

ELiRF-UPV en TASS 2014: Análisis de Sentimientos, Detección de Tópicos y Análisis de Sentimientos de Aspectos en Twitter

ELiRF-UPV at TASS 2014: Sentiment Analysis, Topic Detection and Aspect-based 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 TASS2014. Este taller es un evento enmarcado dentro de la XXX edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural. Este trabajo presenta las aproximaciones utilizadas para las cuatro tareas del taller, los resultados obtenidos y una discusión de los mismos.

Palabras clave: Twitter, Análisis de Sentimientos.

Abstract: This paper describes the participation of the ELiRF research group of the Universitat Politècnica de València at TASS2014 Workshop. This workshop is a satellite event of the XXX edition of the Annual Conference of the Spanish Society for Natural Language Processing. This work describes the approaches used for the four tasks of the workshop, the results obtained and a discussion of these results.

Keywords: Twitter, Sentiment Analysis.

1. *Introducción*

Twitter se ha convertido en una de las plataformas más importantes de las redes sociales. La comunidad científica está dedicando grandes esfuerzos al análisis de los tweets, usando técnicas de procesamiento del lenguaje natural y aprendizaje automático, desarrollando recursos lingüísticos que permitan analizar con éxito las peculiaridades de un lenguaje tan específico como es el de Twitter y el de las redes sociales en general. Mediante las técnicas de análisis de sentimientos se intenta determinar el comportamiento y las opiniones de los usuarios de la red sobre diferentes temas de interés como política, economía, productos comerciales, servicios, aficiones, etc. Esto supone nuevos retos en el área de procesamiento del lenguaje natural para proponer soluciones para problemas no resueltos incluso considerando textos normativos.

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 presenta-

das 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 tercera 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) Identificación de tópicos en tweets. Además, se proponen dos tareas nuevas: 3) Determinación de aspectos en el corpus Social_TV, compuesto por tweets publicados durante la final de la Copa del Rey 2014 y 4) Determinación de la polaridad de los aspectos detectados en la tarea anterior.

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 resul-

tados obtenidos. Finalmente se muestran las conclusiones y posibles trabajos futuros.

2. Descripción de los sistemas

Antes de abordar cada una de las tareas se ha realizado un preproceso de los tweets. En nuestro caso se ha utilizado la estrategia descrita en (Pla y Hurtado, 2013) y que se corresponde con nuestra participación en el TASS 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, excepto la tarea 3 (detección de aspectos), 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).

Para tratar el problema de la clasificación multiclase se ha utilizado una estrategia de uno-contra-todos (one-vs-all) donde se aprende un clasificador binario por cada clase capaz de discriminar entre *esa clase* y *no esa clase* (todas las demás). La tarea 2, clasificación de tópicos, además de ser una tarea multiclase, es una tarea multietiqueta, ya que un tweet puede pertenecer a más de un tópico. Para este problema se han seguido dos aproximaciones.

La primera aproximación considera como

una nueva etiqueta cada conjunto de multi-etiquetas que aparece en el conjunto de entrenamiento (a esta aproximación le llamaremos aglutinativa). Esta aproximación presenta dos limitaciones evidentes: a) las combinaciones de etiquetas no vistas en el entrenamiento no pueden ser predichas; y b) las combinaciones poco habituales tendrán pocas muestras de entrenamiento aunque las etiquetas individuales tengan muchas muestras. A pesar de sus limitaciones esta aproximación suele obtener buenos resultados experimentales.

La segunda aproximación no crea nuevas etiquetas, como resultado de las combinaciones vistas en el entrenamiento. En lugar de esto, se permite seleccionar más de una etiqueta *básica* al clasificador. Para ello, se sustituye el criterio de elegir la etiqueta de mayor confianza por el de elegir todas las etiquetas cuyo modelo (one-vs-all) supera un umbral determinado experimentalmente. En el caso de que, para un tweet, ninguna etiqueta supere el umbral, se puede aplicar un criterio alternativo, por ejemplo, seleccionar la más frecuente, o la proporcionada por otro clasificador que utilice la primera aproximación anteriormente descrita.

Se ha utilizado la aproximación de bolsa de características para representar cada tweet como un vector de características que contiene los coeficientes tf-idf de las características consideradas en el tweet. Los parámetros de los clasificadores se han determinado en la fase de ajuste de parámetros usando una validación cruzada de 10 iteraciones (10-fold cross-validation).

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 el 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

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/>

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 diferentes modelos mediante una técnica de votación para aprovechar su complementariedad y mejorar las prestaciones finales. Estos aspectos proporcionaron mejoras significativas 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).

Se han presentado 3 sistemas para la sub-tarea de 6 etiquetas:

- **run1:** 1-gramas de lemas. Las características son los coeficientes td-idf de los lemas de las palabras que aparecen en el tweet más el número de lemas positivos y negativos según el diccionario de lemas.
- **run2:** 1-gramas de palabras. las características son los coeficiente td-idf de las palabras que aparecen en el tweet más el número de palabras positivas y negativas según el diccionario de palabras.
- **run3:** votación entre 6 sistemas de 1, 2 y 3 gramas de palabras y lemas. Se elige la polaridad mayoritaria entre los 6 sistemas.

También para la sub-tarea de 4 etiquetas se presentaron otros 3 sistemas:

- **run1:** 1-gramas de lemas. Las características son los coeficiente td-idf de los

lemas de las palabras que aparecen en el tweet más el número de lemas positivos y negativos según el diccionario de lemas.

- **run2:** el mismo resultado del *run3* de la sub-tarea de 6 etiquetas pero uniendo P y P+ como P y N y N+ como N.
- **run3:** votación entre 6 sistemas de 1, 2 y 3 gramas de palabras y lemas. Se elige la polaridad mayoritaria entre los 6 sistemas.

La diferencia entre el *run2* y el *run3* radica en que el *run2* está entrenado con 6-etiquetas mientras que el *run3* está entrenado con 4-etiquetas.

En la Tabla 2 se muestran los valores de Accuracy obtenidos para las dos sub-tareas. Los sistemas presentado alcanzaron las 3 primeras posiciones en las 2 sub-tareas.

	Run	Accuracy(%)
6-ETIQUETAS	run1	0.6316
	run2	0.6285
	run3	0.6432
4-ETIQUETAS	run1	0.7053
	run2	0.7089
	run3	0.7035

Tabla 2: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea 1 de la competición TASS-2014 sobre el conjunto de test para 6 y 4 etiquetas.

4. Tarea 2: Clasificación de los tweets en tópicos

La tarea consiste en asignar uno o varios tópicos a un tweet. Los conjuntos de entrenamiento y de test son los mismos que para la Tarea 1.

En la Tabla 3 se muestra la distribución por tópico de los tweets del conjunto de entrenamiento.

Para esta tarea se ha utilizado la tokenización previamente descrita. La única diferencia es que no se han considerado entidades nombradas. Además, se ha construido un diccionario de temáticas, en el que para cada hashtag del conjunto de entrenamiento, se le ha asignado el tópico del tweet en el que aparecía. Si un hashtag podía pertenecer a más de un tópico no se ha considerado como entrada del diccionario.

Se han presentado 3 sistemas:

Tópico	# tweets	%
cine	245	2.56
deportes	113	1.18
economía	942	9.84
entretenimiento	1678	17.53
fútbol	252	2.63
literatura	103	1.08
música	566	5.91
otros	2337	24.41
política	3120	32.59
tecnología	217	2.27
TOTAL	9573	100

Tabla 3: Distribución de los tweets del conjunto de entrenamiento por tópico.

- **run1:** multietiqueta aglutinativo. Un clasificador one-vs-all considerando una nueva etiqueta aglutinativa cada conjunto de multietiquetas que aparece en el entrenamiento.
- **run2:** multietiqueta por umbral. Un clasificador one-vs-all sin etiquetas aglutinativas. Se eligen como resultado todas las clases cuyo modelo supere un umbral de confianza determinado mediante validación cruzada.
- **run3:** votación entre 8 sistemas. Los sistemas se diferencian por tener diferentes valores de n-gramas y de parámetros de la SMV. Se eligen los tópicos predichos por al menos 3 sistemas.

Run	P	R	F_1
run1	0.6823	0.6889	0.6856
run2	0.7001	0.7065	0.7033
run3	0.6660	0.7483	0.7048

Tabla 4: Resultados oficiales de Precision (P), Recall (R) y factor F_1 del equipo *ELiRF-UPV* en la Tarea 2 de la competición TASS-2014 sobre el conjunto de test.

En la Tabla 4 se muestran los resultados obtenidos que ocupan las 3 primeras posiciones de la tarea.

5. Tarea 3: Detección de Aspectos

El objetivo de esta tarea es la detección de los diferentes aspectos expresados por los usuarios en sus tweets. Se proporciona una lista de los aspectos a detectar (jugadores,

equipos, entrenadores, autoridades, partidos, retransmisión, etc).

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 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 no contiene ninguna anotación.

Para el desarrollo de nuestro sistema de detección de aspectos, se han obtenido todas las instancias de cada aspecto que aparecen en el conjunto de entrenamiento. Además, se ha completado con una lista de todos los jugadores del Real Madrid y del Barcelona (con sus apodos) y se ha incorporado una lista de todos los equipos de primera división de la liga española. A partir de estos datos se ha definido un conjunto de expresiones regulares para reconocer estas instancias. Se ha permitido cierta flexibilidad en las mismas para admitir errores gramaticales, posibles nombres de hashtags y menciones de usuario que contengan estos nombres.

En la figura 1 se muestra una expresión regular para reconocer al jugador *Leo Messi*, y algunos ejemplos del corpus que han sido detectados con ella.

Nuestro detector aplica en cascada todo el conjunto de expresiones regulares y devuelve, como resultado, el nombre del aspecto (que se debe corresponder con uno de los de la lista suministrada), la instancia concreta del aspecto detectada dentro del tweet y su posición en el mismo.

Los resultados obtenidos sobre el conjunto de test se muestran en la tabla 5. En esta tarea sólo se ha utilizado un sistema obteniendo el primer lugar de la competición.

Run	P	R	F_1
run1	0.9060	0.9111	0.9086

Tabla 5: Resultados oficiales de Precision (P), Recall (R) y factor F_1 del equipo *ELiRF-UPV* en la Tarea 3 de la competición TASS-2014 sobre el conjunto de test.

```
'([#@]?l[eio(nel)?\S*]? (\s)* ([#@]?me(s)?si\S*)?': 'Jugador-Lionel_Messi'
```

EJEMPLOS: Messi, messi, MESSI, mesi, Leo, Lionel, leoo, LEO, #Messi, @Messi, @TeamMessi, #Messisinvergüenza, @MessiSiempreElMejor, @LeoMessi_Barca

Figura 1: Ejemplo de expresión regular para la detección del *Jugador-Lionel_Messi* y algunas de las instancias detectadas en el corpus.

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

Esta tarea es totalmente dependiente de la anterior y consiste en asignar la polaridad a los aspectos detectados anteriormente. Una de las dificultades de la tarea consiste en, una vez detectado el aspecto, definir qué contexto se le asigna 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. La longitud de la ventana se ha determinado experimentalmente sobre el conjunto de entrenamiento. El valor máximo de la ventana considerado ha sido de 7 palabras a izquierda y derecha del aspecto. 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.

A partir de este corpus, se han entrenado tres sistemas:

- **run1:** 1-gramas de lemas sin detección de entidades nombradas. Las características son los coeficiente td-idf de los lemas de las palabras que aparecen en el segmento del tweet correspondiente al aspecto, más el número de lemas positivos y negativos según el diccionario de polaridad de lemas.
- **run2:** 1-gramas de palabras sin detección de entidades nombradas. Las características son los coeficiente td-idf de las

palabras que aparecen en el segmento del tweet correspondiente al aspecto, más el número de palabras positivas y negativas según el diccionario de polaridad de palabras.

- **run3:** 1-gramas de lemas con detección de entidades nombradas. Las características son los coeficiente td-idf de los lemas de las palabras que aparecen en el segmento del tweet correspondiente al aspecto, más el número de lemas positivos y negativos según el diccionario de polaridad de lemas.

Los resultados obtenidos se muestran en la Tabla 6. Los tres runs ocupan las tres primeras posiciones de la clasificación de la competición.

Run	P	R	F_1
run1	0.5740	0.5899	0.5819
run2	0.5783	0.5956	0.5868
run3	0.5638	0.5820	0.5728

Tabla 6: Resultados oficiales de Precision (P), Recall (R) y factor F_1 del equipo *ELiRF-UPV* en la Tarea 4 de la competición TASS-2014 sobre el conjunto de test.

7. Conclusiones y trabajos futuros

En este trabajo se ha presentado la participación del equipo ELiRF-UPV en las 4 tareas planteadas en TASS 2014. 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*. Los sistemas desarrollados por el equipo ELiRF-UPV han alcanzado la primera posición en todas las tareas.

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 aplicar las técnicas aquí desarrolladas a otros corpus y otros idiomas.

Agradecimientos

Este trabajo ha sido parcialmente subvencionado por los proyectos DIANA: DIScourse ANALysis for knowledge understanding (MEC TIN2012-38603-C02-01) y TÍMPANO: Technology for complex Human-Machine conversational interaction with dynamic learning (MEC TIN2011-28169-C05-01).

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.

Martínez-cámara, E., M. T. Martín-valdivia, M. D. Molina-gonzález, y L. A. Ureñalópez. 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.