

SINAI en TASS 2012

SINAI at TASS 2012

**Eugenio Martínez Cámara, M. Ángel García Cumbreras, M.
Teresa Martín Valdivia, L. Alfonso Ureña López**

Departamento de Informática, Escuela Politécnica Superior de Jaén
Universidad de Jaén, E-23071 – Jaén
{emcamara, magc, maite, laurena}@ujaen.es

Resumen: En el presente artículo se describe la participación del grupo de investigación SINAI de la Universidad de Jaén en la primera edición del taller sobre Análisis de Sentimientos en el congreso de la SEPLN (TASS 2012). El Taller propone dos tareas, una centrada en la determinación de la polaridad de *tweets* en español, y una segunda en la que hay que identificar los temas a los que pertenecen los *tweets*. Para la primera tarea se ha optado por una estrategia de aprendizaje automático supervisado, siendo SVM el algoritmo elegido. En cuanto a la segunda tarea, también se ha utilizado SVM, y con el fin de mejorar el resultado de la clasificación se ha combinado con bolsas de palabras de cada uno de los temas.

Palabras clave: Twitter, Análisis de Sentimientos, Análisis de la Opinión, método supervisado, SVM.

Abstract: In this paper is described the participation of the SINAI research group of the University of Jaén in the first edition of the workshop on Sentiment Analysis at the SEPLN congress (TASS 2012). The Workshop includes two tasks, the first one is focused in the polarity classification of a corpus of Spanish tweets, and the second one involves a topic classification. For the first task, we have chosen a supervised machine learning approach, in which we have used SVM for classifying the polarity. In the second task, we have also used SVM for the topic classification but several bags of words have been used with the goal of improving the classification performance.

Keywords: Twitter, Sentiment Analysis, Opinion Mining, Supervised Machine Learning, SVM

1 *Introducción*

En este artículo se presentan los experimentos y resultados obtenidos en el Taller de Análisis de Sentimientos en la SEPLN (TASS 2012) (Villena-Román et al., 2013). Concretamente se ha participado en las dos tareas propuestas: Sentiment Analysis y Trending Topic Coverage.

El Análisis de Sentimientos (AS), también conocido como Minería de Opiniones (MO) o Análisis de Opiniones (AO), se ha convertido en una prometedora disciplina de investigación que se encuadra dentro del Procesamiento del Lenguaje Natural (PLN) y la Minería de Datos. Se suele definir como el tratamiento computacional de la información subjetiva

presente en cualquier tipo de documento (Pang and Lee, 2008).

La proliferación de contenidos web generados por los propios usuarios en blogs, wikis, foros o redes sociales ha motivado que empresas, investigadores y organizaciones se interesen por analizar y monitorizar toda esta información que circula por la red. Es por ello, que cada vez en más foros se presentan artículos científicos, conferencias o proyectos relacionados con el AO.

Por otra parte, el AO se ha tratado fundamentalmente sobre textos extensos como por ejemplo documentos en blogs o artículos de opinión. Sin embargo, debido al enorme éxito de las redes sociales, el interés para analizar las opiniones en textos cortos

está creciendo de manera exponencial.

Aunque se trata de un área de investigación relativamente nueva, existen una gran cantidad de trabajos relacionados con AO, y más específicamente con la clasificación de la polaridad. Se pueden distinguir dos formas de tratar el problema. La primera se basa en técnicas de aprendizaje automático (Pang et al., 2002), y la segunda se fundamenta en el concepto de orientación semántica, que no necesita el entrenamiento de ningún algoritmo, pero sí debe tener en cuenta la orientación de las palabras (positiva o negativa) (Turney, 2002). En este trabajo se intentan combinar ambas aproximaciones con el fin de mejorar la precisión de los sistemas.

Además, la mayoría de las investigaciones se han centrado en textos en inglés aunque está claro que cada día más, otros idiomas como el chino o el español están más presentes en internet. Precisamente el taller que es la base de este artículo utiliza un corpus en español.

El resto del artículo se organiza como sigue: la siguiente sección incluye una pequeña revisión del estado del arte sobre AO en Twitter. A continuación, se presenta una breve introducción a la categorización de texto en Twitter. La sección 4 muestra el corpus y el proceso de preparación de los datos. En las secciones 5 y 6 se presenta la experimentación y resultados en las dos tareas abordadas.

2 *Análisis de la opinión en Twitter*

Cada día más, los usuarios de Internet utilizan las redes sociales para expresar sus sentimientos y sensaciones sobre cualquier tema. Un ejemplo de esto son las redes de micro-blogging, como Twitter, en la que en tiempo real los usuarios expresan sus opiniones sobre los temas más variados. En España la presencia de Twitter ha ido creciendo paulatinamente, y es a partir de 2010 cuando empresas, políticos y usuarios en general se están dando cuenta del verdadero potencial de esta red social. Sería de gran utilidad el poder determinar de forma automática la polaridad de esas opiniones permitiendo desarrollar sistemas que se encarguen de estudiar y analizar la intención de voto de los ciudadanos, la opinión de consumidores sobre algún

producto o servicio concreto o el estado de ánimo de las personas.

Los *tweets* tienen características que los hacen diferentes de las opiniones y comentarios que hay en foros y páginas web. Normalmente los comentarios u opiniones que se escriben en Internet suelen ser textos más o menos extensos en los que los usuarios intentan resumir lo que piensan sobre un determinado tema, pero los *tweets* suelen estar escritos en un lenguaje informal, y su extensión está limitada a 140 caracteres. Por otra parte, muchos *tweets* no expresan opiniones sino situaciones que les ocurren a los usuarios. Por último, los *tweets* tienen que ser analizados a nivel de frase, y no a nivel de documento.

La mayor parte de la investigación publicada sobre AO en Twitter se ha realizado sobre *tweets* en inglés debido a que la explosión de popularidad de esta red social es relativamente reciente en países de habla no inglesa. Existen, por ejemplo, algunos trabajos que utilizan Twitter a modo de corpus. Petrovic et al. (2010) crean un gran corpus con 97 millones de *tweets*. Pak and Paroubek (2010) describen cómo generar de forma automática un corpus de *tweets* positivos, negativos y neutros. El corpus que crean lo utilizan para entrenar un clasificador de sentimientos. Go et al. (2009) usan técnicas de aprendizaje automático para construir un clasificador que les permita determinar la polaridad de los *tweets*. Para el etiquetado del corpus en *tweets* negativos y positivos siguen la misma estrategia que se describe en (Read, 2005).

Jansen et al. (2009) demuestran como los sitios de micro-blogging son una herramienta muy útil en marketing, e indica que los *tweets* pueden considerarse como Electronic Word Of Mouth (EWOM). Siguiendo esta línea de utilizar Twitter como otra herramienta más en marketing, en (Asur and Huberman, 2010) se utiliza un corpus de *tweets* sobre un conjunto de películas que se estrenaron a finales de 2009 y principios de 2010, para demostrar la correlación existente entre la cantidad de *tweets* y su polaridad sobre una determinada película, con la recaudación en taquilla que ha obtenido en las dos primeras semanas desde su estreno. Bollen et al. (2011) investigan la posible correlación del estado de ánimo que se manifiesta en Twitter con la variación de los

mercados de valores.

El estudio de la tendencia de la opinión política también ha sido un tema que ha atraído a la comunidad científica. O'Connor et al. (2010) analizan opiniones políticas y sobre productos comerciales y comparan los resultados usando Twitter y encuestas. Diakopoulos and Shamma (2010) clasifican la polaridad de los *tweets* durante el debate presidencial en los Estados Unidos del año 2008.

3 Categorización de texto en Twitter

La categorización de textos es una tarea que consiste en la clasificación de los textos en distintas clases predefinidas. Algunas publicaciones han estudiado este problema sobre *tweets*, por ejemplo Sriram et al. (2010) proponen una aproximación que categoriza *tweets* dependiendo del texto que contienen en un conjunto predefinido de clases genéricas como noticias, eventos, opiniones, tratos o mensajes privados. Por su parte Garcia et al. (2010) realizan una comparación entre dos redes sociales (Blippr y Twitter) clasificando en las categorías Movies, Books, Music, Apps y Games. Los resultados para ambas redes sociales en cada una de las categorías son muy parecidos. Por último, distintas técnicas de reconocimiento de entidades (NER) son aplicadas en el trabajo de (Jung, 2011) para categorizar el texto en redes sociales.

4 Preparación de los datos

La organización del taller proporcionó dos conjuntos de datos, uno de entrenamiento y otro para el test. En total más de 70.000 *tweets* en español sobre unas 200 personas conocidas, de diversos ámbitos como política, economía, comunicación o cultura. En (Villena-Román et al., 2013) se puede encontrar una descripción detallada del corpus propuesto.

Antes de aplicar los datos al clasificador de la polaridad, y posteriormente al clasificador de categorías, se le ha aplicado al conjunto de *tweets* un proceso de limpieza, con la intención de reducir la mayor cantidad de ruido posible. Dicho proceso está formado por:

1. Eliminación de caracteres que no sean letras del alfabeto español o números.

La eliminación de los signos de exclamación se produce posteriormente a su tratamiento en aquellos experimentos en los que se ha estudiado su influencia.

2. Se ha llevado a cabo una normalización de las expresiones de risa, de manera que todas ellas se encuentren representadas por una misma expresión.
3. También se han normalizado las palabras que tienen letras repetidas. El proceso ha consistido en reducir a dos repeticiones toda letra que estuviera repetida más de tres ocasiones, de manera que se considera distinta a la palabra original, pero de la misma forma independientemente del número de repeticiones.

Una vez realizado este *preprocesado*, los datos ya están preparados para su aplicación al clasificador tanto de polaridad como de temas.

5 Tarea 1: Clasificación de la polaridad

La primera tarea consiste en el desarrollo de un sistema de clasificación de *tweets* en español en cinco niveles de polaridad: NONE, N+, N, NEU, P, P+, siendo NONE la etiqueta utilizada para aquellos *tweets* que no tienen sentimientos, no son subjetivos, y NEU para los neutros. En (Villena-Román et al., 2013) se describen con mayor detalle cada una de las categorías de opinión.

Para resolver el problema planteado se decidió seguir una estrategia basada en aprendizaje automático supervisado. Primeramente se llevó a cabo un proceso de evaluación del clasificador. El algoritmo de clasificación elegido fue SVM (*Support Vector Machines*) (Vapnik, 1995), y más concretamente la implementación *SVMLight*¹. La elección de SVM se ha fundamentado en los buenos resultados que suele ofrecer en los trabajos de AO, pudiéndose consultar muchos de ellos en (Pang and Lee, 2008). Además, SVM también ha sido utilizado con éxito por nuestro equipo en varios trabajos de AO (Martínez-Cámara et al., 2011a), (Martínez-Cámara et al., 2011b).

Una vez que se han limpiado los datos tal

¹ <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>

y como se ha descrito en la sección 4 , se diseñaron un conjunto de 33 experimentos, que van desde el caso base en el que solamente se *tokenizan* los *tweets*, hasta configuraciones a los que se añaden número de palabras positivas, negativas que aparecen en los *tweets*. Las características que se han utilizado para evaluar la configuración del clasificador han sido:

1. **Unigramas:** Cada *tweet* se tokeniza, y se utiliza la métrica TF para representar cada *unigrama*. TF se refiere a la frecuencia relativa de cada *unigrama* en el *tweet*. La elección de TF, y no por ejemplo TF-IDF, se basa en trabajos previos en el taller en el que siempre los mejores resultados se han obtenido con TF.
2. **Emoticonos:** Se añade como característica el número de emoticonos positivos o negativos que aparecen en el *tweet*. Para ello se ha utilizado una bolsa de emoticonos positivos y negativos, más concretamente los que aparecen en la Tabla 1.
3. **Palabras positivas y negativas:** En algunas de las evaluaciones que se han llevado a cabo se han incluido como características el número de palabras positivas y negativas que aparecen en los *tweets*. Esto se ha realizado siguiendo un enfoque basado en bolsa de palabras. En español no se ha encontrado ninguna bolsa de palabras positivas y negativas, por lo que se decidió traducir automáticamente la presentada en (Hu and Liu, 2004), y que se puede descargar desde la web² de uno de sus autores.
4. **Intensidad:** Aprovechando la bolsa de palabras, se decidió añadir como característica adicional el número de palabras positivas y negativas con caracteres repetidos, de manera que se pudiera modelizar la intensidad de la opinión o emoción que el autor quiere expresar.
En otro conjunto de experimentos las palabras indicadoras de opinión y que contaban con caracteres repetidos no se les da ningún tratamiento especial, mientras que otro conjunto de

evaluaciones del método al contador de palabras positivas y negativas se le añadía una unidad más en el caso de que el término tuviera letras repetidas. La intensidad puede ser expresada mediante signos de exclamación, por lo que en otro conjunto de experimentos se aumentaba con una unidad adicional el contador de palabras positivas o negativas en el caso de fueran acompañadas por un signo de exclamación.

También se ha experimentado con el efecto de la presencia de partículas negativas en el *tweet*. La presencia de una partícula negativa delante de alguna palabra indicadora de opinión, hacía que su aportación se sumara al contador opuesto a la categoría que pertenece, es decir, que si una palabra positiva va acompañada de un elemento negativo, su aportación se registra en el contador de palabras negativas, en lugar del de positivas.

Emo. Positivos	:) :) :-) ;) ;)=) ^_^ :- D :D :d =D C: Xd XD xD (x (= ^^ ^o^ 'u' n_n *_* *O* *O* *_* _
Emo. Negativos	:-(: (:(((: (D: Dx 'n' :\n /:):-/ :'= '[:_(/T_T TOT ;,;

Tabla 1 Emoticonos positivos y negativos

En las tareas de aprendizaje automático relacionadas con procesamiento de texto se suele utilizar *stopper* y *stemmer* para reducir el número de características léxicas. Para los experimentos que se han llevado a cabo para la participación en el taller, solo se ha probado la eficacia de la aplicación de *stemmer*, ya que en investigaciones previas se ha llegado a la conclusión que para la tarea de Análisis de la Opinión en Twitter, el uso de *stopper* es contraproducente.

El procedimiento de evaluación del clasificador elegido ha sido el de *K-Cross-Validation* con un valor de $K=10$.

La evaluación de la combinación de todas las características anteriores originó 33 experimentos, de los cuales se escogieron cuatro para presentarlos al taller.

Los resultados que se obtuvieron en el proceso de evaluación del clasificador se pueden ver en la Figura 1. De todas esas configuraciones del clasificador se seleccionaron los cuatro con mayor valor de

² <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>

F1. Las características de esas configuraciones son:

1. **EXP 2:** En este caso cada *tweet* se representa como un conjunto de *unigramas*, cuya importancia es indicada por su valor de TF. Antes de su clasificación se le aplica un proceso de *stemmer*.
2. **EXP 6:** Lo mismo que el anterior pero además se normalizan las direcciones web, de manera que cada url se sustituye por URL. En este caso también se normalizan las menciones, por lo que cada expresión de la forma *@nombre_usuario*, se sustituye por MENTION.
3. **EXP 4:** Igual que en el experimento *pol-sinai-2-5l*, pero en esta ocasión solo se normalizan las menciones.
4. **EXP 19:** En este experimento, además de las características léxicas incluidas en las tres anteriores configuraciones, se incluyen como características el número de emoticonos positivos y negativos, el número de palabras positivas y negativas, y en el caso de que alguna de esos términos tengan caracteres repetidos se considera dicha palabra como doble.

Los mejores resultados se han obtenido con las configuraciones del sistema más simples. Únicamente el experimento EXP 19 es el que incluye características que indican una cierta información semántica, como es el número de emoticonos y palabras positivas y negativas.

El que no haya funcionado el enfoque basado en bolsa de palabras positivas y negativas, como se pensaba en un principio, se cree que es debido a la calidad de la traducción del listado de términos original, y a las lógicas diferencias entre expresiones que en inglés pueden ser indicativas de una opinión, y en español no. Esto indica, que tiene que haber una mejora de los recursos lingüísticos en español relacionados con el AO, y mientras tanto seguir investigando en

el uso de otros recursos de mayor calidad en inglés.

A continuación se presentan los resultados obtenidos con los datos de *test*, al mismo tiempo que se asocia el nombre del experimento con el nombre asignado por los organizadores del taller. La configuración *EXP-19 (pol-sinai-4-5l)* fue la que mejores resultados obtuvo (Tabla 2).

Experimento	Precisión
P	0,94%
P+	60,99%
NEU	0,38%
N	33,61%
N+	32,00%
NONE	71,52%
Total	54,68%

Tabla 2: Resultados del experimento *pol-sinai-4-5l*

Los mayores errores se encuentran entre clases cercanas, como son NONE y NEU, y entre P y P+. El error entre las clases NONE y NEU es muy probable que sea debido a la diferencia tan grande entre el número de *tweets* NEU y NONE, por lo que seguramente el clasificador ha sobreentrenado la clase NONE. El error en las clases positivas es muy probable que esté debido al conjunto de palabras positivas utilizadas, así como a la asignación de una doble importancia a aquellas palabras positivas con letras repetidas.

En la Tabla 3 se muestran los resultados para la configuración correspondiente al experimento *EXP 6 (pol-sinai-2-5l)*. En este caso, al no haber ninguna característica semántica, el sobre-ajuste sobre la clase NONE es mucho más evidente.

La Tabla 4 contiene los resultados del experimento *EXP 2 (pol-sinai-1-5l)*. En este caso existe un mayor sobre-ajuste sobre la clase NONE. Además, al tratarse de una configuración en la que solo se tiene en cuenta las características léxicas, se puede decir que las únicas clases con un vocabulario más determinante son P y P+, ya

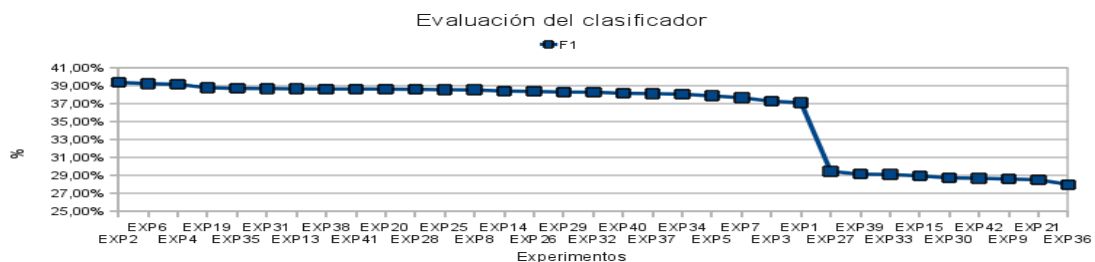


Figura 1: Valores F1 obtenidos durante el proceso de evaluación del clasificador

que son las únicas en las que unos pocos *tweets* se han clasificado correctamente.

En la Tabla 5 se muestran los resultados de la configuración del experimento *EXP 4 (pol-sinai-3-5l)*. Este experimento tiene un comportamiento similar al experimento *pol-sinai-1-5l*, pero en esta ocasión sí se consigue clasificar algún *tweet* de las clases N y N+. Esto puede ser debido a la no normalización de las direcciones web, que haya favorecido a esas clases, y haya perjudicado algo la clasificación de la clase NONE.

Experimento	Precisión
P	0,27%
P+	4,69%
NEU	0%
N	0,12%
N+	0,37%
NONE	96,48%
Total	35,65%

Tabla 3: Resultados del experimento *pol-sinai-2-5l*

Experimento	Precisión
P	0,87%
P+	0,81%
NEU	0%
N	0%
N+	0%
NONE	99,31%
Total	35,28%

Tabla 4: Resultados del experimento *pol-sinai-1-5l*

6 Tarea 2: Categorización de temas

En esta tarea se evalúa el rendimiento de un clasificador que identifique correctamente el *topic* de un *tweet*, y en función de ese *topic* analice la polaridad de dicho *tweet*. La evaluación se realiza conforme a las mismas métricas definidas en la tarea anterior (precisión, *recall* y F1).

Experimento	Precisión
P	0,47%
P+	0,58%
NEU	0%
N	0,01%
N+	0,04%
NONE	98,65%
Total	34,97%

Tabla 5: Resultados del experimento *pol-sinai-3-5l*

Para esta tarea se ha desarrollado un sistema de clasificación de temas que tomando como base las categorías de los *tweets* de entrenamiento, categorice correctamente un nuevo *tweet*. Las categorías identificadas son: cine, deportes, economía, entretenimiento, fútbol, literatura, música, otros, política y tecnología.

El sistema de clasificación utilizado para los experimentos oficiales es un sistema de aprendizaje automático basado en SVM, que toma diversas características para el entrenamiento. Como características a procesar se han utilizado los *unigramas* de los *tweets*, procesados y sin procesar. Se realizaron experimentos previos con otros sistemas de aprendizaje automático, siendo SVM el que obtuvo los mejores resultados. El procesamiento de los *tweets* es el mismo descrito en la tarea 1.

Además, se han generado dos bolsas de palabras relevantes para cada categoría para mejorar el rendimiento del sistema. La primera bolsa de palabras se ha obtenido a partir de Google AdWordsKeywordTool³, que permite introducir un término y devuelve las *n* ideas directamente relacionadas. La segunda bolsa de palabras se ha obtenido a partir de los *hashtags* de los *tweets* de entrenamiento, tomando para cada categoría los *hashtags* que únicamente aparecen en los *tweets* de dicha categoría.

En función del procesamiento o no de los *tweets* y las bolsas de palabras, se realizaron diversos experimentos con el conjunto de *tweets* de entrenamiento y evaluación cruzada (*10-fold cross validation*). Según los resultados previos obtenidos se presentaron los siguientes experimentos oficiales:

- **Top-sinai-1** (caso base). Los *tweets* no se han procesado ni se utilizan bolsas de palabras.
- **Top-sinai-2**. Los *tweets* se han procesado pero no se aplica ni *stopper* ni *stemmer*. No se utilizan bolsas de palabras.
- **Top-sinai-3**. Los *tweets* se han procesado y se aplica *stopper* y *stemmer*. No se utilizan bolsas de palabras.
- **Top-sinai-4**. Los *tweets* se han procesado y se aplica *stopper* y

³ Disponible en <https://adwords.google.com/o/KeywordTool>

stemmer. Como bolsas de palabras se utilizan los *hashtags*, y se añaden a cada *tweet* en entrenamiento, dependiendo de su *topic*.

- **Top-sinai-5.** Los *tweets* se han procesado y se aplica *stopper* y *stemmer*. Como bolsas de palabras se utilizan los *hashtags* y las palabras de Adwords, y se añaden a cada *tweet* de entrenamiento, dependiendo de su *topic*.

Los resultados obtenidos con estos experimentos se muestran en la Tabla 6.

Analizando las categorías etiquetadas automáticamente por nuestro sistema observamos que la mayoría de los *tweets* de evaluación han sido etiquetados en las categorías “otros”, “política” y “entretenimiento”. Por este motivo se han generado estadísticas de etiquetado de la colección de entrenamiento, obteniendo los datos que se pueden ver en la Tabla 7.

Run id	Precisión
top-sinai-1	32.34%
top-sinai-2	34.76%
top-sinai-3	34.06%
top-sinai-4	37.79%
top-sinai-5	39.37%

Tabla 6: Resultados oficiales de la tarea 2

Observando estos resultados no es difícil concluir que la distribución de categorías está muy desbalanceada para un sistema de aprendizaje automático al uso, sin contar con información adicional de la categoría a la hora de entrenar, ya que, por ejemplo, con un subconjunto de entrenamiento del 1,11% (literatura) resultará imposible que un *tweet* de evaluación lo clasifique en dicha categoría.

Se está realizando un análisis más profundo de los resultados con el fin de obtener más conclusiones y trabajo a realizar para mejorar el sistema, aunque es muy probable que la mejora del clasificador pase por entrenar diferentes categorías generales y específicas con otro material externo.

7 Conclusiones y trabajo futuro

La experimentación que se ha presentado pone de manifiesto que la inclusión de características semánticas al conjunto de características ayuda al proceso de la clasificación de la polaridad. Por otro lado,

también es muy importante la limpieza y normalización de los datos, como demuestran los resultados *pol-sinai-2-5l*, *pol-sinai-1-5l* y *pol-sinai-3-5l*.

Categoría	# tweets	%
cine	183	2,53%
deportes	101	1,40%
economía	525	7,27%
entretenimiento	1.209	16,75%
fútbol	225	3,12%
literatura	80	1,11%
música	411	5,69%
política	2.715	37,61%
otros	1.625	22,51%
tecnología	145	2,01%
TOTAL:	7.219	100%

Tabla 7: N° de *tweets* por categoría

En cuanto a la tarea de categorización de temas, la incorporación al proceso de información de cada categoría mejora considerablemente los resultados.

Actualmente se está trabajando en el uso de diversos recursos semánticos, de listas de palabras, de aprovechamiento de la información del contexto y del tratamiento de la negación para mejorar la clasificación de la polaridad. Para mejorar la categorización, se está apostando por mejorar el enfoque basado en bolsa de palabras combinándolo con un sistema supervisado cuyo modelo se fuera actualizando periódicamente.

Agradecimientos

Esta investigación ha sido subvencionada parcialmente por el Fondo Europeo de Desarrollo Regional (FEDER), a través del proyecto TEXT-COOL 2.0 (TIN2009-13391-C04-02) y el proyecto ATTOS (TIN2012-38536-C03-0) por el gobierno español, y por la Comisión Europea bajo el Séptimo programa Marco (FP7 - 2007-2013) a través del proyecto FIRST (FP7-287607).

Bibliografía

Asur, Sitaram, Huberman, Bernardo A. (2010). Predicting the Future with Social Media. 2010 IEEE/WIC/ACM International Conf. on Web Intelligence and Intelligent Agent Technology, 1, pp.492-499.

- Bollen, J. Mao, H., Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 21, pp. 1-8.
- Diakopoulos, N. A. and D. A. Shamma. (2010). Characterizing debate performance via aggregated twitter sentiment. *CHI '10: Proc. of the 28th International Conf. on Human Factors in Computing Systems*. New York, NY, USA. ACM. pp 1195–1198.
- García, S. O'Mahony, M.P., Smyth, B. Towards tagging and categorization for micro-blog. *21st National Conf. on Artificial Intelligence and Cognitive Science (AICS 2010)*, Galway, Ireland.
- Go, A., R. Bhayani, and L. Huang. (2009). Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project
- Hu, Minqing, Liu, Bing. (2004). Mining and Summarizing Customer Reviews. *Proc. of the ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD-2004)*. Seattle, Washington, USA.
- Jansen, B., M. Zhang, K. Sobel, and A. Chowdury (2009). Twitter power: tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*.
- Jung, J.J. Towards Named Entity Recognition Method for Microtexts in Online Social Networks: A Case Study of Twitter, *International Conf. on Advances in Social Networks Analysis and Mining*, pp.563-564
- Martínez-Cámara E., Martín-Valdivia M. T., Perea-Ortega, J. M., Ureña-López, L. A. (2011b). Técnicas de clasificación de opiniones aplicadas a un corpus en español. *Procesamiento de Lenguaje Natural*. 47, pp. 163-170.
- Martínez-Cámara E., Martín-Valdivia M. T., Ureña-López, L. A. (2011a). Opinion classification techniques applied to a Spanish corpus. *Natural Language Processing and Information Systems*. Springer, pp 169-176.
- O'Connor, B., R. Balasubramanyan, B. R. Routledge, and N. A. Smith (2010). From Tweets to polls: Linking text sentiment to public opinion time series. *International AAAI Conf. on Weblogs and Social Media*, Washington, D.C.
- Pak, A., P. Paroubek (2010). Twitter as a corpus for sentiment analysis and opinion mining. *Proc. of the Seventh Conf. on International Language Resources and Evaluation (LREC'10)*, (ELRA), Valletta, Malta, pp. 19–21.
- Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. *Foundation and Trends in Inf. Retrieval* 2(1-2) 1-135
- Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *Proc. of the Conf. on Empirical Methods in Natural Language Processing (EMNLP)*. ACL. pp. 79–86.
- Petrovic, S., Osborne, M., Lavrenko, V. (2010). The Edinburgh Twitter corpus. *SocialMedia Workshop: Computational Linguistics in a World of Social Media*, pp. 25–26.
- Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. *Proc. of the ACL Student Research Workshop*, pp. 43–48.
- Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H. and Demirbas, M. 2010. Short text classification in Twitter to improve information filtering. In *Proc. of the 33rd International ACM SIGIR Conf. on Research and development in Inf. Retrieval (SIGIR '10)*. ACM
- Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *Proc. of the 40th Annual Meeting on ACL*. Morristown, NJ, USA. pp. 417–424.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Villena-Román, J. Lana-Serrano, S. González-Cristóbal, J.C. Martínez-Cámara, E. (2013). TASS – Workshop on Sentiment Analysis at SEPLN. *Procesamiento de Lenguaje Natural*, 50.