

Desafíos del Análisis de Sentimientos

Salud M^a Jiménez, Eugenio Martínez, M. Teresa Martín, L. Alfonso Ureña

SINAI - Sistemas Inteligentes de Acceso a la Información
Departamento de Informática
Universidad de Jaén
{sjzafra, emcamara, maite, laurena}@ujaen.es

1 Introducción

La minería de opiniones (MO), también conocida como análisis de sentimientos (AS), es una disciplina que se centra en detectar la información subjetiva de un texto y clasificarla. Existen muchos trabajos centrados en la MO, pero la mayor parte de las investigaciones se han realizado sobre opiniones escritas en inglés. Sin embargo, cada vez es mayor la presencia de otros idiomas en Internet, entre los que se encuentra el español, lo que pone de manifiesto la necesidad de su tratamiento. El grupo SINAI de la Universidad de Jaén lleva unos años trabajando en el análisis de opiniones en español [1], [2], [3] y actualmente continúa con esta temática, debido a su gran importancia.

Por otra parte, en las revisiones del estado del arte del AS de Bing Liu [4] y de Pang y Lee [5] se muestran como desafíos el tratamiento de la negación, de la ironía y del sarcasmo, la adaptación al dominio, el análisis a nivel de aspecto, la detección de opiniones spam, etc. En este momento nos encontramos trabajando en dos de los frentes abiertos en el análisis de opiniones, como son el tratamiento de la negación y el análisis a nivel de aspecto o característica.

2 Tratamiento de la negación en español

La negación es un elemento fundamental en el análisis de opiniones, que requiere un tratamiento especial, ya que una opinión negativa puede ser expresada con términos positivos negados o, por el contrario, una opinión positiva puede expresarse a partir de la negación de términos negativos. Las oraciones “La película no me gustó” y “El personaje principal no era una mala persona” son claros ejemplos de los dos casos mencionados anteriormente.

Debido a la importancia de este fenómeno decidimos realizar una primera aproximación al estudio de la negación en opiniones escritas en español. Este trabajo se ha enviado a revisión para el número 53 de la revista Procesamiento del Lenguaje Natural.

En el trabajo mencionado se propone un sistema no supervisado para la clasificación de opiniones teniendo en cuenta la influencia de la negación. La negación es una característica particular de cada idioma que debe ajustarse a las singularidades de la lengua en estudio. En nuestro caso se han estudiado

las partículas negativas más importantes según la RAE [6]: “no”, “tampoco”, “nadie”, “jamás”, “ni”, “sin”, “nada”, “nunca” y “ninguno”. Para resolver este fenómeno es necesario, en primer lugar, identificar el ámbito de la negación y, posteriormente, modificar la polaridad del fragmento de la oración que se ve afectado por ella.

Para determinar el ámbito de la negación se han definido una serie de reglas que permiten generalizar el tratamiento de las distintas partículas negativas. Para construir estas reglas se ha llevado a cabo un estudio de los árboles de dependencias de diferentes oraciones en las que está presente alguna de las partículas abordadas. Para ello, se ha utilizado el analizador de dependencias de Freeling [7], que permite generar el árbol de dependencias de una oración en base a su estructura sintáctica. En la Tabla 1 se muestran las reglas obtenidas tras realizar el estudio.

Partícula	Regla
no, tampoco, nadie, jamás, ninguno	Afecta al nodo padre y al árbol formado por el hermano de la derecha (incluido).
ni, sin	Afecta a todos los hijos y a todos los árboles formados por ellos hasta llegar a nodos hoja.
nada, nunca	Afecta al nodo padre.

Tabla 1. Reglas ámbito de la negación.

Estas reglas permiten marcar las palabras de la oración afectadas por alguna de las partículas de estudio, de manera que se tenga en cuenta esta información a la hora de calcular la polaridad de la opinión. Nuestra propuesta consiste en invertir la polaridad de las palabras marcadas, es decir, de las palabras pertenecientes al ámbito de la negación. Por ejemplo, en la oración “El personaje principal no era una mala persona”, la partícula negativa “no”, según las reglas definidas, afecta a las palabras “era”, “mala” y “persona” (Figura 1), las cuales llevarán la marca de la negación, pero sólo “mala” expresa opinión (-1, negativa) por lo que su polaridad se verá invertida $(-1 * -1) = (1, \text{positiva})$.

Para demostrar que la capacidad de predicción de la polaridad de un sistema de clasificación de opiniones mejora al incluir un módulo que se encargue del tratamiento de la negación, en este trabajo, se han realizado experimentaciones sobre un corpus en español formado por 3.878 críticas de cine [8] recogidas de la web MuchoCine¹, utilizando diferentes recursos lingüísticos para el cálculo de la polaridad. Las experimentaciones se han llevado a cabo tanto sin el módulo de identificación de la negación como con él, llegando a la conclusión de que se produce una mejora en la clasificación.

No obstante, este trabajo ha sido una primera aproximación al desafío del tratamiento de la negación, en el cual seguimos trabajando para abordar el resto de partículas negativas utilizadas en español.

¹ <http://www.muchochine.net>

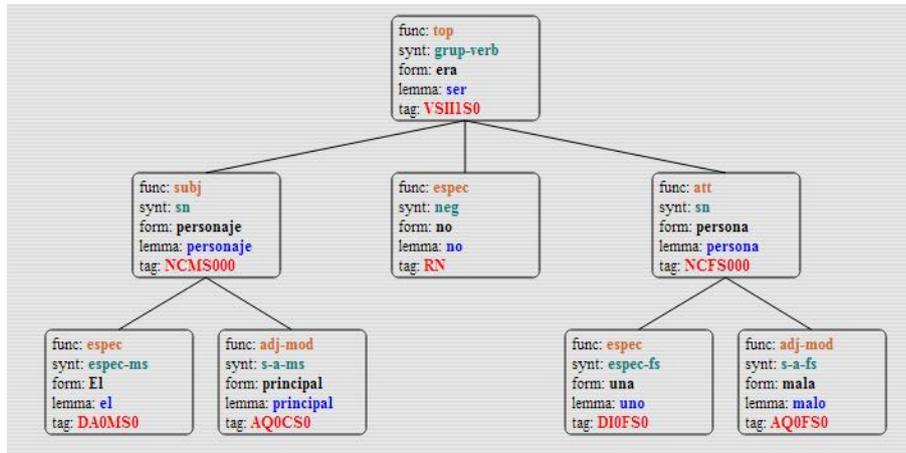


Figura 1. Árbol de dependencias en el que se analiza la partícula “no” en la oración “El personaje principal no era una mala persona”.

3 Análisis a nivel de aspecto en inglés

En el AS se distinguen tres niveles de estudio de un texto: nivel de documento, de frase o de aspecto. El análisis a nivel de documento determina el sentimiento general expresado en una opinión, mientras que el análisis a nivel de frase específica, para cada una de las oraciones de un texto, si expresan una opinión positiva, negativa o neutra [4]. Sin embargo, estos dos tipos de análisis no llegan al nivel de detalle que desea un usuario cuando busca información sobre algún producto, ya que el hecho de que la opinión general de un producto sea positiva no significa que el autor tenga una opinión positiva de todos los aspectos de dicho producto, ni el hecho de que sea negativa implica que todo lo relacionado con el producto sea malo.

La gran cantidad de fuentes y el elevado volumen de textos con opiniones hacen que resulte complicado para el usuario seleccionar información de su interés. Por ello, es necesario desarrollar sistemas de clasificación de opiniones a nivel de aspecto, que ayuden a los usuarios a tomar decisiones y que, por otro lado, muestren a las empresas la opinión que los consumidores tienen acerca de sus productos, para ayudarles a decidir qué deben mantener, qué deben eliminar o qué deben mejorar.

Actualmente nos encontramos trabajando en este tema sobre opiniones escritas en inglés. Nuestra primera aproximación se ha realizado para la participación en la tarea 4 (Aspect Based Sentiment Analysis) del congreso SemEval 2014. Esta tarea se centra en la extracción de los aspectos relacionados con una entidad (ej. ordenadores, restaurantes) y en el cálculo de la polaridad expresada sobre dichos aspectos en la opinión.

Nuestra propuesta para identificar los aspectos relacionados con el dominio en cuestión (ordenadores, restaurantes) se basa en la utilización de una bolsa

de palabras, construida a partir de los datos de entrenamiento proporcionados y de los datos extraídos de forma automática de la base de conocimiento colaborativa Freebase². Una vez extraídos los aspectos relacionados con la entidad de estudio, el siguiente paso es determinar qué palabras modifican a cada aspecto. Para ello se ha realizado un análisis de dependencias utilizando el analizador de Stanford [9], considerando las principales formas de expresar opinión acerca de un aspecto: utilizando un verbo (“nsubj” o “nsubjpass”), empleando un adjetivo (“amod”) o por medio de una relación de dependencia con otra palabra (“dep”). Tras identificar los modificadores de los aspectos se ha calculado su polaridad mediante un sistema de voto formado por tres clasificadores basados en listas de palabras: Bing Liu [10], SentiWordNet [11] y MPQA [12].

Tras esta primera aproximación seguimos trabajando en el análisis a nivel de aspecto en inglés, considerando otros posibles modificadores. En un futuro cercano trataremos de extrapolarlo a textos en español.

Referencias

1. Martín-Valdivia, M. T., Martínez-Cámara, E., Perea-Ortega, J. M., Ureña-López, L. A. (2013). Sentiment polarity detection in spanish reviews combining supervised and unsupervised approaches. *Expert Systems with Applications*, 40(10), 3934-3942.
2. Martínez Cámara, E., Martín Valdivia, M. T., Perea Ortega, J. M., Ureña López, L. A. (2011). Técnicas de clasificación de opiniones aplicadas a un corpus en español.
3. Molina-González, M. D., Martínez-Cámara, E., Martín-Valdivia, M. T., Perea-Ortega, J. M. (2013). Semantic orientation for polarity classification in Spanish reviews. *Expert Systems with Applications*, 40(18), 7250-7257.
4. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
5. Pang, B., Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
6. Española, Real Academia. 2009. Nueva gramática de la lengua española, volumen 1. Espasa Calpe Madrid, Spain.
7. Padró, L. (2012). Analizadores multilingües en freeling. *Linguamática*, 3(2), 13-20.
8. Cruz, F. L., Troyano, J. A., Enriquez, F., Ortega, J. (2008). Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento de Lenguaje Natural*, 41.
9. De Marneffe, M. C., Manning, C. D. (2008). Stanford typed dependencies manual. URL http://nlp.stanford.edu/software/dependencies_manual.pdf.
10. Hu, M., Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.
11. Baccianella, S., Esuli, A., Sebastiani, F. (2010, May). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In *LREC* (Vol. 10, pp. 2200-2204).
12. Wilson, T., Wiebe, J., Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing* (pp. 347-354). Association for Computational Linguistics.

² <http://www.freebase.com/>