

Nuevos retos en el Análisis de Sentimientos

Salud M. Jiménez-Zafra, M. Teresa Martín-Valdivia, L. Alfonso Ureña-López

*Programa de Doctorado "Tecnologías de la Información y la Comunicación".
Departamento de Informática. Escuela Politécnica Superior de Jaén.
Universidad de Jaén, Campus Las Lagunillas, s/n, 23071, Jaén, España.
 {sjzafra, maite, laurena}@ujaen.es*

Introducción

La aparición de la web en 1990 supuso una revolución social y desde ese momento su evolución no ha cesado. En sus comienzos el usuario era prácticamente un sujeto pasivo que visualizaba la información o la publicaba, sin demasiadas posibilidades de que se generara una interacción. Con la llegada de la Web 2.0 estas limitaciones han desaparecido, ya que ofrece a los usuarios la posibilidad de generar y compartir contenidos, opiniones, experiencias, etc.

Hasta hace unos años lo más común era preguntar a familiares y amigos para conocer su opinión sobre un tema determinado, pero tras la aparición de la Web 2.0 se ha incrementado el número de usuarios que consultan Internet para este fin. 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 automáticos de extracción, clasificación y presentación de opiniones. La disciplina conocida como Minería de Opiniones (MO) o Análisis de Sentimientos (AS) surge para dar solución a este problema.

El AS es la disciplina que combina técnicas de Procesamiento del Lenguaje Natural y de la Lingüística Computacional para detectar la información subjetiva de un texto y clasificarla. Existen muchos trabajos centrados en la MO, la mayoría de ellos en inglés, pero son muchos los frentes que aún siguen abiertos y que se consideran un desafío, como el tratamiento de la negación, el análisis a nivel de aspecto, el tratamiento de la ironía y del sarcasmo, etc. (Liu, 2012) (Pang and Lee, 2008). En este momento nos encontramos trabajando en dos de estos frentes abiertos, el tratamiento de la negación y el análisis a nivel de característica.

Tratamiento de la negación

La negación es un elemento fundamental en el AS que requiere un tratamiento especial, ya que una opinión negativa puede ser expresada a partir de la negación de términos positivos o, por el contrario, una opinión positiva puede expresarse con términos negativos negados (Ej. "El ordenador no funciona bien". "El personaje principal no era una mala persona".).

La mayoría de las investigaciones en este dominio se han realizado sobre opiniones escritas en inglés (Polanyi and Zaenen, 2004) (Wilson et al., 2005) (Jia et al., 2009) (Wiegand et al., 2010) (Taboada et al., 2011) (Díaz, 2014). 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. Debido a la importancia de este fenómeno y a la escasez de investigaciones en español, estamos trabajando en la clasificación de opiniones en español teniendo

en cuenta la influencia de la negación. Nuestra propuesta pretende seguir un enfoque sintáctico, ya que la negación es una característica particular de cada idioma que debe ajustarse a las singularidades de la lengua en estudio.

Para resolver este fenómeno es necesario detectar la presencia de negación en una opinión, identificar su ámbito y modificar la polaridad del fragmento de la oración que se ve afectado por ella. En nuestra primera aproximación, para determinar el ámbito de la negación, hemos estudiado las partículas negativas más importantes según la RAE: “no”, “tampoco”, “nadie”, “ni”, “sin”, “nada”, “nunca” y “ninguno” y hemos definido una serie de reglas que permiten generalizar su tratamiento. Para construir estas reglas se ha llevado a cabo un estudio de los árboles de dependencias de diferentes oraciones en las que se hace uso de alguna de las partículas negativas abordadas. En la Tabla 1 se muestran las reglas inferidas tras realizar el estudio.

Partícula negativa	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 para determinar el ámbito de la negación.

Estas reglas permiten marcar las palabras que se encuentran en el ámbito de las principales partículas negativas utilizadas en español, de manera que se tenga en cuenta esta información a la hora de calcular el grado de subjetividad de una opinión.

En las experimentaciones realizadas se ha comprobado que la inclusión de un módulo para el tratamiento de la negación en un sistema de clasificación de opiniones mejora la predicción del grado de subjetividad de las opiniones. Por ello, seguimos trabajando en este desafío para abordar el resto de partículas negativas utilizadas en español y para mejorar nuestro enfoque.

Análisis a nivel de característica o aspecto

En el AS se pueden distinguir tres niveles de estudio de un texto: documento (Pang et al., 2002), oración (Wilson et al., 2005) y aspecto (Hu and Liu, 2004). Los análisis a nivel de documento y de oración determina de forma general el grado de subjetividad de la opinión expresada sobre un tema, producto, monumento, persona... Sin embargo, el hecho de que la opinión general de una persona sobre un producto sea positiva no quiere decir que esa persona tenga una opinión positiva sobre cada uno de los aspectos de ese producto, ni el hecho de que sea negativa implica que todo lo relacionado con este producto le parezca negativo. El análisis a nivel de aspecto permite solucionar las limitaciones de los análisis a nivel de documento y oración, ya que lleva a cabo un análisis con mayor detalle que ayuda a conocer los aspectos positivos y negativos del objeto de estudio (una marca, un lugar, un producto...).

Cada día son más las compañías interesadas en conocer la opinión que los usuarios tienen acerca de sus productos y cada día son más los usuarios que consultan Internet para conocer la opinión que otros usuarios tienen acerca de un lugar,

monumento, producto... Por este motivo, las investigaciones en el área del análisis a nivel de aspecto se han incrementado en los últimos años.

El análisis a nivel de aspecto se centra en la extracción de los aspectos relacionados con la entidad de estudio y en determinar si la opinión expresada sobre ellos es objetiva, positiva o negativa. Nuestra propuesta para abordar este fenómeno se basa en el uso de la base de conocimiento colaborativa Freebase para identificar los aspectos relacionados con la entidad de estudio, en la utilización del analizador de dependencias de Stanford (De Marneffe and Manning, 2008) para determinar qué palabras se han utilizado para hablar de esos aspectos (para lo cual hemos considerado 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") y en la definición de un sistema de voto formado por tres clasificadores no supervisados para determinar si la opinión expresada sobre dichos aspectos es objetiva o subjetiva (positiva o negativa).

Este es otro de los desafíos del AS en el que nos encontramos trabajando. Nuestra primera aproximación la hemos realizado sobre opiniones escritas en inglés, pero pretendemos extrapolar nuestras ideas a opiniones escritas en español ya que la investigación en nuestra lengua origen resulta escasa.

Bibliografía

- [1] De Marneffe, M. C., & Manning, C. D. (2008). Stanford typed dependencies manual. URL http://nlp.stanford.edu/software/dependencies_manual.pdf
- [2] Díaz, N. P. C. (2014). Detección de la Negación y la Especulación en Textos Médicos y de Opinión. http://www.sepln.org/wp-content/uploads/2014/10/Tesis_final1.pdf
- [3] Jia, L., Yu, C., & Meng, W. (2009, November). *The effect of negation on sentiment analysis and retrieval effectiveness*. In *Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 1827-1830). ACM.
- [4] 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.
- [5] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- [6] Pang, B., Lee, L., & Vaithyanathan, S. (2002, July). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*.
- [7] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135.
- [8] Polanyi, L., & Zaenen, A. (2004). Context Valence Shifters. In *Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text*.
- [9] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- [10] Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010, July). A survey on the role of negation in sentiment analysis. In *Proceedings of the workshop on negation and speculation in natural language processing* (pp. 60-68).
- [11] 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).