de tweets relacionados con una serie de aspectos políticos (como economía, sanidad, etc.) 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).

4.1. 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
STOMPOL	run1	0.633

Tabla 3: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea 2 de la competición TASS-2016 para el corpus STOMPOL.

En la Tabla 3 se presentan los resultados obtenidos para la Tarea 2 con lo que nuestra aproximación ha obtenido la primera posición en dicha tarea.

5. Conclusiones y trabajos futuros

En este trabajo se ha presentado la participación del grupo *ELiRF-UPV* en las 2 tareas planteadas en TASS 2016. Nuestro equipo ha utilizado aproximaciones basadas en máquinas de soporte vectorial y se ha centrado principalmente en combinar diferentes sistemas.

Haciendo un análisis del número de participantes y de los resultados obtenidos en las

dos últimas ediciones del TASS, creemos que se está cerca de alcanzar los mejores resultados posibles en la tarea de Análisis de sentimientos tal y como se ha venido planteando hasta el momento.

A la vista de los buenos resultados que se han obtenido mediante la combinación de sistemas, como trabajo futuro nos planteamos desarrollar nuevos métodos de combinación de sistemas más sofisticados así como la inclusión de otros paradigmas de clasificación más heterogéneos (distintos de los SVM) para aumentar la complementariedad de los sistemas combinados.

Además, se pretende extender el sistema para otros idiomas. El sistema descrito ya ha sido utilizado, con ligeras modificaciones, en tareas de análisis de sentimientos para el Inglés en la competición Semeval (Martínez, Pla, y Hurtado, 2016) aunque con resultados no tan satisfactorios como en las tareas del TASS.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el MINECO mediante el proyecto ASLP-MULAN: Audio, Speech and Language Processing for Multimedia Analytics (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.
- García-Cumbreras, Miguel Ángel, Julio Villena-Román, Eugenio Martínez-Cámara, Manuel Carlos Díaz-Galiano, M^a. Teresa Martín-Valdivia, y L. Alfonso Ureña-López. 2016. Overview of tass 2016. En *Proceedings of TASS 2016: Workshop on Sentiment Analysis at SEPLN co-located with the 32nd SEPLN Conference (SEPLN 2016)*, Salamanca, Spain, September.
- 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., Ferran Pla, y Davide Boscaldi. 2015. Elirf-upv en tass 2015: Análisis de sentimientos en twitter. En *SEPLN*.
- 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, Víctor, Ferran Pla, y Lluís-F Hurtado. 2016. Dsic-elirf at semeval-2016 task 4: Message polarity classification in twitter using a support vector machine approach.
- 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 Maguelonne 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.