

Social Rankings: análisis visual de sentimientos en redes sociales

Social Rankings: Visual Sentiment Analysis in Social Networks

Javi Fernández, Yoan Gutiérrez, José M. Gómez, Patricio Martínez-Barco

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante
Carretera San Vicente del Raspeig S/N, 03690 San Vicente del Raspeig
{javifm,ygutierrez,jmgomez,patricio}@dlsi.ua.es

Resumen: *Social Rankings* es una aplicación web que realiza un seguimiento en tiempo real de entidades en las redes sociales. Detecta y analiza las opiniones sobre estas entidades utilizando técnicas de análisis de sentimientos para generar un informe visual de su valoración y su evolución en el tiempo.

Palabras clave: análisis de sentimientos, minería de opiniones, redes sociales

Abstract: *Social Rankings* is a web application that follows different entities in the social networks in real time. It detects and analyses the opinions about these entities using sentiment analysis techniques, to generate a visual report of their reputation and evolution in time.

Keywords: sentiment analysis, opinion mining, social networks

1 Introducción

En la actualidad millones de personas expresan cada día sus opiniones públicamente a través de las redes sociales. Empresas y organizaciones analizan manualmente esta información subjetiva para realizar estudios y obtener una visión global de la valoración de diferentes marcas, productos o personajes públicos. Pero la cantidad de información disponible es tan grande que es necesario crear herramientas que realicen esta tarea de manera automática y en tiempo real. Este es el objetivo de la aplicación web *Social Rankings*.

Social Rankings es una aplicación web que realiza un seguimiento en tiempo real de entidades en las redes sociales. Detecta y analiza las opiniones sobre estas entidades utilizando técnicas de análisis de sentimientos para generar un informe visual de su valoración y su evolución en el tiempo.

2 Configuración

El único paso a realizar por el usuario para poder utilizar *Social Rankings* es elegir las entidades a seguir. Pueden ser tanto personas como empresas, organizaciones, temas o, en general, cualquier conjunto de palabras de interés. Para iniciar su seguimiento, estas deben ser definidas por el usuario, indicando su nombre, descripción y palabras clave. A partir de ese momento, las redes sociales son rastreadas periódicamente, obteniendo todas

las publicaciones que mencionen a estas entidades, es decir, que contengan sus palabras clave.

3 Funcionamiento

Una vez elegidas las entidades y sus palabras clave, comienza a realizarse el seguimiento en las redes sociales. En este momento realizamos el seguimiento en Twitter¹, por la facilidad de uso de sus APIs² y la disponibilidad de herramientas para su manipulación³. En el futuro planeamos añadir nuevas redes sociales, además de blogs y foros.

A medida que se van encontrando diferentes publicaciones mencionando a las entidades especificadas, un sistema de análisis de sentimientos detecta aquellas que contienen una opinión y las clasifica como positivas o negativas. En la actualidad utilizamos una combinación dos sistemas diferentes.

El primero está basado en una de las aproximaciones de (Balahur, 2011). Este sistema utiliza unos diccionarios de palabras que contienen sentimientos, obtenidos combinando otros recursos de opinión como *WordNet Affect* (Strapparava y Valitutti, 2004) o *SentiWordNet* (Esuli y Sebastiani, 2006). La polaridad de un texto dependerá de la suma total de los pesos de las palabras de opinión que

¹<http://twitter.com>

²<https://dev.twitter.com/streaming/overview>

³<http://twitter4j.org/>

contiene ese texto.

El segundo está basado en la aproximación híbrida de Fernández et. al (2013) y (2014). En esta aproximación se utilizan *skipgrams* como unidades de información (Guthrie et al., 2006). A partir de un corpus se genera un diccionario de skipgrams con pesos de positividad y negatividad asociados. Utilizando el corpus y este diccionario, se entrena un modelo de aprendizaje automático que aprende a combinar los pesos de los skipgrams para obtener la polaridad de nuevos textos. En este momento el corpus que hemos utilizado es el proporcionado en la edición de 2013 del workshop TASS⁴ (Villena Román et al., 2013).

Estas aproximaciones han sido elegidas por sus buenos resultados en competiciones (Fernández et al., 2013) y su velocidad de procesamiento, apta para aplicaciones en tiempo real.

4 Valoración

Con el fin de generar un informe visual en el tiempo para cada entidad, necesitamos obtener una puntuación numérica que tenga en cuenta la cantidad de opiniones positivas y negativas detectadas, y la cantidad de gente a la que han llegado estas opiniones, en un periodo de tiempo concreto. A esta puntuación le hemos llamado *valoración* y se ha calculado utilizando la ecuación 1.

$$v_{e,t} = \frac{\sum_{p \in P_{e,t,+}} 2a_p + \sum_{p \in P_{e,t,0}} a_p - \sum_{p \in P_{e,t,-}} 3a_p}{d_t + \sum_{p \in P_{e,t}} 3a_p} \quad (1)$$

Donde $v_{e,t}$ es la valoración de la entidad e en el período de tiempo t ; $P_{e,t}$ es el conjunto de todas las publicaciones que contienen una mención a la entidad e en el período de tiempo t (y p es una publicación dentro de ese conjunto); $P_{e,t,+}$, $P_{e,t,0}$ y $P_{e,t,-}$ son subconjuntos de $P_{e,t}$ cuyas publicaciones han sido clasificadas como positivas, neutrales o negativas respectivamente; a_p es la audiencia o número de usuarios a los que ha llegado la publicación p ; y d_t es la duración en milisegundos del periodo t . Utilizando esta ecuación obtenemos un valor dentro del intervalo $[-1, +1]$, donde -1 sería la peor valoración y $+1$ sería la mejor valoración dada por el sistema.

Consideramos una mención neutral ($p \in P_{e,t,0}$) como algo positivo, ya que el hecho de

que una entidad sea mencionada en las redes sociales aumenta su valoración. Por eso las menciones neutrales influyen de manera positiva en la fórmula. Las menciones positivas ($p \in P_{e,t,+}$) indican que no sólo se está mencionando a la entidad sino que se está diciendo algo bueno sobre ella. Por eso consideramos que deben influir en mayor medida que las menciones neutrales, aumentando su valor en un factor de 2. Finalmente, a las menciones negativas ($p \in P_{e,t,-}$) les aumentamos su valor en un factor de 3, ya que para los usuarios suele ser más interesante conocer los malos comentarios sobre una entidad (por ejemplo, para encontrar una solución). Estos parámetros han sido establecidos manualmente según lo descrito anteriormente.

La suma de la duración del período en el denominador de la ecuación (d_t) es una forma de dar una mayor valoración a las publicaciones que hayan llegado a más seguidores. Por ejemplo, si una entidad tiene menciones negativas que ha llegado a 100 personas en un minuto, su valoración sería de $-3 \cdot 100 / (3 \cdot 100 + 60000) = -0,005$. Sin embargo, si las publicaciones ha llegado a 10.000 personas, la valoración sería de $-3 \cdot 10000 / (3 \cdot 10000 + 60000) = -0,333$.

5 Informe visual

En la Figura 1 podemos ver un ejemplo de la interfaz visual de *Social Rankings*. Las entidades que comparamos en este ejemplo son los cuatro principales partidos políticos en España en este momento. Concretamente, se muestra el valor de reputación entre el 17 y el 24 de marzo de 2015.

Esta interfaz podemos dividirla en cuatro partes, que describiremos a continuación:

- *Selector de entidades*. Es un campo desplegable que contiene todas las entidades a las que se les está realizando el seguimiento. Se deben seleccionar una o varias entidades para obtener el informe visual. En la Figura 2 podemos ver el proceso de selección de entidades en funcionamiento.
- *Selector de vista*. Es otro campo desplegable mediante el cual se puede cambiar la vista actual. Las vistas disponibles son *número de tweets*, *número de tweets positivos*, *número de tweets negativos* y *valor de reputación*. En las primeras tres vistas simplemente se realiza un conteo

⁴<http://www.daedalus.es/TASS2013>

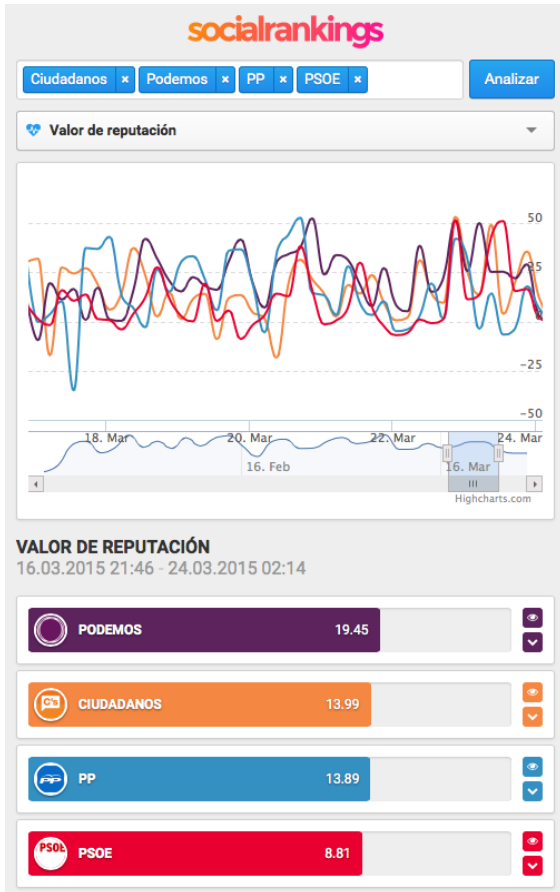


Figura 1: Interfaz visual de Social Rankings

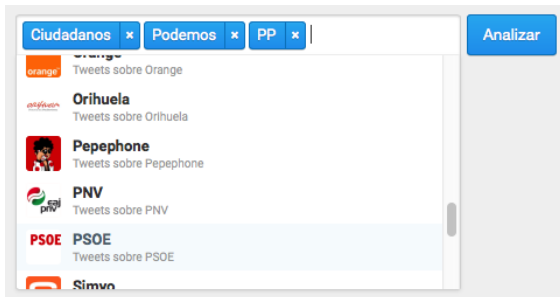


Figura 2: Selector de entidades

del número de publicaciones total, publicaciones positivas y publicaciones negativas respectivamente. La última vista muestra la valoración descrita en la sección 4. En la Figura 3 se puede observar el desplegable con las vistas disponibles.

- *Ranking de entidades.* En esta sección se muestra una lista de las entidades elegidas con un valor asociado para un rango de tiempo dado, que depende de la vista elegida. En el ejemplo de la Figura 4 la vista elegida es el número de tweets positivos. Para que sea más sencillo compa-

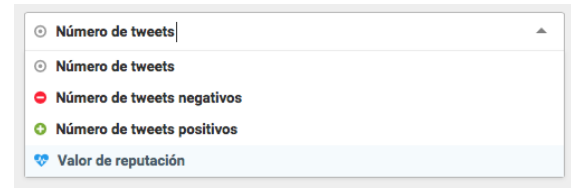


Figura 3: Selector de vista

rar los valores para las distintas entidades, se muestra una barra con el color de la entidad cuyo tamaño depende del valor asignado. Para elegir seleccionar un rango de fechas diferente, se utiliza la gráfica de evolución, que explicamos a continuación.

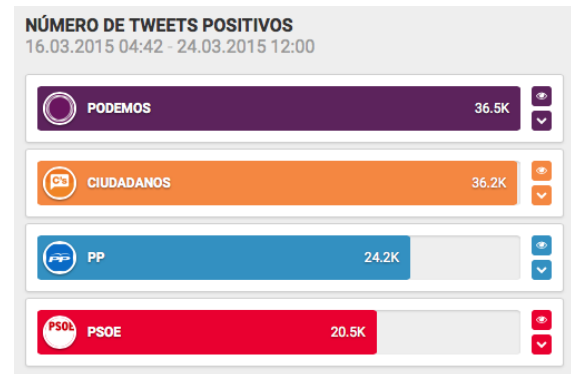


Figura 4: Ranking de entidades

- *Gráfica de evolución.* Es una gráfica en la que se muestra la evolución de los valores de la vista elegida para las entidades seleccionadas durante todo el periodo de seguimiento. Por defecto sólo se muestra el último día, pero es posible seleccionar un rango de fechas diferente a lo largo de toda la línea de evolución. Al cambiar el rango de fechas, el ranking de entidades explicado anteriormente también actualizará sus valores. En la Figura 5 podemos ver un ejemplo de evolución del valor de reputación de las entidades elegidas. Cada punto representa un periodo de tiempo concreto y, al pasar el ratón por encima, es posible ver los datos para ese periodo.

También se ha añadido una funcionalidad adicional experimental cuando se visualiza una única entidad. Podremos ver los términos y expresiones más utilizados en cada período de tiempo. De esta forma es posible obtener una visión general de lo que ha ocurrido y lo que más se ha comentado en un momento

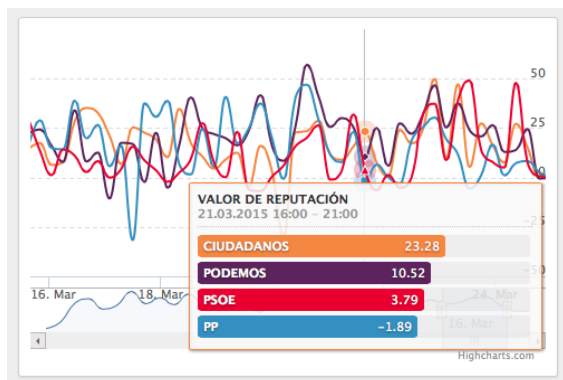


Figura 5: Gráfica de evolución

concreto. Esto se ha realizado mediante técnicas de clustering de texto utilizando la herramienta Carrot2⁵. En la Figura 6 se muestra un ejemplo.



Figura 6: Términos y expresiones más relevantes

Esta herramienta se puede utilizar públicamente⁶. En esta versión pública sólo se visualizan los datos de las entidades predefinidas, no es posible añadir nuevas entidades. Si desea añadir alguna entidad contacte con los autores de este artículo.

6 Agradecimientos

Social Rankings ha sido desarrollada por el Grupo de Procesamiento del Lenguaje Natural y Sistemas de Información (GPLSI)⁷ de la Universidad de Alicante⁸. Esta aplicación ha sido financiada parcialmente por el Gobierno Español a través de los proyectos *ATTOS* (TIN2012-38536-C03-03) y *LEGOLANG* (TIN2012-31224), la Comisión Europea a través del proyecto *SAM*

⁵<http://project.carrot2.org>

⁶<http://gplsi.dlsi.ua.es/demos/socialrankings>

⁷<http://gplsi.dlsi.ua.es/>

⁸<http://www.ua.es/>

(FP7-611312), la Generalitat Valenciana a través del proyecto *DIIM2.0* (PROMETEOII/2014/001) y la Universidad de Alicante a través del proyecto emergente “*Explotación y tratamiento de la información disponible en Internet para la anotación y generación de textos adaptados al usuario*” (GRE13-15).

Bibliografía

- Balahur, A. 2011. *Methods and resources for sentiment analysis in multilingual documents of different text types*. Universidad de Alicante.
- Esuli, A. y F. Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. En *Proceedings of LREC*, volumen 6, páginas 417–422. Cite-seer.
- Fernández, J., Y. Gutiérrez, J. M. Gómez, y P. Martínez-Barco. 2014. A Supervised Approach for Sentiment Analysis using Skipgrams. En *Workshop on Natural Language Processing in the 5th Information Systems Research Working Days (JISIC 2014)*.
- Fernández, J., Y. Gutiérrez, J. M. Gómez, P. Martínez-Barco, A. Montoyo, y R. Muñoz. 2013. Sentiment Analysis of Spanish Tweets Using a Ranking Algorithm and Skipgrams. En *XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural (SEPLN 2013)*, páginas 133–142.
- Guthrie, D., B. Allison, W. Liu, L. Guthrie, y Y. Wilks. 2006. A closer look at skipgram modelling. En *Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)*, páginas 1–4.
- Strapparava, C. y A. et al. Valitutti. 2004. WordNet Affect: an Affective Extension of WordNet. En *LREC*, volumen 4, páginas 1083–1086.
- Villena Román, J., S. Lana Serrano, E. Martínez Cámara, y J. C. González Cristóbal. 2013. TASS-Workshop on Sentiment Analysis at SEPLN.