

Minería de Opiniones basada en características guiada por Ontologías*

Ontology-guided approach to Feature-Based Opinion Mining

Isidro Peñalver-Martínez
Villa Virtual S.L.
C/Molina de Segura 5, Bloque
3, 1ºB, 30007 Murcia, España
ipmartinez@villavirtual.com

Rafael Valencia-García
Facultad de Informática
Campus de Espinardo
30100 Murcia
valencia@um.es

Francisco García-Sánchez
Escuela Técnica Superior
de Ingeniería
Universidad de Valencia
46100 Burjassot, Valencia
francis6@uv.es

Resumen: El éxito de la Web Social ha tenido un gran impacto en la sociedad actual y en distintas áreas de investigación. En este trabajo se propone un nuevo método para la minería de opiniones que emplea técnicas tradicionales de procesamiento de lenguaje natural junto con procesos de análisis sentimental y tecnologías de la Web Semántica. Los principales objetivos de la metodología propuesta son mejorar la minería de opiniones basada en características empleando ontologías en la selección de las mismas, así como proporcionar un nuevo método para el análisis sentimental basado en análisis vectorial.

Palabras clave: Minería de opiniones, Ontologías, Análisis Sentimental, Extracción de características, identificación de la polaridad.

Abstract: The boom of the Social Web has had a big impact on a number of different research topics. In particular, the possibility to extract various kinds of added-value, informational elements from users' opinions has attracted researchers from the information retrieval and computational linguistics fields. However, current approaches to so-called opinion mining suffer from a series of drawbacks. In this paper we propose an innovative methodology for opinion mining that brings together traditional natural language processing techniques with sentimental analysis processes and Semantic Web technologies. The main goals of this methodology is to improve feature-based opinion mining by employing ontologies in the selection of features and to provide a new method for sentimental analysis based on vector analysis.

Keywords: Opinion mining, Ontology, Sentimental Analysis, Feature extraction, Polarity identification.

1 Introducción

Con la llegada de la Web Social, el número de opiniones online en las que se expresan pensamientos acerca de una gran variedad de temas está en constante crecimiento. Los contenidos producidos por la participación de los usuarios en sitios Web tales como Amazon (<http://www.amazon.com>), o IMDB (<http://www.imdb.com>) son útiles tanto para los proveedores de información como para los

consumidores de la misma. La minería de opiniones es una subdisciplina derivada de la recuperación de información y la lingüística computacional, que no trata de detectar el tema abordado en un texto, sino las opiniones y sentimientos expresados en el mismo (Esuli y Sebastiani, 2006). Estas opiniones pueden referirse a un producto dado, comentarios sobre un libro, etc., que aparecen en foros online, blogs o comentarios. Sin embargo, para llevar a cabo el proceso de minería de opiniones de

* Este trabajo ha sido financiado por el Ministerio de Ciencia e Innovación a través del proyecto SeCLOUD (TIN2010-18650)

manera automática, las técnicas existentes que implican contrastar la similitud entre un texto y una lista de palabras semilla no son suficientes.

Las Ontologías pueden utilizarse para estructurar la información. La semántica formal que subyace en los lenguajes ontológicos permite el procesamiento automático de la información representada en forma de ontologías y habilita el uso de razonadores semánticos para inferir nuevo conocimiento. En este trabajo, una ontología se define como “una especificación formal y explícita de una conceptualización compartida” (Studer et al, 1998). Las ontologías proporcionan una representación del conocimiento estructurado y formal, con la ventaja de ser reutilizable y compartible. También proporcionan un vocabulario común dentro de un dominio y definen, con diferentes niveles de formalidad, el significado de los atributos de los términos y las relaciones entre ellos. En este trabajo, se ha utilizado el lenguaje de ontologías OWL, que es el estándar propuesto por la W3C para la representación de ontologías en la Web Semántica, para representar los conceptos y características del dominio.

El objetivo de este trabajo es describir un nuevo método de minería de opiniones basado en características que utiliza una ontología del dominio para la identificación de dichas características. Este método ha sido validado en el dominio de las opiniones de los usuarios sobre películas cinematográficas.

El resto del artículo está organizado de la siguiente manera: En la sección 2 se presentan trabajos relacionados; el método propuesto se explica en detalle en la Sección 3; la sección 4 muestra la validación de dicho método y finalmente, las conclusiones y trabajos futuros se exponen en la Sección 5.

2 *Trabajos Relacionados*

La clasificación sentimental está atravesando un gran momento y muchos investigadores del ámbito de la recuperación de información y la lingüística computacional han focalizado sus investigaciones en este tópico. Por ejemplo, Hatzivassiloglous y McKeown trabajaron en la orientación semántica (polaridad) de adjetivos (Hatzivassiloglous y McKeown, 1997). Desde entonces, se han empleado varias técnicas en la minería de opiniones tales como: medidas probabilísticas de asociación de palabras (Turney y Litman, 2003), técnicas que usan

información sobre relaciones léxicas (Kamps y Marx, 2002), (Kim y Hovy, 2004), y técnicas que usan bases de datos léxicas (Esuli y Sebastiani, 2005; Andreevskaja y Bergler, 2006). Todas ellas usan datos provenientes de las opiniones sobre automóviles, bancos, películas o destinos de viaje. Además, clasifican las palabras en dos categorías (positiva o negativa) y calculan la puntuación positiva y negativa total para el texto. Si los documentos contienen más términos positivos que negativos, se asume que el documento es positivo, en caso contrario negativo. Estos mecanismos de clasificación son útiles y mejoran la eficiencia de la clasificación sentimental, pero no pueden determinar la polaridad de cada característica particular dentro de una opinión determinada.

También existen estudios que determinan el nivel de polaridad como alto/medio/bajo positivo/negativo (Dave, Lawrence, y Pennock, 2003), (Goldberg y Zhu, 2006), (Shimada y Endo, 2008), (Baccianella, Esuli, y Sebastiani, 2009).

En (Pang et al., 2002) y (Mukras y Carrol, 2004) se explotan los datos provenientes de las opiniones sobre películas, el feedback de clientes y opiniones sobre productos. Usan un conjunto completo de métodos de selección de características estadísticos y aplican técnicas de aprendizaje computacional. Estos experimentos demuestran que las técnicas de aprendizaje computacional no obtienen resultados satisfactorios en la clasificación sentimental. En (Hu y Liu, 2004) se centran en obtener las opiniones de los usuarios sobre características de un producto. Estas características se obtienen usando minería de asociaciones entre palabras.

Más recientemente, se pueden destacar varios trabajos importantes sobre ontologías y minería de opiniones basadas en características (Zhou y Chaovalit, 2008), (Zhao y Li, 2009). El objetivo de estos trabajos es calcular la polaridad teniendo en cuenta las características de un concepto en una ontología. A diferencia de Hu y Liu, en esos trabajos se utilizan ontologías del dominio que modelan el contenido de los textos en el corpus y, a partir de dichas ontologías, extraen los adjetivos que describen dichas características. En el trabajo de Zhou y Chaovalit, la polaridad de un texto se calcula basándose en los pesos de las características, y el valor de la polaridad de la característica se calcula mediante la “estimación de máxima verosimilitud”. En el

trabajo de Zhao y Li, se obtiene la orientación semántica del texto desde la jerarquía de la ontología. Además, el método que proponen es capaz de obtener la positividad, negatividad y neutralidad de un texto. Estos dos trabajos demuestran que el uso de ontologías mejora los resultados en el problema del análisis sentimental.

Este trabajo se centra en obtener un método de análisis sentimental alternativo basado en ontologías que mejore los resultados obtenidos en estos dos trabajos citados.

3 Minería de opiniones basada en ontologías

Los principales objetivos de la aproximación propuesta en este artículo son: mejorar la minería de opiniones basada en características empleando ontologías en la selección de las mismas y proporcionar un nuevo método para el análisis sentimental basado en análisis vectorial.

El sistema presentado se compone de cuatro módulos principales: (ver Fig. 1): el módulo de Procesamiento del lenguaje natural (PLN), el módulo de identificación de características basado en ontologías, el módulo de identificación de la polaridad y el módulo de minería de opiniones. Cada uno de estos componentes se describe a continuación.

3.1 Módulo PLN

El principal objetivo de este módulo es obtener la estructura morfológica y sintáctica de cada frase de las opiniones del usuario. Para ello se han desarrollado un conjunto de herramientas (detección de frases, tokenizador, POS, lematizador y analizadores sintácticos) utilizando GATE (<http://gate.ac.uk/>). GATE es una infraestructura para desarrollar y desplegar componentes de software para el procesamiento del lenguaje natural.

Por ejemplo, el análisis morfológico llevado a cabo por el módulo NLP para la frase (en inglés) “Return of the Jedi is the latest episode of Star Wars but it’s the one I liked. I love the soundtrack.” obtiene el siguiente resultado:

Return_of_the_Jedi_NNP is_VBZ the_DT late_JJS episode_NN of_IN Stars_War_NNP but_CC it_PRP be_VBZ the_DT one_NN I_PRP like_VBD. I_PRP love_VBP the_DT soundtrack_NN.

3.2 Módulo de identificación de características basado en ontologías

En esta fase se utiliza una ontología del dominio para extraer las características a las que los usuarios hacen referencia en sus opiniones. En primer lugar, la ontología del dominio se enriquece con sinónimos utilizando la base de datos léxica para inglés de WordNet (<http://wordnet.princeton.edu/>).

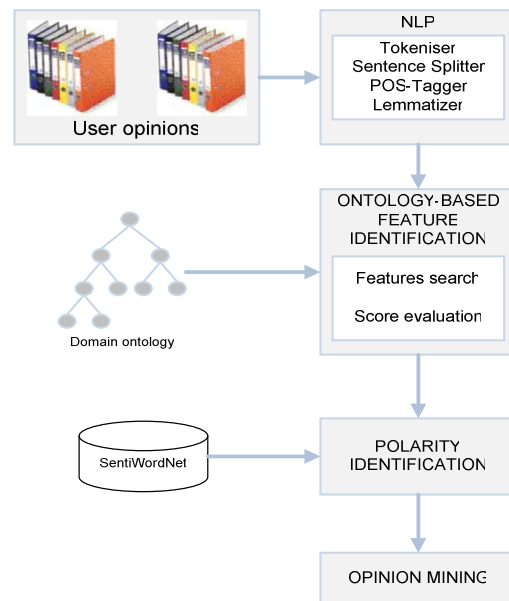


Figura 1: Arquitectura del sistema

Este módulo recibe el corpus de opiniones y la ontología del dominio como entrada. Seguidamente, las frases que contienen clases, instancias, y propiedades de la ontología son identificadas en el texto. Una vez que estas características han sido reconocidas son agrupadas por distancias semánticas y relacionadas con el concepto principal de la ontología.

Por ejemplo, vamos a suponer que la ontología del dominio modela el dominio de las películas cinematográficas (ver Fig. 2). Esta ontología ha sido poblada previamente con la información disponible en *The Internet Movie Database* (IMDb). Las características identificadas en el ejemplo mostrado en la sección anterior son {“Return of the Jedi”, “Star Wars”, “Soundtrack”}. Más concretamente, “Return of the Jedi” y “Stars War” son identificados como instancias de la

clase “*Movie*”, y “*Soundtrack*” es una clase relacionada con “*Movie*”. En este punto, gracias a la estructura semántica de las ontologías, las opiniones que tienen que ver con las clases que están directamente relacionadas con la clase “*Movie*”, como por ejemplo “*Actor*”, “*Producer*”, “*Genre*” y sus instancias, se tienen en cuenta para calcular la polaridad global.

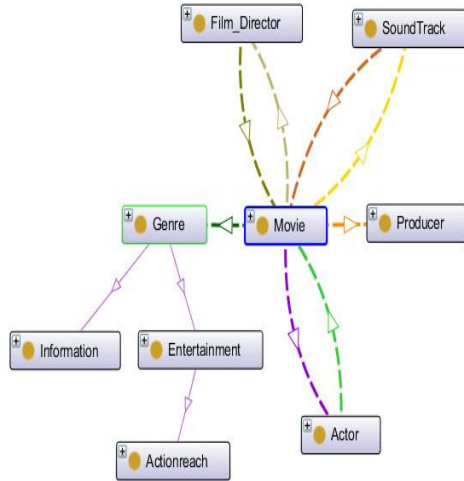


Figura 2: Un extracto de la ontología MovieOntology

Los métodos tradicionales de identificación de características dan igual importancia a todas las características identificadas en el texto como los siguientes trabajos: (Baccianella, Esuli, y Sebastiani, 2009), (Zhou y Chaovalit, 2008), (Zhao y Li, 2009) y (Moreno, Pineda e. Hidalgo, 2010). En nuestra aproximación, calculamos una puntuación para cada característica tomando como base las siguientes suposiciones:

1. No todas las características relacionadas con la misma clase tienen la misma relevancia. Por ejemplo, la nacionalidad de una película es previsiblemente menos relevante que los premios que ha conseguido, y por lo tanto será menos referenciada por los usuarios en sus opiniones.
2. Las características citadas más a menudo por los usuarios en sus opiniones tendrán más relevancia.
3. Desde el punto de vista del discurso periodístico, la polaridad del último párrafo

coincide aproximadamente con la polaridad global del texto (Moreno, Pineda e. Hidalgo, 2010). Empezar y terminar con una perspectiva positiva o negativa realza la polaridad positiva o negativa del texto. En este sentido, el lugar en el que una característica aparece en las opiniones de los usuarios debe ser considerada para calcular la puntuación de la característica en cada opinión. Por ejemplo, si la última frase de una opinión sobre una película es la siguiente: “indudablemente una película con una banda sonora impresionante”, esto realza la polaridad positiva de la característica “banda sonora” en la opinión del usuario, y por lo tanto esta característica debe tener una puntuación mayor.

Una vez que las características han sido identificadas en las opiniones, se calcula la puntuación de las características en cada opinión de usuario. Para cada característica recuperada la puntuación total, llamada *Tscore* (ver ecuación 1), se calcula como la media aritmética de la puntuación de la característica en cada opinión de los usuarios.

$$Tscore(f) = \sum_{i=1}^n \frac{score(f, userop_i)}{n} \quad (1)$$

En la fórmula para calcular la puntuación de una característica dada en una opinión de usuario, se tiene en cuenta la posición de la expresión lingüística que representa a la característica dentro del texto. Para esto, el texto se divide en tres partes iguales: (1) el comienzo, (2) la parte media, y (3) el final de la opinión. Esta puntuación se define en la ecuación 2.

$$score(f, userop_i) = z_1 * |O_1| + z_2 * |O_2| + z_3 * |O_3| \quad (2)$$

donde $|O_i|$ es el número de ocurrencias de la característica f en la parte del texto i de la opinión $userop_i$, y Z_i es un parámetro que representa la importancia de la ocurrencia de la característica en esta parte del texto. En este trabajo se segmenta el texto i de la opinión $userop_i$ en tres partes iguales basándonos en el número de palabras del texto. Se han considerado tres partes en cada opinión correspondientes a las partes: inicial, central y

final. Este método contempla dar más peso en la opinión a las características que aparecen en las partes inicial y final.

Más concretamente, en este trabajo se han utilizado los valores mostrados en la ecuación 3.

$$Z_1 = 0,3 \quad Z_2 = 0,2 \quad Z_3 = 0,5 \quad (3)$$

3.3 Módulo de identificación de características basado en ontologías

La polaridad de las características se calcula usando SentiWordNet 3.0 (Baccianella, Andrea y Sebastiani., 2010). Esta herramienta permite obtener valores de positividad, negatividad y neutralidad de los nombres, adjetivos y verbos localizados a continuación de la expresión lingüística que representa a una característica dada en la opinión de un usuario. En este trabajo se utilizan fórmulas basadas en la media aritmética de cada valor en SentiWordNet (positivo, negativo). Además, el valor neutral se calcula como: $sObj = 1 - sPos - sNeg$. Entonces, para cada valor de polaridad (positivo, negativo y neutro) para cada característica se define un vector como muestra la ecuación 4.

$$\vec{V}(f) = Tscore(f) * (a * sPos, a * sNeg, sObj) \quad (4)$$

donde “a” indica cuando una cláusula negativa se ha encontrado en la frase. Si existe esa cláusula negativa “a” tomaría el valor -1, en otro caso el parámetro “a” se establece a 1.

3.4 Módulo de minería de opiniones

En esta sección, se describe un nuevo método de minería de opiniones basado en análisis vectorial para la clasificación sentimental basada en características. Como se muestra en la sección anterior, las características se expresan por tres coordenadas, es decir, cada característica se determina por su vector euclidiano (V) en R^3 . Un vector euclidiano se representa por dos puntos, el punto de origen y el punto de destino. Dado que para el punto de origen utilizaremos siempre el punto $(0, 0, 0)$, la expresión del vector de posición se reduce a expresar el punto de destino. Por tanto, un vector de posición se expresa por $V = (x, y, z)$. Una característica con un sentido positivo estricto en SentiWordNet se determina por el

vector de posición $(1, 0, 0)$. Por analogía, los vectores $(0, 1, 0)$ y $(0, 0, 1)$ representan respectivamente los sentidos negativo y neutral estrictos.

La polaridad de una característica f en el conjunto de opiniones de los usuarios, $Polarity(f)$, se calcula como se muestra en la ecuación 5:

$$Polarity(f) = \begin{cases} +x \text{ (positive) if } inside(P;P_p) = true \\ -y \text{ (negative) if } inside(P;P_n) = true \\ 0 \text{ (neutral) if } inside(P;P_o) = true \end{cases} \quad (5)$$

conde:

- $P=(x, y, z)$, es el punto de destino de $V(f)$.
- P_p es la pirámide geométrica cuyo volumen está compuesto por todos los puntos de destino de los vectores de posición con dirección positiva.
- P_n es la pirámide geométrica cuyo volumen está compuesto por todos los puntos de destino de los vectores de posición con dirección negativa.
- P_o es la pirámide geométrica cuyo volumen está compuesto por todos los puntos de destino de los vectores de posición con dirección neutra.
- $inside(P;P_x)$ es una función que devuelve *true* si el punto P está contenido dentro de la pirámide geométrica P_x . En otro caso, devuelve *false*.

Para implementar la función $inside(P;P_x)$ es necesario saber si un punto $P = (x, y, z)$ es un punto interior de P_p , P_n o P_o . Para ello, se utilizan los “elementos definidos en topología de conjuntos” (Apostol, 2006):

- *Punto interior*: sea S un subconjunto de R^n y P un punto de R^n . Asumimos que $P \in S$. Entonces P se llama punto interior de S si existe una N -bola abierta centrada en P y contenida en S .
- *N-bola*: sea P un punto de R^n y R un número positivo dado. El conjunto de todos los puntos X en R^n tales que $\|X - P\| < R$ se llama N -bola abierta de radio R y centro P . Denotaremos este conjunto por $B(P)$ o $B(P;R)$.

En otras palabras, cada P , punto interior de S , puede ser rodeado por una n -bola $B(P) \subseteq S$. El conjunto de todos los puntos interiores de S se llama *interior de S*. La Figura 3 muestra una representación de R^3 donde se delimitan las pirámides P_{POS} , P_{NEG} y P_{NEU} .

Por otro lado, todas las características tienen asignadas una puntuación en concordancia con su valor $V(f)$. Entonces, la polaridad de todas las opiniones de usuarios (el análisis sentimental global), llamado *Polarity*, se calcula como el vector de posición resultante de la suma de todos los vectores de posición (puntuados) para cada característica identificada como muestra la fórmula 6.

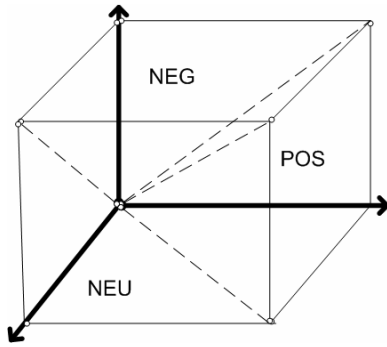


Figura 3: Pirámides geométricas de polaridad en R^3

$$Polarity = \sum_{i=1}^n \vec{V}(f_i) \quad (6)$$

El valor *Polarity* resultante será un vector euclídeo con tres coordenadas (x, y, z). Sea P el punto de destino del vector de posición resultante, $P=(x, y, z)$, entonces:

$$Polarity = \begin{cases} +x \text{ (positive) if } inside(P; P_p) = true \\ -y \text{ (negative) if } inside(P; P_n) = true \\ 0 \text{ (neutral) if } inside(P; P_o) = true \end{cases} \quad (7)$$

4 Caso de aplicación. Opiniones sobre películas

En esta sección, presentaremos los resultados obtenidos por el método propuesto en el dominio de las opiniones de los usuarios sobre películas cinematográficas.

El número de opiniones relativas al mundo del cine en general, y a las películas en particular, está en constante crecimiento. Para poder manipular la gran cantidad de datos relacionados con las opiniones sobre películas, existe la creciente necesidad de utilizar herramientas de minería de opiniones que puedan evaluar algunas de las características relevantes de las películas.

Teniendo en cuenta las dificultades de desarrollar una nueva ontología desde cero, para los propósitos de nuestra investigación hemos reutilizado la ontología sobre películas Movie Ontology MO¹. Esta ontología pretende proporcionar un vocabulario controlado para describir semánticamente los conceptos relacionados con las películas, tales como Película (*Movie*), Género (*Genre*), Director (*Director*) o Actor (*Actor*). Un total de 100 películas diferentes han sido insertadas en la ontología como instancias para los propósitos experimentales.

El corpus del experimento contiene 15,323 palabras y comprende 64 opiniones (31 negativas y 33 positivas). Este corpus ha sido extraído del corpus usado en (Pang y Lee, 2004).

En el experimento, se estudian los resultados obtenidos etiquetando manualmente las características como línea base en contraste con los resultados de nuestra aproximación, y las aproximaciones de (Zhou y Chaovalit, 2008) y (Zhao y Li, 2009). Los resultados para la identificación de características se muestran en la Tabla 1. Estos resultados indican que la precisión en nuestra aproximación es superior que la que se puede obtener mediante los métodos anteriores. Creemos que la clave se encuentra en el desarrollo de la ontología, cuidadosamente completada y poblada para el dominio que nos ocupa.

A continuación, se ha llevado a cabo la evaluación del método de análisis sentimental propuesto en la sección 3.5. Como el conjunto de datos seleccionado está ya etiquetado, usamos la etiqueta de la clase origen (Pos y Neg) como línea base. Los resultados se muestran en la Tabla 2. Los resultados para la precisión en el análisis sentimental parecen prometedores obteniendo un valor del 84,8 % y 87,1% para los textos positivos y negativos respectivamente. Estos valores superan con mucho los que se obtienen mediante el empleo de los otros métodos.

Después de una reflexión, encontramos que los factores clave para la obtención de estos resultados han sido dos: (i) método de asignación de pesos a características en función de su posición y frecuencia en el texto, y (ii) método de análisis vectorial para la minería de

¹ Amancio Bouza: "MO – the Movie Ontology". MO – the Movie Ontology. 2010. movieontology.org

opiniones. En la Tabla 3 se muestran cuales habrían sido los resultados obtenidos si se hubiera realizado la minería de opiniones con nuestro método sin ponderar las características identificadas en función de su número de ocurrencias y su posición relativa en cada opinión. Claramente, la precisión que obtenemos es inferior.

	Línea Base	Nuestra aproxim.	Zhou y Chaovalit	Zhao y Li
Caract.	15	14	13	11
Precis.		93.3%	86.6%	73.3%

Tabla 1: Resultados de identificación de características

	Línea Base	Nuestra aproxim.	Zhou y Chaovalit	Zhao y Li
Pos	33	28 Precisión: 84.8%	24 Precisión: 72.7%	26 Precisión: 78.7%
Neg	31	27 Precisión: 87.1%	23 Precisión: 74.1%	25 Precisión: 80.6%

Tabla 2: Resultados del análisis sentimental

	Línea Base	Nuestra aproxim. (sin pesos ni frecuencia)
Pos	33	25 Precisión: 75,8%
Neg	31	23 Precisión: 74.1%

Tabla 3: Resultados sin asignación de pesos ni frecuencia de las características

5 *Discusión y conclusión*

La minería de opiniones es un problema cambiante y determina cuándo esas opiniones son positivas, negativas o neutras (Zhao y Li, 2009). En este trabajo, nos hemos centrado en definir un nuevo método para la minería de opiniones basada en características. Este método proporciona un conjunto completo de novedades tales como i) un mecanismo basado en ontologías para identificar características, ii) un método para asignar un peso a cada característica basado en su número de

ocurrencias y su posición relativa en cada opinión de usuario, y (iii) una nueva aproximación para la minería de opiniones basada en cálculos con análisis vectorial.

En nuestra aproximación, la ontología contiene no solo conceptos básicos, sino también conceptos específicos, atributos, relaciones e instancias. Además, toda la información disponible para representar las características está representada en la ontología.

La aproximación propuesta ha sido validada en el dominio de las películas y los resultados parecen prometedores. Se pretende realizar una evaluación con todo el corpus utilizado en (Pang y Lee, 2004) previa identificación manual de todas las características de cada opinión de usuario, ya que en este corpus únicamente se identifica la polaridad global del texto y no la polaridad de cada característica.

Además, como trabajo futuro se ha planeado realizar una mayor validación en otros dominios, tales como el dominio de los productos, y usar métodos estadísticos para analizar los resultados obtenidos.

La minería de opiniones basada en características para las opiniones sobre productos es una tarea difícil, dada la gran variedad semántica de la expresión de las opiniones, y también la diversidad de características y sub-características que describen los productos, y la gran cantidad de palabras usadas en las opiniones sobre ellos (Balahur y Montoyo, 2009).

Estamos trabajando actualmente en actualizar este sistema y validarlo en las opiniones sobre productos en el idioma español.

Bibliografía

- Andreevskaia, A. y S. Bergler. 2006. "Mining wordnet for a fuzzy sentiment: Sentiment tag extraction from wordnet glosses," Proceedings of the 11rd Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006), pp. 209–216.
- Apostol, T.M. 2006. "Mathematical Analysis," Addison-Wesley Publishing Company, Inc. Reading, Massachusetts, U.S.A.
- Balahur A. y A. Montoyo. 2009. "Semantic Approaches to Fine and Coarse-Grained Feature-Based Opinion Mining", In Proceedings NLDB 2009, LNCS 5723, pp. 142–153

- Baccianella, S., E. Andrea y F. Sebastiani. 2010. SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Proceedings of the Seventh Conference on International Language Resources and Evaluation, 2200–2204. European Language Resources Association.
- Baccianella, S., A. Esuli, y F. Sebastiani. 2009. “Multi-facet rating of product reviews,” ECIR '09: Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, pp. 461–472.
- Dave, K., S. Lawrence, y D. M. Pennock. 2003. “Mining the peanut gallery: opinion extraction and semantic classification of product reviews,” WWW '03: Proceedings of the 12th international conference on World Wide Web, pp. 519–528.
- Esuli, A. y F. Sebastiani. 2005. “Determining the semantic orientation of terms through gloss classification,” CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management, pp. 617–624.
- Esuli, A. y F. Sebastiani. 2006. “Sentiwordnet: A publicly available lexical resource for opinion mining”. En: Proceedings of 5th Conference on Language Resources and Evaluation, LREC 2006.
- Goldberg, A. B. y X. Zhu. 2006. “Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization,” TextGraphs '06: Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing on the First Workshop on Graph Based Methods for Natural Language Processing, pp. 45–52.
- Hatzivassiloglou, V. y K. R. McKeown. 1997. “Predicting the semantic orientation of adjectives,” Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics, pp. 174–181.
- Hu, M. y B. Liu. 2004. “Mining and summarizing customer reviews,” KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168–177.
- Kamps, J. y M. Marx. 2002. “Words with attitude,” 1st International WordNet Conference, pp. 332–341.
- Kim, S. y E. Hovy. 2004. “Determining the sentiment of opinions,” COLING '04: Proceedings of the 20th international conference on Computational Linguistics, pp. 1267–1373.
- Moreno, A., F. Pineda y R. Hidalgo. 2010. “Análisis de Valoraciones de Usuario de Hoteles con Sentitext: un sistema de análisis de sentimiento independiente del dominio”, Procesamiento del Lenguaje Natural, vol 45.
- Mukras, R. y J. Carroll. 2004. “A comparison of machine learning techniques applied to sentiment classification”.
- Pang et al. 2002. “Sentiment Classification using machine learning methods”. En EMNLP-2002.
- Pang, B. y L. Lee. 2004. “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”, In Proceedings of the ACL, pp. 271--278.
- Shimada, K. y T. Endo. 2008. “Seeing several stars: A rating inference task for a document containing several evaluation criteria,” PAKDD 2008: Proceedings of Advances in Knowledge Discovery and Data Mining, 12th Pacific-Asia Conference, pp. 1006–1014.
- Studer, R., V. R. Benjamins y D. Fensel. 1998. Knowledge engineering: Principles and methods, Data Knowl. Eng., vol. 25, pp. 161-197.
- Turney, P. D. y M. L. Littman. 2003. “Measuring praise and criticism: Inference of semantic orientation from association,” ACM Trans. Inf. Syst., vol. 21, no. 4, pp. 315–346.
- Zhao, L. y C. Li. 2009. “Ontology based opinion mining for movie reviews,” in Proc. of KSEM, pp. 204–214.
- Zhou, L. y P. Chaovalit. 2008. “Ontology-supported polarity mining,” Journal of the American Society for Information Science and Technology, vol. 59, no. 1, pp. 98–110.