

ELiRF-UPV en TASS-2013: Análisis de Sentimientos en Twitter*

ELiRF-UPV at TASS-2013: Sentiment Analysis in Twitter

Ferran Pla y Lluís-F. Hurtado

Universitat Politècnica de València

Camí de Vera s/n, 46022 València

(fpla, lhurtado)@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 sobre Análisis de Sentimientos (TASS-2013). Este taller es un evento enmarcado dentro de la XXIX edición del Congreso Anual de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). El TASS-2013 es la segunda edición de esta competición y se proponen cuatro tareas. La primera consiste en determinar el análisis de sentimientos a nivel global de un tweet; la segunda, consiste en determinar el tópico de un tweet (política, economía, deportes, ...); la tercera consiste en el análisis de sentimiento a nivel de entidad dentro de un tweet y la última tarea consiste en la identificación de las tendencias políticas (derechas, centro, izquierdas, neutral) de un usuario basándose en sus tweets. Nuestro equipo ha participado en todas las tareas obteniendo buenos resultados en todas ellas. Este trabajo presenta las aproximaciones utilizadas, los resultados obtenidos y una discusión de los mismos.

Palabras clave: Twitter, Análisis de Sentimientos, Aprendizaje Automático.

Abstract: This paper describes the participation of the ELiRF research group of the Universitat Politècnica de València in the Workshop (TASS, 2013). This workshop is a satellite event of the XXIX edition of the Annual Conference of the Spanish Society for Natural Language Processing (SEPLN). TASS-2013 is the second edition of this competition and proposes four tasks. The first one is to determine the analysis of the global sentiment of a tweet, the second task is to determine the topic of a tweet (politics, economy, sports, ...), the third task is to determine the sentiment analysis at entity level within a tweet, and the fourth task is to identify the political trends of a given user (right, center, left, neutral) based on his tweets. Our team has participated in all the tasks with good results in all them. This work describes the approaches used, the results obtained and a discussion of these results.

Keywords: Twitter, Sentiment Analysis, Machine Learning.

1. *Introducción*

La gran cantidad de información contenida en los medios sociales ha propiciado que la comunidad científica dedique grandes esfuerzos a analizar, estructurar y procesar esta información. Entre otras cosas, estos medios se utilizan para expresar opiniones y sentimientos diversos sobre diferentes aspectos de la sociedad, productos, servicios, aficiones, etc. Por eso, empresas, organizaciones, gobiernos

y diferentes colectivos en general, han mostrado también su interés en conocer las opiniones y sentimientos que los usuarios tienen sobre sus actividades. En ese sentido, Twitter se ha convertido en una excelente herramienta para conocer en tiempo real las opiniones que los usuarios expresan sobre una gran variedad de temas. La naturaleza de los textos usados en Twitter (tweets), textos cortos de un máximo de 140 caracteres y un lenguaje, en algunos casos, agramatical y específico del medio (emojiconos, abreviaturas, terminología específica, jergas, etc.), plantea nuevos retos para su procesamiento. Las técnicas usuales de procesamiento del lenguaje na-

* 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)

tural (PLN) deben adaptarse a estas características del lenguaje y se deben proponer nuevas aproximaciones para poder afrontar con éxito el problema.

El análisis de opiniones (sentiment analysis, SA) viene siendo estudiado desde hace más de una década. El problema se ha abordado usando técnicas de aprendizaje automático, en el trabajo de (Pang, Lee, y Vaithyanathan, 2002) o desde aproximaciones basadas en el conocimiento o no supervisadas en (Turney, 2002). Estos trabajos pioneros en el área, estudian y proponen métodos para determinar la polaridad, en concreto, sobre el dominio de opiniones sobre películas. Muchos trabajos posteriores han seguido usando estas técnicas, combinaciones de ambas y nuevas propuestas, sobre este y otros dominios. Una buena revisión de las mismas la podemos encontrar en (Liu, 2012).

Los trabajos de SA sobre Twitter son mucho más recientes. Twitter aparece en el año 2006 y los primeros trabajos en este campo son de 2009, cuando Twitter empieza a alcanzar popularidad. Algunos de los trabajos más significativos son (Jansen et al., 2009), (Connor, Krieger, y Ahn, 2010), (Barbosa y Feng, 2010). Una buena recopilación de las aproximaciones más relevantes al SA en Twitter se pueden consultar en los siguientes trabajos: (Vinodhini y Chandrasekaran, 2012), (Martínez-Cámara et al., 2012). La competición SemEval-2013 también ha dedicado una tarea específica para el SA en Twitter (Wilson et al., 2013), lo cual demuestra el notable interés despertado en la comunidad científica en esta temática.

En las siguientes secciones se describen los sistemas utilizados para las cuatro tareas planteadas en el taller TASS-2013. En primer lugar se presentan las herramientas utilizadas para el preproceso de los tweets (tokenizadores, lematizadores, analizadores morfosintácticos, desambiguadores morfosintácticos, etc.) y la adaptación realizada para abordar las tareas. A continuación se describe cada una de las tareas abordadas en este trabajo, las soluciones propuestas así como la presentación y análisis de los resultados obtenidos. Finalmente se incluyen una serie de conclusiones y posibles trabajos futuros.

2. Preproceso de los tweets

Antes de abordar cualquiera de las tareas planteadas en TASS-2013, es necesario reali-

zar una adecuada tokenización de los tweets que constituyen el corpus de entrenamiento y de evaluación. Aunque existen una gran cantidad de tokenizadores disponibles en la red, es necesaria una adaptación de los mismos para abordar la segmentación en tokens de un tweet. Esto es debido, como se ha comentado anteriormente, a la naturaleza del lenguaje empleado en Twitter (frases agramaticales, ausencia o mal uso de los símbolos de puntuación, terminología específica, jergas, etc.). Además, la mayor parte de estos recursos son para el inglés, lo cual añade un grado de dificultad para su uso.

Por otra parte, la utilización de otros recursos básicos en el PLN de textos normativos, como lematizadores, analizadores morfosintácticos, analizadores sintácticos, detectores de entidades, desambiguadores semánticos, etc., se hace impracticable si no se tienen en cuenta las características de los tweets y se proponen mecanismos correctivos que hagan factible el uso de esas herramientas. Resolver adecuadamente la tokenización es de suma importancia para obtener buenos resultados en las tareas planteadas.

En nuestro caso, se decidió utilizar y adaptar herramientas disponibles en la red para la tokenización, la lematización y el etiquetado morfosintáctico (POS tagging). No se han utilizado otro tipo de análisis más profundo para abordar las tareas del TASS-2013 debido a que se estimó que no aportarían ningún beneficio a las propuestas planteadas en este trabajo.

Se ha utilizado y se ha adaptado para el procesamiento del español, el tokenizador para tweets *Tweetmotif* que se describe en (Connor, Krieger, y Ahn, 2010)¹. También se ha utilizado como tokenizador la herramienta *Freeling* (Padró y Stanilovsky, 2012)², en concreto para realizar la agrupación de ciertos grupos de palabras de interés. Además, *Freeling* se ha usado como lematizador y como etiquetador morfosintáctico, con las correspondientes modificaciones para las tareas.

Las modificaciones realizadas en *Tweetmotif* fueron las siguientes. Se reescribieron algunas expresiones regulares para contemplar el uso de los acentos, diéresis y letras específicas del español como la ñ. Respecto al tratamiento de los emoticonos, se realizó una función que clasificaba los emoticonos detec-

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

²<http://nlp.lsi.upc.edu/freeling/>

tados en cinco categorías: *happy*, *sad*, *tongue*, *wink* y *other*. Se incorporó una función con el fin de realizar una normalización de ciertas palabras y abreviaciones usuales en Twitter a una forma canónica. Por ejemplo, cambiar $d \rightarrow de$, $q \rightarrow que$, $pq \rightarrow porque$, $dl \rightarrow del$, etc.

Con respecto a *Freeling* se modificaron los ficheros de configuración para que mantuviera la tokenización producida por *Tweetmotif*, como por ejemplo *hashtags* (#tema), menciones de usuarios (@user), emoticonos, *http*, *www*, signos de puntuación. Sí que se permitía que se produjeran agrupamientos entre palabras (Por_lo.tanto), nombres propios (Antonio_Pérez), fechas (12.de.enero.de.2013), lugares (Santiago.de.Compostela), etc.

Además, se modificaron los diccionarios para que a ciertos tokens se le asignara una categoría gramatical fija, por ejemplo, #tema, @user: Nombre, emoticonos: Intersección, http, web: Signo de puntuación (Fz), etc. Finalmente, con todo esto se realizó la tokenización y lematización de los tweets.

También se desarrollaron funciones que permitían la unificación de ciertos tokens. Por ejemplo, agrupar todos los *hashtags* en un único token, agrupar todas las direcciones *www* en una o agrupar todas las *url* en una. Considerar los números o las fechas o los signos de puntuación, como un único token, etc. Estas funcionalidades se utilizaron principalmente en el proceso de selección de características en cada una de las tareas abordadas.

3. Tarea 1: Análisis de sentimientos en tweets

Esta tarea consiste en determinar la polaridad de los tweets en seis niveles: 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.

Los datos proporcionados por la organización del TASS-2013 consta de un conjunto de entrenamiento, compuesto por 7219 tweets etiquetados con esta polaridad, 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.

La tarea se ha abordado como un problema de clasificación, utilizando la herramienta

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.

ta WEKA³, que incluye, entre otras utilidades, una colección de algoritmos de aprendizaje automático para el problema de clasificación. En esta tarea se ha utilizado el algoritmo SVM usando la librería externa *LibSVM*⁴, por ser un software muy eficiente para construir clasificadores SVM y por permitir una fácil integración con WEKA y el uso de todas sus funcionalidades.

Se ha utilizado la aproximación *bag-of-words* para representar cada tweet como un vector de características que contiene la frecuencia de las características seleccionadas del conjunto de entrenamiento. A partir del proceso de tokenización, lematización, etiquetado morsosintáctico y normalización presentado en la Sección 2, se ha determinado experimentalmente el mejor conjunto de características para la tarea.

El proceso de selección de características se ha realizado utilizando el conjunto de entrenamiento mediante una validación cruzada (5-fold validation) para optimizar los parámetros (características y parámetros del algoritmo de aprendizaje SVM).

El conjunto de características tenidas en cuenta en esta tarea ha sido:

1. Se consideraron sólo unigramas de lemas obtenidos en el preproceso de los tweets de cierta frecuencia mínima preestablecida (f).
2. Se consideraron todos los *hashtags* (#tag) como una sola característica.
3. Se consideraron todas las menciones a los usuarios (@user) como una sola característica.
4. Se unificaron todos los números en un sola característica.

³<http://www.cs.waikato.ac.nz/ml/weka/>

⁴<http://www.cs.iastate.edu/~yasser/wlsvm/>

5. Se unificaron todas las fechas en una sola característica.
6. Se unificaron todos los signos de puntuación de un mismo tipo en una sola característica.
7. Se sustituyó cada emoticono por su correspondiente categoría: *happy, sad, tongue, wink y other*.
8. Se consideraron sólo como característica los tokens pertenecientes a ciertas categorías morfosintácticas preestablecidas (*selPOS*).
9. Se utilizaron como recurso externo, léxicos de polaridad de lemas y palabras (*DIC*).

Aparte de las características anteriormente enumeradas, se tuvieron en cuenta otras, que no mejoraron los resultados finales, y que por lo tanto, se descartaron. Éstas fueron: considerar la etiqueta morfosintáctica, incluir la negación, uso de bigramas de características, inclusión del tipo y del nombre del usuario del tweet, etc.

De la lista de características enumeradas anteriormente, algunas necesitaban un ajuste adicional. Por ejemplo, en la característica (1) hay que estimar cuál es la frecuencia de los lemas a considerar (f). En la (8) hay que determinar qué lemas se consideran, dependiendo de su categoría morfosintáctica en ese contexto. Por último hay que determinar si se utilizan los léxicos externos (9) de palabras y/o de lemas (*DIC*).

Los léxicos externos utilizados estaban constituidos por listas de palabras y lemas con su correspondiente polaridad a priori. Uno de los léxicos utilizados estaba originalmente para el inglés (Wilson et al., 2005) por lo que fue automáticamente traducido al español usando diccionarios, y el otro (Perez-Rosas, Banea, y Mihalcea, 2012) es una lista de palabras ya en español⁵.

En la Tabla 2 se puede observar cómo afecta a la precisión de la tarea la elección de las características. La primera fila indica que sólo se han considerado lemas de frecuencia mayor o igual que dos en el conjunto de entrenamiento ($f=2$). La segunda fila indica que se consideran los lemas de $f=2$ que además

pertenezcan a un conjunto restringido de etiquetas morfosintácticas. En nuestro caso, la mejor selección fue considerar sólo *nombres, verbos, adjetivos, adverbios y exclamaciones* (los emoticonos pertenecen a esa categoría). Las siguientes filas representan las mismas características pero incluyendo los léxicos de polaridad de palabras y lemas (*DIC*).

Selección	#Caract.	P (%)
f=2	4855	46.05
f=2+selPOS	4737	44.69
f=2+DIC	4859	47.35
f=2+selPOS+DIC	4741	47.49
f=1	12201	45.30
f=1+selPOS	12071	44.83
f=1+DIC	12205	47.23
f=1+selPOS+DIC	12075	46.37

Tabla 2: Resultados de Precisión(P), y el número de características del modelo (#Caract.) dependiendo de las características elegidas y realizando una validación cruzada sobre el conjunto de entrenamiento (*5-fold validation*).

Con las características que mejor resultados se obtuvieron sobre el conjunto de entrenamiento, se construyeron nuevos modelos y se aplicaron al conjunto de test proporcionado por la organización del TASS-2013. Los resultados para las dos salidas enviadas sobre el conjunto de test se muestran en la Tabla 3. A partir de estas salidas, la organización calculó los resultados para 3 niveles de polaridad, juntando N y N+ en N, y P+ y P en P. Estos resultados también se muestran en la Tabla 3.

Run	Selección	P (%)
R3_5niveles	f=1+DIC	57.30
R2_5niveles	f=2+selPOS+DIC	57.60
R3_3niveles	f=1+DIC	67.40
R2_3niveles	f=2+selPOS+DIC	67.40

Tabla 3: Resultados oficiales del equipo *ELiRF-UPV* en la Tarea 1 de la competición TASS-2013 sobre el conjunto de test para 5 y 3 niveles de polaridad respectivamente.

⁵Gracias a Joan Puig Cerver Pérez, estudiante de la asignatura Aplicaciones de la Lingüística Computacional del máster IARFID del DSIC, UPV, junio 2013, por recopilar y traducir estos recursos.

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 la Tarea 1. El conjunto de entrenamiento, compuesto por 7219 tweets etiquetados con la polaridad y los tópicos a que pertenece, y el conjunto de test de 60798 tweets, al cual se le debe asignar el tópico.

En la Tabla 4 se muestra la distribución por tópico de los tweets del conjunto de entrenamiento.

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 4: Distribución de los tweets del conjunto de entrenamiento por tópico.

La tarea se ha abordado como un problema de clasificación, utilizando la herramienta WEKA, pero en este caso se ha utilizado el método *SMO*. Se entrenó un clasificador binario para cada uno de los tópicos, de tal manera que clasificaba en ese *tópico* o en la clase *no_tópico*. Estos clasificadores se aplicaron en cascada a todo el conjunto de test y posteriormente se proporcionaba como salida para cada tweet los tópicos que habían sido asignados por cada clasificador.

Como podía darse el caso de que a algún tweet no se le asignara ningún tópico, se entrenaron una serie de clasificadores auxiliares para cada tópico, pero en este caso utilizando la librería *libSVM*.

SMO no proporciona una probabilidad a la clase elegida, la salida es 1 para la clase elegida, y 0 para las demás. Por eso se utilizó como alternativa la librería *libSVM* para construir los clasificadores auxiliares, ya que ésta sí que asigna una probabilidad a todas las clases. En caso de que a un tweet no se le asignara una clase de las esperadas (un tópi-

co), utilizando el método *SMO*, se elegía la clase más probable proporcionada por los clasificadores auxiliares obtenidos con *libSVM*.

Los tweets fueron preprocesados de la misma manera que en la Tarea 1, y se utilizó la misma técnica descrita anteriormente para la elección del conjunto de características óptimo utilizando el conjunto de entrenamiento.

Las características que mejores resultados proporcionaron en el proceso de ajuste de los clasificadores fue utilizar los *lemas* de los tweets cuya frecuencia en el conjunto de entrenamiento fuera mayor o igual a 1 ($f=1$) y cuya categoría gramatical fuera *nombre* o *verbo*. En esta tarea no se ha utilizado ningún otro recurso externo.

Los resultados obtenidos con esta configuración sobre el conjunto de test son del 75.60% de Precisión, Cobertura y valor de F_1 .

5. Tarea 3: Análisis de sentimientos a nivel de entidad en los tweets

Esta tarea consiste en realizar un análisis de sentimientos, similar a la Tarea 1, pero determinando la polaridad de las entidades que aparecen en el tweet, no la polaridad global del mismo. Para ello, la organización proporcionó un nuevo corpus de test, el corpus *Politics*, que consta de 2079 tweets y 2596 entidades extraídas de esos tweets. También se proporcionó una lista de las entidades⁶ a analizar, para que el problema fuera la clasificación de las mismas y no el de su detección.

El problema consiste en determinar, para cada entidad de la lista proporcionada, su polaridad, pero en este caso a tres niveles (P, N, NEU, NONE). Como recurso para entrenar y validar los modelos, se podía utilizar el corpus de entrenamiento de las tareas 1 y 2, en el que también estaban detectadas y etiquetadas las entidades con su polaridad.

La tarea se ha abordado como la Tarea 1, usando WEKA y el algoritmo SVM con la librería externa *LibSVM*. Como tokens se usaron los lemas de $f=2$, utilizando los léxicos externos (DIC) y sin realizar en este caso

⁶En esta lista había algunos errores que nuestro equipo corrigió. En concreto se trataba de entidades que no aparecían exactamente igual en el tweet. Según nuestros experimentos eran 398 entidades. Algunos ejemplos de estos errores son: psoe→@psoe, #votapsoe; pp→#votapp, #paremosalPP; #rajoy→#RajoyenlaSER; #rubalcaba→#rubalcabaenlaser, etc.

ninguna selección por tipo de categoría morfosintáctica (selPOS).

La complejidad adicional de la tarea radica en cómo determinar qué parte del tweet se refiere a cada una de las entidades presentes en él. Para resolver este problema se debería hacer un análisis sintáctico del tweet y estudiar este tipo de dependencias. En nuestro caso se ha optado por establecer una serie de heurísticas para determinar qué segmento del tweet se refiere a cada una de las entidades presentes.

La Tabla 6 ilustra algunos ejemplos genéricos de los criterios que se han considerado para la citada segmentación.

1 Entidad	
1.1	xxxxxx[E1xxxxxx]
1.2	[E1xxx.]xxxxxxxx
1.3	[xxxxxxxxxxxxxE1]
2 Entidades	
2.1	[E1xxx][E2xxxxxx]
2.2	[E1E2xxxxxxxxxx]
2.3	[E1xxxxxxxxxxE2]
2.4	[xxxxxxxxxxE1E2]
2.5	[xxx[E1xxxxxx]E2]
2.6	xx[E1xxx][E2xxxx]
...	

Tabla 5: Ejemplos de segmentos de tweets y su correspondiente entidad.

Para el caso de que el tweet contenga sólo una entidad, en 1.1 se indica que si está a mitad de tweet el contexto utilizado para determinar la polaridad de la entidad E1 es desde la entidad hasta el final del tweet. En 1.2 se expresa que si la entidad E1 está al principio del tweet, su contexto será hasta el primer signo de puntuación, en nuestro caso ”.”, y ”;”. En 1.3, si la entidad está al final del tweet, el contexto se considera todo el tweet.

Para el caso de dos entidades en el tweet, la casuística es mayor. Si las entidades E1 y E2 están juntas al principio o al final del tweet (2.2, 2.4) se considera todo el tweet como contexto para ambas entidades. Por el contrario, si están separadas, y tienen suficiente contexto, se segmentan como en los ejemplos 2.1 y 2.6. También se puede dar los casos 2.3 y 2.5 en los que o bien una entidad se le puede asignar un segmento y a la otra

no, o las 2 entidades tienen el mismo contexto.

Aparte de los problemas de segmentación comentados, debido a la longitud de los tweets, muchas veces el contexto de una entidad es tan pequeño que no contiene suficiente información para poderlo clasificar correctamente. En esos casos la opción que se ha elegido es establecer un umbral de contexto, y si está por debajo de ese umbral, se le asigna la misma polaridad a todas las entidades del tweet. Relacionado con este problema, cuando el número de entidades es mayor que dos, los contextos son muy cortos, por lo que en esos casos a todas las entidades se le asigna la misma polaridad global del tweet.

Nuestra participación en esta tarea se limita a dos salidas, la primera (*run0*), en la que se asigna la misma polaridad global a todas las entidades del tweet y la segunda (*run1*) en la que se aplica el método de segmentación anteriormente expuesto.

Los resultados oficiales de *Precisión*, *Cobertura* y valor de F_1 , proporcionados por la organización, son de 39.50 % para el *run0* y de 35.80 % para el *run1*. Como se puede observar con los criterios de segmentación presentados se obtienen resultados globales inferiores a utilizar el caso base en el que se asigna la misma polaridad a todas las entidades de un tweet.

6. Tarea 4: Determinación de la tendencia política de los usuarios

El objetivo de la tarea 4 era clasificar a cada uno de los usuarios del corpus general en función de su orientación política. A cada usuario se le debía asignar una etiqueta entre las cuatro posibles: *LEFT*, *RIGHT*, *CENTRE* y *UNDEFINED*.

Para realizar la clasificación de los usuarios partimos de la siguiente hipótesis: las opiniones positivas sobre un partido político supone una orientación política similar por parte del usuario que realiza esa opinión a la del partido; por el contrario, una opinión negativa sobre un partido supone una orientación política contraria a la representada por el partido.

De esta forma, para clasificar a los usuarios por su orientación política, en primer lugar identificamos las entidades relacionadas con partidos políticos y en segundo lugar analizamos la polaridad de esas entidades en los

tweets de cada usuario.

Consideramos tres tipos de entidades:

- Entidades etiquetadas por *Freeling* como nombres propios (NP), por ejemplo: comité_de_dirección_del_pp_de_madrid
- Usuario de Twitter, por ejemplo: @38congresopsoe
- Etiquetas (hashtags) de Twitter, por ejemplo: #upydfacts

De entre todas las posibles entidades del vocabulario seleccionamos aquellas que contuvieran las siglas de algún partido político o el nombre de algún dirigente. Los partidos considerados fueron: *CIU*, *IU*, *PP*, *PSOE*, *UPYD*; y los dirigentes: *Cayo Lara*, *Rajoy*, *Rubalcaba*, *Zapatero* y *González Pons*. En total se identificaron 864 entidades asociadas a partidos políticos. Asociamos un valor *Tendency* de -1 a aquellas entidades asociadas a partidos identificados como de izquierdas, un valor de +1 a aquellas entidades asociadas a partidos de derechas y un valor de 0 a los partidos considerados de centro.

A continuación detallaremos la manera en que se ha calculado numéricamente la orientación política de los usuarios. Para cada usuario A_i del corpus general definimos un conjunto T_i formado por todos los tweets de este usuario que contienen entidades políticas. A aquellos usuarios que no tienen ningún tweet que contenga entidades políticas se le asigna la etiqueta *UNDEFINED*.

Para cada tweet $T_{i_j} \in T_i$, $j = 1 \dots |T_i|$, identificamos las entidades políticas que contiene. Al conjunto de entidades del tweet T_{i_j} lo denominamos E_{i_j} y a cada una de las entidades que contiene la llamamos $E_{i_{j_k}} \in E_{i_j}$, $k = 1 \dots |E_{i_j}|$.

La determinación de la polaridad de cada entidad se realizó igual que para el *run1* de la Tarea 3 descrito en la Sección 5. La única diferencia ha consistido en asignar un valor numérico a la polaridad en lugar de una etiqueta. De este modo a las entidades con polaridad positiva (etiqueta P en la tarea 3) se les asignaba un valor de *Polarity* igual a +1, a las entidades con polaridad negativa (etiqueta N en la tarea 3) se les asignaba un valor de -1 y a las entidades sin polaridad (NEU o NONE en la tarea 3) se les asignaba un valor de 0.

Los valores de orientación política (valor *Tendency*, T) y polaridad (valor *Polarity*,

P) de cada entidad en cada Tweet del usuario se combinan para obtener la tendencia política (Political Tendency, PT) del usuario. Para nuestra participación en esta tarea se diseñaron 4 combinaciones distintas de *Tendency* y *Polarity*. En la ecuación 1 se detalla la combinación que obtuvo el mejor resultado en la competición (run1).

$$PT(A_i) = \frac{\sum_{j=1 \dots |T_i|} \sum_{k=1 \dots |E_{i_j}|} P(E_{i_{j_k}}) \cdot T(E_{i_{j_k}})}{\sum_{j=1 \dots |T_i|} \sum_{k=1 \dots |E_{i_j}|} 1} \quad (1)$$

A aquellos usuarios cuyo PT es cercano a 0 se les asigna la etiqueta *CENTRE*, a aquellos con un PT menor que 0 se les asigna la etiqueta *LEFT* y aquellos con un PT mayor que cero se les asigna la etiqueta *RIGHT*.

En la tabla 6 se presentan los resultados para todos los runs enviados para la Tarea 4.

Runs	Precision (%)
run1	70.30
run4	69.60
run2	67.70
run3	65,80

Tabla 6: Resultados oficiales del grupo ELiRF-UPV para la Tarea 4.

7. Conclusiones

En este trabajo se ha presentado la participación del equipo ELiRF-UPV en las 4 tareas planteadas en TASS-2013. Nuestro equipo ha utilizado básicamente técnicas de aprendizaje automático, en concreto, la aproximación de SVMs. Para ello, ha utilizado la herramienta WEKA y la librería externa LibSVM. Los resultados obtenidos han sido similares, y en algunos casos, superiores a los de los mejores equipos de la competición.

Un problema primordial en todas las tareas de TASS-2013, es realizar una adecuada y correcta tokenización de los tweets. En ese sentido, sería interesante que la organización proporcionara los tweets ya tokenizados con el fin de poder comparar las diferentes aproximaciones a los problemas planteados, sin que este aspecto influyera en la comparativa, considerando únicamente el método y las características tenidas en cuenta en su resolución.

Otro aspecto importante de las tareas del TASS2013 es el conjunto de herramientas lingüísticas disponibles que se pueden utilizar. El tipo de textos a procesar, textos cortos, muchas veces agramaticales, con jergas específicas de Twitter, requiere de una adaptación para poder utilizarlas con éxito. La comunidad científica debería dedicar un esfuerzo especial a estos problemas, desarrollando y/o adaptando recursos que se adecuen a las características del lenguaje usado en los medios sociales. Así se hace necesario herramientas específicas como lematizadores, normalización del vocabulario, desambiguación morfosintáctica, detectores y reconocedores de entidades, etc., adaptados a las tareas abordadas.

Bibliografía

- Barbosa, Luciano y Junlan Feng. 2010. Robust sentiment detection on twitter from biased and noisy data. En *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, páginas 36–44. Association for Computational Linguistics.
- 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.
- Jansen, Bernard J, Mimi Zhang, Kate Sobel, y Abdur Chowdury. 2009. Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60(11):2169–2188.
- Liu, Bing. 2012. *Sentiment Analysis and Opinion Mining. A Comprehensive Introduction and Survey*. Morgan & Claypool Publishers.
- Martínez-Cámara, Eugenio, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López, y Arturo Montejo-Raéz. 2012. Sentiment analysis in twitter. *Natural Language Engineering*, 1(1):1–28.
- 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.
- Pang, Bo, Lillian Lee, y Shivakumar Vaithyanathan. 2002. Thumbs up? sentiment classification using machine learning techniques. En *IN PROCEEDINGS OF EMNLP*, páginas 79–86.
- Perez-Rosas, Veronica, Carmen Banea, y Rada Mihalcea. 2012. Learning sentiment lexicons in spanish. En Nicoletta Calzolari (Conference Chair) Khalid Choukri Thierry Declerck Mehmet Uğur Doğan Bente Maegaard Joseph Mariani Jan Odijk, y Stelios Piperidis, editores, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may. European Language Resources Association (ELRA).
- Turney, Peter D. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. En *ACL*, páginas 417–424.
- Vinodhini, G y RM Chandrasekaran. 2012. Sentiment analysis and opinion mining: A survey. *International Journal*, 2(6).
- Wilson, Theresa, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff, y Siddharth Patwardhan. 2005. Opinionfinder: A system for subjectivity analysis. En *Proceedings of HLT/EMNLP on Interactive Demonstrations*, páginas 34–35. Association for Computational Linguistics.
- Wilson, Theresa, Zornitsa Kozareva, Preslav Nakov, Sara Rosenthal, Veselin Stoyanov, y Alan Ritter. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. *Proceedings of the International Workshop on Semantic Evaluation, SemEval*, 13.