

Tipo de artículo: Artículo original  
Temática: Reconocimiento de patrones  
Recibido: 05/10/2015 | Aceptado: 14/12/2015

## Método para la Clasificación de Polaridad basado en Aspectos de Productos

### *Aspect-Based Polarity Classification Method on Products Review*

Carlos E. Muñiz-Cuza<sup>1\*</sup>, Reynier Ortega-Bueno<sup>1</sup>

<sup>1</sup>Centro de Reconocimiento de Patrones y Minería de Datos. Ave Patricio Lumumba S/N, Universidad de Oriente, Santiago de Cuba, Cuba. {carlos, reynier.ortega}@cerpamid.co.cu

\*Autor para correspondencia: [carlos@cerpamid.co.cu](mailto:carlos@cerpamid.co.cu)

---

#### Resumen

Este artículo presenta un método para la clasificación de la polaridad de aspectos de productos. La característica más relevante de la propuesta radica en la construcción automática de recursos de polaridad dependientes del dominio a través del empleo de la técnica de *Análisis de Semántica Latente*. El método permite generar recursos de polaridad para varias unidades textuales como bigramas y trigramas y es independiente del idioma. La clasificación de la polaridad de los aspectos es realizada en dos fases fundamentales: extracción de las palabras y frases de opinión, y la clasificación de la polaridad. La etapa de extracción de las palabras y frases de opinión consiste en extraer de un contexto lineal y sintáctico relacionado con el aspecto las unidades textuales para las cuales fueron generadas recursos de polaridad. Finalmente, la polaridad del aspecto, en una crítica dada, es determinada por los valores de polaridad positivo y negativo de cada una de las palabras y frases de opinión extraídas. Los resultados obtenidos por la propuesta son alentadores si consideramos que el proceso de construcción de los recursos se realiza completamente de manera automática.

**Palabras claves:** polaridad, análisis de semántica latente, contexto lineal y sintáctico

#### Abstract

*This paper presents a method for aspect-based sentiment analysis on user products reviews. The most outstanding feature in this proposal is the automatic building of domain-depended sentiment resource using Latent Semantic Analysis. The proposed method can be adapted to different textual units such as bigrams and trigrams and is language independent. The aspect term polarity classification is carried out in two phases: opinion words and phrases extraction and polarity classification. The extraction phase involve the search of surface and parse feature of the aspect, getting polarities scores of the textual units generated on the previous phase. Finally, the polarity of the aspect, in a given review, is determined from the positive and negative scores of each words and phrases extracted. The results obtained by the approach are encouraging if we consider that the construction of the domain-dependent polarity lexicon is performed fully automatic.*

**Keywords:** *polarity, latent semantic analysis, surface feature, parse feature*

---

## Introducción

Conocer el estado de opinión respecto a un producto, la reputación de una celebridad o la valoración sobre un candidato político resulta de mucha utilidad en el mundo actual. Por ejemplo, en el caso particular de una empresa, realizar un análisis sobre las opiniones de sus clientes resulta muy beneficioso, pues le permite conocer las inconformidades y, a partir de estas, definir nuevas estrategias que le garanticen obtener un mayor grado de satisfacción en sus usuarios, así como un éxito superior dentro del mercado. En muchas aplicaciones es importante asociar la opinión a una entidad en específico o a los aspectos de esta entidad. En la oración: “*El nuevo modelo de teléfonos iPhone 6 tiene una gran calidad de sonido, pero el tiempo de vida de la batería es muy corto*” se puede conocer que el cliente tiene un criterio positivo respecto a la calidad de sonido, pero negativo con respecto al tiempo de vida de la batería. El *Análisis de Sentimientos a Nivel de Aspecto* (HU and LIU, 2004) tiene como objetivo identificar los aspectos (ej., *batería, pantalla, comida, servicio, tamaño, peso*) de entidades específicas (ej., *laptop, restaurantes, películas, cámaras*) y la opinión que se emite con respecto a estas (ej., *positivo, negativo, neutro*). La misma está compuesta por dos fases principales: extracción de los aspectos y clasificación de los mismos.

Uno de los grandes retos que presenta la tarea, es la dependencia del dominio de las palabras que se emplean para emitir una opinión. Por ejemplo, la palabra *impredecible* puede ser considerada positiva en un dominio de películas, sin embargo es muy negativa en un dominio relacionado con el control de aviones. Por esta razón, muchos autores proponen la creación de recursos de polaridad dependientes del dominio como paso previo en la clasificación de la polaridad de los aspectos. El uso de estos recursos de polaridad ha probado ser útil en la clasificación, entrenamiento y evaluación de los sistemas para el análisis de sentimientos (MATA, 2011). Uno de los primeros trabajos, presentado por (HATZIVASSILOGLOU and MCKEOWN, 1997), predice la polaridad de los adjetivos, analizándolos en pares (unidos por conjunciones: *and, or, but, either-or, or, neither-nor*). La idea subyacente es el hecho de tener adjetivos unidos por conjunciones, está sujeto a restricciones lingüísticas con respecto a la orientación de los adjetivos involucrados (ej., *and* usualmente une dos adjetivos con la misma polaridad, mientras que *but* une adjetivos con diferente polaridad).

A pesar de la gran variedad de métodos para la generación de recursos de polaridad (TURNERY, 2002; EDUARD C. DRAGUT and MENG, 2010; KAMPS et al., 2004), los resultados están aún lejos de alcanzar los resultados esperados. Como se explicó anteriormente, la polaridad de las palabras que se emplean para emitir una opinión es altamente dependiente del dominio de aplicación y los recursos de polaridad de propósito general como *General Inquirer* (STONE et al., 1968) y *SentiWordNet* (BACCIANELLA et al., 2010) no capturan esta información. Por otra parte, no se puede construir de manera manual un nuevo recurso para cada producto en el mercado. Por tanto, proponer métodos para la construcción de estos recursos es todavía una tarea desafiante.

Este artículo se enfoca solo en el problema de clasificación de la polaridad de los aspectos tomando en cuenta el problema anteriormente planteado. Con tal objetivo, se construye de manera automática recursos de polaridad dependientes del dominio empleando *Análisis de Semántica Latente* (DEERWESTER et al., 1990) (*LSA*, por sus siglas en inglés) sobre colecciones de críticas de productos recolectadas de sitios como *Ciao*<sup>1</sup> y *Epinions*<sup>2</sup>. Para la etapa de clasificación es considerado el contexto lineal y sintáctico del aspecto. En el contexto lineal son extraídos  $n$ -gramas de palabras ( $n$  en el rango de 1 a 3). Por otra parte, en el contexto sintáctico son extraídos los términos cercanos al aspecto en el grafo de dependencia y  $n$ -gramas sintácticos no continuos (SIDOROV, 2013a,b) ( $n$  en el rango de 2 a 3). Finalmente, estos valores de polaridad son combinados para determinar la polaridad del aspecto.

## Descripción del Método Propuesto

Nuestra propuesta está compuesta por tres fases. La primera, consiste en descargar críticas de usuarios desde los sitios Web *Ciao* y *Epinions* (en este trabajo se empleó además, las colecciones de críticas de usuarios de *Yelp*<sup>3</sup>). Para cada página, se extraen los campos de texto *pros*, *cons* y el campo de texto que corresponde al texto libre. En estos campos de texto los usuarios comentan brevemente lo que consideran relevante respecto al producto y sus aspectos más importantes. Por tal motivo, se puede asumir que los términos empleados en estos campos de textos están muy relacionados semánticamente con los términos que se emplean habitualmente en el dominio para emitir una opinión positiva y negativa en el caso de los campos *pros* y *cons* respectivamente. Análogamente, para aquellas críticas que no presentan *pros* y *cons* pero que si están clasificadas por los usuarios a través de estrellas, asumimos que una clasificación de cinco o cuatro estrellas es positivo y una clasificación de dos y una estrella es negativo. Las oraciones contenidas en estos fragmentos de textos serán consideradas en lo adelante como ejemplos positivos y negativos. A la vez, las oraciones que no se encuentren bajo esta clasificación serán clasificadas como texto libre.

Una vez construida la colección de ejemplos positivos y negativos, cada oración es segmentada por palabras y se le aplica un análisis morfosintáctico con el propósito de obtener, para cada una de ellas, su lema y la categoría gramatical, eliminando las palabras sin contenido semántico o palabras auxiliares (*stopwords*, en inglés). Este preprocesamiento es realizado empleando la herramienta de Procesamiento del Lenguaje Natural *FreeLing* (PADRÓ and STANILOVSKY, 2012). Luego de eliminadas las *stopwords* cada oración es representada usando el modelo clásico de espacio vectorial. Para el caso particular de este trabajo el modelo de espacio vectorial fue generado para cinco unidades textuales,  $n$ -gramas de términos ( $n$  en el rango de 1 a 3) y bigramas y

---

<sup>1</sup><http://www.ciao.com>

<sup>2</sup><http://www.epinions.com>

<sup>3</sup>[http://www.yelp.com/dataset\\_challenge](http://www.yelp.com/dataset_challenge)

trigramas sintácticos no continuos (para generar el grafo de dependencia se empleó la herramienta *Stanford Parser* (MARIE-CATHERINE DE MARNEFFE and MANNING, 2006). En este modelo, cada componente de los vectores representa la frecuencia de aparición de la unidad textual en la oración. Una vez generados estos vectores, con la intención de evaluar la asociación semántica entre las unidades textuales y las clases, se adicionan dos términos artificiales. En el caso de los ejemplos positivos es adicionado  $t_{pos}$ , mientras que en los ejemplos negativos es adicionado  $t_{neg}$ . El peso otorgado a estas nuevas componentes es igual a la cantidad de términos del vector. Por último, se aplica *Análisis de Semántica Latente* (en el trabajo se utilizó el paquete de Python, Gensim) para calcular la asociación semántica entre las unidades textuales y las clases.

La *Descomposición en Valores Singulares* (SVD, por sus siglas en inglés) es empleada por LSA para medir la relación estadística entre las unidades textuales dentro de una colección de documentos. El primer paso para ello es construir una matriz  $M_{n \times m}$ , en la cual los vectores filas  $v_i$  representan las unidades textuales y los vectores columnas  $s_i$  las oraciones. El próximo paso es la aplicación de SVD a la matriz  $M_{n \times m}$  para descomponerla en el producto de las matrices  $U \Sigma V^T$ . Luego, se seleccionan los  $k$  primeros valores singulares y sus correspondientes vectores singulares en  $U$  y  $V$ , obteniendo las matrices reducidas de rango  $k$ :  $U_k$ ,  $\Sigma_k$  y  $V_k^T$ . Posteriormente, se calcula la matriz  $\hat{M} = U_k \Sigma_k$  y sobre esta se determina el grado de relación semántica entre las unidades textuales  $u_i$  y  $u_j$  asociadas a los vectores  $v_i$  y  $v_j$  de la matriz  $\hat{M}$  a través de la función de similitud coseno (ver ecuación 1).

$$LSA(u_i, u_j) = \frac{\langle v_i, v_j \rangle}{\|v_i\| \|v_j\|} \quad (1)$$

Finalmente, los recursos de polaridad contienen para cada unidad textual  $u_i$  un valor de polaridad positivo y negativo. Este valor es computado mediante  $LSA(u_i, t_{pos})$  y  $LSA(u_i, t_{neg})$  respectivamente.

Para el caso particular del recurso generado para unigramas como unidad textual, el proceso termina con la eliminación de un conjunto de términos con baja carga subjetiva. La idea subyacente consiste en que la presencia de estos términos (ej., *self*, *hi*, *breakfast*, *laptop*) con carga de polaridad en los recursos generados afecta de manera significativa la etapa de clasificación de la polaridad. Con el propósito de reducir este impacto negativo, se emplea *SentiWordNet* para medir el grado de objetividad de los términos y eliminar aquellos que no presentan carga subjetiva en ninguna de sus acepciones.

## Clasificación de la polaridad de los aspectos

Con el objetivo de emplear los recursos generados, se propone un método no supervisado basado en reglas para clasificar la polaridad del aspecto. La propuesta consiste en la extracción de las unidades textuales generadas en la etapa anterior presentes en un contexto lineal y sintáctico del aspecto. Por una parte, el contexto lineal consiste en una ventana de términos alrededor del aspecto en la oración. Mientras que el contexto sintáctico consiste en aquellos términos (representados como nodos en el grafo de dependencias sintáctica) que

se encuentren a una distancia menor igual que la especificada del nodo que representa al aspecto.

En el contexto lineal, son extraídos los unigramas, bigramas y trigramas de términos, mientras que en el contexto sintáctico son extraídos los unigramas, bigramas y trigramas sintácticos no continuos que se generan a partir del aspecto. Para cada una de estas unidades textuales encontradas en ambos contextos se obtiene a partir de los recursos generados sus valores de polaridad positivo y negativo.

Los modificadores de polaridad (XIAODAN ZHU and KIRITCHENKO, 2014) fueron tomados en cuenta a partir de un diccionario *Neg* de palabras de negación,  $Neg = \{not, never, no, don't, n't, nothing\}$ . En el caso de la búsqueda en el contexto lineal, se invierte la polaridad de aquellos términos que presentan en una ventana de tamaño dos a su izquierda una de las palabras de negación anteriores. Por otra parte, en el caso de la búsqueda en el contexto sintáctico, es invertida la polaridad de los términos relacionados directamente con una de las palabras de negación anteriores. Además, se asumió que mientras más alejado se encuentran los términos del aspecto menor será su impacto en la polaridad del mismo. Para modelar esta idea en el caso del contexto lineal se dividen los valores de polaridad por la cantidad de términos entre el término y el aspecto. El caso del contexto sintáctico se divide por la distancia mínima en el grafo de dependencia entre el término y el aspecto. Finalmente, es calculado un valor de polaridad global positivo y negativo de la siguiente manera:

$$SO_{pos}(f_i) = \sum_{u \in U} Pos(u) \quad (2)$$

$$SO_{neg}(f_i) = \sum_{u \in U} Neg(u) \quad (3)$$

Donde  $U$  es el conjunto de todas las unidades textuales extraídas en ambos contexto y las funciones  $Pos(u)$  y  $Ne(u)$  retornan los valores de polaridad otorgados para la unidad textual en los recursos de polaridad generados. A partir de los valores  $SO_{pos}$  y  $SO_{neg}$ , se realiza una combinación lineal de los valores de polaridad obtenidos para ambos contextos, empleando las ecuaciones 4 y 5.

$$P_{score} = \alpha \times SO_{pos_{cl}} + (1 - \alpha) \times SO_{pos_{cs}} \quad (4)$$

$$N_{score} = \alpha \times SO_{neg_{cl}} + (1 - \alpha) \times SO_{neg_{cs}} \quad (5)$$

Una vez calculado estos valores la regla de decisión clasifica la polaridad del aspecto en positiva si  $P_{score} > N_{score}$  y supera un umbral  $\epsilon$ . Por el contrario, si  $P_{score} < N_{score}$  y este valor es menor que  $-\epsilon$  entonces la polaridad del aspecto es negativa. Finalmente, si  $|P_{score} - N_{score}| < \epsilon$  la polaridad del aspecto es considerada

como neutro.

## Entorno Experimental y Resultados

La Octava Conferencia de Análisis Semántico (SemEval 2014) propuso el *Análisis de Sentimientos a Nivel de Aspectos* como la cuarta tarea en la agenda (PONTIKI et al., 2014). Para la evaluación de esta tarea, los organizadores distribuyeron dos colecciones, divididas en entrenamiento y prueba, anotadas manualmente a nivel de aspectos, para los dominios de restaurante y laptop. Estas colecciones fueron las empleadas en los experimentos para evaluar el método de clasificación de polaridad propuesto. En estas colecciones las categorías empleadas para determinar la polaridad de los aspectos fueron *positivo*, *negativo*, *neutro* y *conflictivo*. Las tres primeras categorías corresponden a las clases frecuentemente empleadas para la tarea, mientras que la clase *conflictivo* corresponde a aquellos aspectos sobre los cuales se emite una opinión positiva y negativa simultáneamente. La tabla 1 muestra la distribución de las colecciones en correspondencia a las categorías empleadas. Teniendo en cuenta las colecciones de entrenamiento y prueba expuestas anteriormente, fue nece-

Tabla 1. Aspectos de productos y sus polaridades por dominio

Colecciones	Pos	Neg	Con	Neu	Total
LaptopTrain	987	866	45	460	2358
LaptopTest	341	128	16	169	654
RestauranteTrain	2164	805	91	633	3693
RestauranteTest	728	196	14	196	1134

sario para la clasificación de polaridad crear recursos de opinión para los dominios de laptop y restaurantes. Con tal objetivo, se descargaron 3007 críticas de usuarios relacionadas con laptop y 6053 para restaurante, las cuales fueron agregadas a las colecciones de sus respectivos dominios. Además como ejemplos de la clases positivo y negativo fueron agregadas las oraciones de los conjuntos de entrenamiento de SemEval 2014 en las cuales todos los aspectos extraídos contenían la misma polaridad. Los recursos fueron generados para todas las unidades textuales para los valores de  $k = \{100, 200, 300\}$  en el algoritmo LSA. Con estos recursos se realizaron varias corridas del método sobre el conjunto de entrenamiento para evaluar la calidad de los valores de polaridad generados. Después de este proceso, y luego de algunas revisiones manuales, se decidió emplear para la clasificación de polaridad los recursos generados por LSA empleando un  $k = 100$  en el caso de unigramas y trigramas sintácticos, en el caso de los bigramas  $k = 200$  y para los bigramas sintácticos y trigramas de términos  $k = 300$ . La tabla 2 muestra los diez primeros términos positivos y negativos para el dominio de laptop.

Tabla 2. Los diez primeros unigramas negativos y positivos para el dominio de laptop

Negativos+	Score	Positivos+	Score
<b>get-go</b>	0,976685	<b>unparalleled</b>	0,923310
<b>arse</b>	0,955711	<b>substantive</b>	0,916410
<b>cranky</b>	0,946873	amp	0,834991
<b>faceless</b>	0,936896	<b>lightweight</b>	0,735993
<b>inefficient</b>	0,856263	<b>enrage</b>	0,710885
indirectly	0,848760	<b>dependability</b>	0,695059
<b>fiend</b>	0,844514	intuition	0,689393
verity	0,827673	coupon	0,626067
vibration	0,815749	<b>inexpensive</b>	0,624734
<b>overpriced</b>	0,795222	<b>cachet</b>	0,589612

Para evaluar el aporte de los recursos de polaridad generados para diferentes unidades textuales se realizaron varias pruebas sobre el conjunto de entrenamiento de SemEval 2014. En la tabla 3 se muestra el impacto de utilizar los unigramas, bigramas y trigramas en el contexto lineal<sup>4</sup> y los unigramas, y bigramas y trigramas sintácticos en el contexto sintáctico<sup>5</sup>. Los resultados fueron evaluados en términos de exactitud (*accuracy*, en el inglés), definidos como la cantidad de aspectos bien clasificados sobre el total de aspectos. Como se

Tabla 3. Impacto de considerar los valores de polaridad de las unidades textuales generadas para ambos contextos

<i>unidad textual usada</i>	Contexto Lineal		Contexto Sintáctico	
	Laptop	Restaurante	Laptop	Restaurante
<b>unig</b>	66,41 %	64,95 %	63,35 %	61,02 %
<b>unig + big</b>	73,75 %	71,49 %	<b>66,07 %</b>	<b>65,09 %</b>
<b>unig + big + trig</b>	<b>73,97 %</b>	<b>73,11 %</b>	65,81 %	64,98 %

puede apreciar, los mejores resultados se alcanzaron mediante la combinación de los valores de polaridad en el contexto lineal de los unigramas, bigramas y trigramas. Este resultado demuestra que los valores de polaridad generados para los bigramas y trigramas están en correspondencia con los valores esperados. Por otra parte, en el contexto sintáctico los mejores resultados son alcanzados cuando se combinaron los unigramas y bigramas, mientras que el empleo de los trigramas sintácticos no continuos tuvo un impacto negativo en ambos dominios. A consideración de los autores, este comportamiento se debe a las pocas coincidencias encontradas en los recursos generados. Por otra parte, puede ser producto de valores de polaridad artificiales derivados del proceso de generación automática como resultado de la poca coocurrencia de esta unidad textual en las colecciones

<sup>4</sup>Una evaluación empírica sobre el conjunto de entrenamiento determinó que 6 términos es el mejor tamaño de ventana

<sup>5</sup>De manera similar al contexto lineal, se determinó que 4 términos es el mejor tamaño de ventana.



empleadas para la generación de los recursos. Por este motivo, se descarta el uso de esta unidad textual en el método de clasificación de polaridad y no se utiliza en las siguientes evaluaciones de la propuesta.

Con el objetivo de evidenciar el aporte de los recursos generados en la fase clasificación de la polaridad, se realizó una ejecución del algoritmo sobre el conjunto de prueba de SemEval 2014 empleando la combinación de los contextos lineales y sintácticos. Además, se realizó una prueba para una variante del método que solo tiene en cuenta los valores de polaridad otorgados por *SentiWordNet* (SWN) a los unigramas en el contexto lineal y sintáctico del aspecto. Por último, se diseñó un método que combina los valores de polaridad que asigna SWN y los valores de polaridad en los nuestros recursos generados de manera automática.

En la tabla 4, *BrLexGen* indica la variante del método que considera solamente los recursos generados.

Tabla 4. Aporte de los recursos de polaridad generados

Variantes del Método	F1pos	F1neg	F1neu	Macro-F1	Exactitud
<b>Dominio Laptop</b>					
<b>BrLexGen</b>	<b>79,51 %</b>	55,26 %	2,28 %	34,26 %	61,16 %
<b>BrSWN</b>	70,94 %	50,00 %	7,77 %	32,18 %	55,81 %
<b>BrLexGenSWN</b>	78,94 %	56,35 %	2,29 %	34,37 %	61,46 %
<b>BrLexGenSWN*</b>	78,97 %	<b>59,42 %</b>	<b>46,19 %</b>	<b>46,14 %</b>	<b>64,98 %</b>
<b>Dominio Restaurante</b>					
<b>BrLexGen</b>	85,08 %	57,63 %	13,6 %	39,10 %	72,75 %
<b>BrSWN</b>	80,55 %	49,80 %	16 %	36,10 %	66,84 %
<b>BrLexGenSWN</b>	85,59 %	<b>60,70 %</b>	13,69 %	40,00 %	73,45 %
<b>BrLexGenSWN*</b>	<b>86,05 %</b>	58,72 %	<b>39,34 %</b>	<b>46,03 %</b>	<b>73,63 %</b>

El elemento *BrSWN* representa el algoritmo considerando los valores de polaridad de SWN. Finalmente, *BrLexGenSWN* considera la combinación de los valores de polaridad de ambos recursos. Como se puede apreciar, el uso individual de los recursos generados y la combinación de estos con SWN supera los resultados alcanzados por la variante usando solamente los valores de polaridad de SWN. Nótese, que en sentido general, la variante del método que utiliza valores de polaridad de ambos recursos, supera los resultados para la variante que solo emplea los recursos generados automáticamente. Es por ello, que esta variante es la empleada para el ajuste de los parámetros  $\alpha$  y  $\epsilon$ , en la tabla 4 se identifica como *BrLexGenSWN\**. Después de varias ejecuciones sobre el conjunto de entrenamiento los mejores valores de estos parámetros fueron  $\alpha = 0,4$  y  $\epsilon = 0,1$ . Como se puede apreciar, el ajuste de estos parámetros tiene un impacto favorable en los resultados finales para ambos dominios. Véase, el incremento considerable de los valores de *F1* alcanzados para la clase neutral respecto a los valores obtenidos por *BrLexGenSWN*. A pesar del incremento de los resultados medidos en términos de MacroF1 respecto al método *BrLexGenSWN*, parte del deterioro de estos valores se debe a que no se diseñó ninguna estrategia en el método para la clasificación respecto a la clase conflictiva.



## Comparación con otros sistemas del método propuesto

Con el objetivo de establecer una comparativa del método propuesto con otras aproximaciones reportadas en la literatura se comparó la propuesta con los sistemas participantes en la subtarea de clasificación de polaridad de aspecto de SemEval 2014. En la tabla 5, se puede observar que el método propuesto se encuentra para cada dominio dentro de los 8 primeros sistemas de un total de 31 enviados<sup>6</sup>, resultado que evaluamos como significativo considerando el enfoque no supervisado de nuestra propuesta.

Tabla 5. Comparación de los métodos propuestos con los sistemas participantes en SemEval 2014

Lugar	Dominio Laptop		Dominio Restaurante	
	Equipo	Exactitud	Equipo	Exactitud
1	DCU	70,48 %	DCU	80,95 %
2	NRC-Can	70,48 %	NRC-Can	80,15 %
3	SZTE-NLP	66,97 %	UWB	77,68* %
⋮	⋮	⋮	⋮	⋮
6	BrLexGenSWN	64,98* %	UNITOR	74,95* %
7	lsis lif	64,52 %	UBham	74,6 %
8	USF	64,52 %	BrLexGenSWN	73,63* %
⋮	⋮	⋮	⋮	⋮
26	UO-UA	55,19* %	Blinov	63,58*
⋮	⋮	⋮	⋮	⋮
31	DLIREC	36,54* %	DLIREC	41,71 %

Por otra parte, el método supera en un 9,79% al método *UO-UA* (ORTEGA et al., 2014) que corresponde a uno de los sistemas participantes en la competición SemEval 2014 para la tarea cuatro y constituye el antecedente directo de este trabajo. En el mismo se emplea LSA para la generación de los recursos tomando en consideración solamente los unigramas como unidad textual. Además, en este sistema solo se consideran los términos que están directamente relacionados al aspecto en el árbol sintáctico para la clasificación de polaridad del aspecto.

## Conclusiones

En este artículo se presenta y evalúa un método para la clasificación de la polaridad de los aspectos en críticas de productos de usuarios. De manera general, se presenta una metodología para la generación de recursos de polaridad dependientes del dominio aplicando Análisis de Semántica Latente y de manera concreta se

<sup>6</sup>El asterisco implica el empleo de recursos externos por parte de los métodos

generaron estos recursos para los dominios de laptop y restaurante para  $n$ -gramas de términos ( $n$  en el rango de 1 a 3) y  $n$ -gramas sintácticos no continuos ( $n$  en el rango de 2 a 3). Con excepción del recurso generado para los trigramas sintácticos, los recursos demostraron generar información valiosa que puede ser tomada en cuenta para la clasificación de la polaridad de los aspectos. Los resultados obtenidos por nuestra propuesta son alentadores considerando que el proceso de construcción de los recursos se realiza de forma automática.

## Referencias

- BACCIANELLA, S., ESULI, A., and SEBASTIANI, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *International Conference on Language Resources and Evaluation (LREC)*, volume 10, pages 2200–2204.
- DEERWESTER, S. C., DUMAIS, S. T., LANDAUER, T. K., FURNAS, G. W., and HARSHMAN, R. A. (1990). Indexing by latent semantic analysis. *JAsIs*, 41(6):391–407.
- EDUARD C. DRAGUT, CLEMENT YU, P. and MENG, W. (2010). Construction of a sentimental word dictionary. *International Conference on Information and Knowledge Management (CIKM) Toronto Ontario Canada*, pages 1761–1764.
- HATZIVASSILOGLOU, V. and MCKEOWN, K. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the Joint ACL/EACL Conference*, pages 174–181.
- HU, M. and LIU, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 168–177.
- KAMPS, J., MARX, M., MOKKEN, R. J., and DE RIJKE, M. (2004). Using wordnet to measure semantic orientations of adjectives. In *International Conference on Language Resources and Evaluation (LREC)*, volume 4, pages 1115–1118.
- MARIE-CATHERINE DE MARNEFFE, B. M. and MANNING, C. D. (2006). Generating typed dependency parses from phrase structure parses. *Proceeding of the fifth International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy, pages 449–454.
- MATA, F. L. C. (2011). *Extracción de Opiniones sobre Características: Un Enfoque Práctico Adaptable al Dominio*. PhD thesis, Departamento de Lenguajes y Sistemas Informáticos Universidad de Sevilla.
- ORTEGA, R., FONSECA, A., MUÑIZ, C., GUTIÉRREZ, Y., and MONTOYO, A. (2014). Using latent semantic analysis to build a domain-dependent sentiment resource. *SemEval 2014*, pages 773–779.

- PADRÓ, L. and STANILOVSKY, E. (2012). Freeling 3.0: Towards wider multilinguality. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012), Istanbul, Turkey, May 23-25, 2012*, pages 2473–2479.
- PONTIKI, M., PAPAGEORGIOU, H., GALANIS, D., ANDROUTSOPOULOS, I., PAVLOPOULOS, J., and MANANDHAR, S. (2014). Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 27–35.
- SIDOROV, G. (2013a). N-gramas sintácticos no-continuos. *Polibits*, 48:69–78.
- SIDOROV, G. (2013b). N-gramas sintácticos y su uso en la lingüística computacional. *Vectores de investigación*, (6):13–27.
- STONE, P., DUNPHY, D. C., SMITH, M. S., and OGILVIE, D. (1968). The general inquirer: A computer approach to content analysis. *Journal of Regional Science*, 8(1):113–116.
- TURNEY, P. D. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia*, pages 417–424.
- XIAODAN ZHU, HONGYU GUO, S. M. and KIRITCHENKO, S. (2014). An empirical study on the effect of negation words on sentiment. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, USA*.