

Un método para el análisis de sentimientos bajo un enfoque supervisado usando recursos léxicos

Antonio Hernández Ambrocio, Gabriela Ramírez de la Rosa,
Esaú Villatoro Tello

Universidad Autónoma Metropolitana (UAM), Unidad Cuajimalpa,
Departamento de Tecnologías de la Información,
México

antonio.hdza12@gmail.com,
{gramirez,evillatoro}@correo.cua.uam.mx

Resumen. La tarea de análisis de sentimientos ha sido un tema de interés desde hace algunos años. Conocer si un texto es positivo o negativo es relevante para tener una opinión general sobre diversas entidades como: productos, personajes públicos, y empresas. En este artículo presentamos un método de clasificación supervisada que toma ventaja de recursos léxicos existentes en español e inglés. Los experimentos realizados están orientados a contestar en qué medida la información externa de recursos léxicos, es útil bajo un enfoque de clasificación supervisada. Los resultados obtenidos muestran que el método propuesto es estable para ambos idiomas y diferentes recursos léxicos empleados.

Palabras clave: recursos léxicos, aprendizaje supervisado, análisis de sentimientos, procesamiento del lenguaje natural.

A Sentiment Analysis Method using a Supervised Approach based on Lexical Resources

Abstract. Sentiment analysis task has been an interest topic in recent years. Knowing if a text is positive or negative is relevant to have a general opinion about different entities such as products, public personalities or companies. In this paper we present a supervised classification method that takes into account existing Spanish or English lexicons. The experiments carried out are aimed at answering how much the external information of lexical resources is useful in a supervised classification approach for sentiment analysis. The obtained results show that the proposed method is stable for both languages and for different lexicons used.

Keywords: lexical resources, supervised learning, sentiment analysis, natural language processing.

1. Introducción

A diario, usuarios activos en la Web generan inmensas cantidades de información en forma de opiniones sobre diversos temas. Estas opiniones son vertidas en medios digitales; por ejemplo en: foros de opinión, blogs, redes sociales, sitios de reseñas o evaluación de productos, sitios de comercio electrónico, entre otros. Un panorama de esta constante actividad de millones de personas al rededor del mundo en redes sociales se puede ver en estadísticas de Facebook y Twitter. Por un lado, Facebook tiene hasta 1.2 billones de usuarios activos al día y 1.9 billones al mes¹. Por otro lado Twitter reporta un promedio de 500 millones de tweets escritos al día². Así mismo, el sitio de comercio electrónico Amazon cuenta con 2.4 billones de visitas mensuales³.

Ante esta creciente cantidad de información es importante contar con métodos automáticos que nos permitan realizar un análisis rápido y eficiente de la información para la toma de decisiones. El área encargada del análisis de este tipo de textos es el Procesamiento del Lenguaje Natural, una sub-área de la Inteligencia Artificial. Así, la tarea del análisis de sentimientos se puede definir como una tarea de clasificación donde a cada texto se le puede asignar una de tres posibles etiquetas: negativa, positiva, y en algunos casos la etiqueta de opinión neutral.

Existen dos principales enfoques de solución para la tarea del análisis de sentimientos: i) enfoques basados en recursos léxicos y ii) enfoques basados en aprendizaje supervisado. En el primer enfoque se necesita de un diccionario de palabras asociadas a un sentimiento (positivo o negativo). Este tipo de diccionarios son compilaciones que capturan conocimiento previo de las palabras que en él aparecen. En el segundo enfoque, aprendizaje supervisado, no se requieren recursos léxicos pero sí un conjunto de ejemplos de opiniones previamente etiquetados.

Obtener ejemplos etiquetados es una tarea costosa, pues usualmente estos ejemplos se etiquetan de forma manual; adicionalmente, es preferible contar con ejemplos específicos para el dominio de clasificación que se necesite. Por ejemplo, si se requiere saber el sentimiento de revisiones de libros, posiblemente los ejemplos etiquetados en este dominio no sean adecuados para determinar el sentimiento hacia un personaje político.

Por otro lado, actualmente se cuentan con diversos recursos léxicos que contienen palabras asociadas a un valor de sentimiento u opinión. Estos recursos pueden ser útiles para clasificar nuevos textos sin importar el dominio. Sin embargo, usar solamente la información de estos recursos léxicos puede no ser adecuado para todos los dominios pues normalmente cada uno tienen formas específicas de expresar opiniones positivas o negativas.

Dado lo anterior, en este artículo se trata de responder a la siguiente pregunta de investigación: ¿en qué medida se puede combinar información de los recursos

¹ All Facebook Statistics In One Place. <https://www.socialbakers.com/statistics/facebook/>

² All Twitter Statistics In One Place. <https://www.socialbakers.com/statistics/twitter/>

³ Web Visitor traffic to Amazon.com. <https://www.statista.com/statistics/623566/web-visits-to-amazoncom/>

léxicos previamente construidos bajo un enfoque de clasificación supervisado en la tarea de análisis de sentimientos?

El resto de este artículo está organizado como sigue: en la sección 2 se hace un repaso de trabajos pertinentes al tema de análisis de sentimientos. En la sección 3 se presenta la idea general del método propuesto. Luego, en la sección 4 se describen los recursos léxicos usados, y una propuesta de estandarización. Los experimentos realizados se presentan en la sección 5 y finalmente en la sección 6 se exponen ideas de trabajo futuro y conclusiones.

2. Trabajo relacionado

Se han realizado numerosos trabajos empleando métodos basados en recursos léxicos, algunos con buenos resultados. Tal es el caso del método que reportó Ding et al. [7] en el 2008, en el que proponen un léxico conformado por *palabras de opinión*. En el método propuesto en ese trabajo se trata de determinar el contexto y reglas semánticas de la oración.

Por otro lado, bajo un enfoque de aprendizaje supervisado, Pang B. et al.[9] experimentaron con tres algoritmos de aprendizaje: Naive Bayes, *Maximum Entropy* y *Support Vector Machines*. Sus resultados sugieren que se puede notar que los clasificadores tienen, en general, una exactitud aceptable.

Adicionalmente, se han propuesto métodos que involucran recursos léxicos y métodos de aprendizaje supervisado. Por ejemplo, Madhavi D. et al.[6] trabajaron con dos enfoques: el primero es la construcción de clasificadores con ensambles que obtienen de la combinación de tres léxicos de cuatro que usan (SentiWordNet, Bing Liu Lexicon, SenticNet, MPQA), haciendo un clasificador por cada léxico, además uno por cada ensamble; el segundo enfoque mencionado consiste en aunar el conocimiento que proporcionan los recursos léxicos a un clasificador. Reportan no haber encontrado una mejora en usar ensambles a usar los léxicos originales, sin embargo, se notó una mejoría al hacer uso del conocimiento proporcionado por los léxicos.

3. Método propuesto

En la figura 1 se presenta el esquema general del método de clasificación propuesto. De manera similar a un enfoque de clasificación supervisada, tiene dos etapas: entrenamiento y validación.

Etapa 1. Entrenamiento usando recursos léxicos. En esta etapa se construye un modelo de clasificación mediante un conjunto de documentos previamente etiquetados. De la misma forma que en el enfoque tradicional, la representación de dichos textos se hace considerando únicamente información de los documentos etiquetados, es decir, una representación vectorial donde cada vector corresponde a un documento en el corpus etiquetado D , de tal forma que cada $d_i \in D$ se define con un vector $\langle w_{ij}, \dots, w_{im} \rangle$, usualmente la dimensión

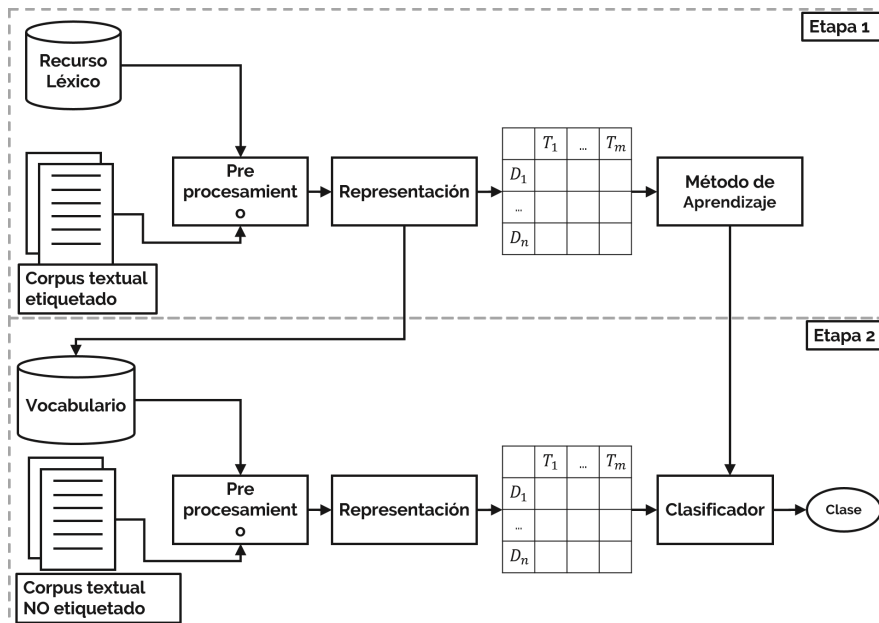


Fig. 1. Esquema general del método propuesto.

del vector corresponde al tamaño de los términos diferentes en el conjunto de documento D . En nuestra propuesta, la dimensión del vector corresponde al conjunto de términos T donde $\forall t \in T \mid t \in (V \cap L)$ siendo V el vocabulario de todos los documentos en D y L el conjunto de términos en un recurso léxico dado. Por lo tanto, una salida parcial de esta etapa es un vocabulario compuesto por el conjunto T . Note que en nuestra propuesta se incorpora una comparación adicional, que es que el término a considerar en la representación también esté en el conjunto de términos del recurso léxico.

Una vez que se tiene la representación vectorial de cada documento etiquetado se utiliza algún algoritmo de aprendizaje, como Naïve Bayes o SVM, para generar el modelo de aprendizaje específico para esta tarea.

Etapa 2. Validación. El propósito de la etapa de validación es usar el mismo conjunto T de términos para representar los documentos a clasificar, luego se utiliza el modelo de aprendizaje construido en la etapa anterior para determinar la clase de cada documento. Cabe mencionar que este proceso es similar al del enfoque tradicional.

4. Recursos léxicos para el análisis de sentimientos

Un recurso léxico es un conjunto de términos en la que cada uno de estos términos está asociado a un sentimiento (positivo o negativo). El sentimiento

puede ser representado con la polaridad, i.e., positivo; o con un valor numérico que refleja la probabilidad afectiva.

En este artículo utilizamos un conjunto de recursos léxicos tanto para el idioma Inglés como para el Español que han sido usados previamente en la literatura [6,10]. En la tabla 1 se muestran cinco recursos léxicos con el número de palabras asociada a cada sentimiento. Los recursos léxicos que se analizaron y posteriormente usaron en la evaluación experimental de la sección 5 son: SentiWordNet [1], SenticNet⁴ y ML-SentiCon [5] para el idioma inglés. Además, para el español se usan: la versión en español de SentiCon [5] y una adaptación de ANEW [3] (Por sus siglas en inglés Affective Norms for English Words) en español.

Tabla 1. Número de palabras positivas y negativas en cinco recursos léxicos: SentiWordNet, SenticNet, ML-SentiCon, ANEW y MS-SentiCon (versión en español).

Recurso léxico	Idioma	Términos positivos	Términos negativos
SentiWordNet	EN	25,508	26,440
SenticNet	EN	27,405	22,595
ML-SentiCon	EN	12,999	12,424
ANEW	ES	511	523
ML-SentiCon	ES	5,568	5,974

La mayoría de los recursos léxicos mostrados en la tabla 1 cuentan con información similar. Sin embargo, cada recurso léxico puede presentar su lista de términos en diferentes formatos, diferente orden, o algunos tienen información adicional como su etiqueta POS (part-of-speech). Con el propósito de homogeneizar estos recursos se presenta una propuesta de estandarización a un formato compuesto por los pares término - polaridad. Esta polaridad se representa con un valor numérico en un rango de -1 a 1, siendo -1 polaridad negativa y 1 positiva. La intención de esta homogeneización es que el método propuesto pueda ser adaptado para cualquier léxico con este formato.

5. Evaluación experimental

En esta sección, mediante una evaluación empírica, se busca evaluar el uso de las palabras de recursos léxicos en el enfoque supervisado de clasificación de sentimientos.

Para todos los experimentos se utilizó la representación vectorial descrita en la sección 3 usando tres diferentes esquemas de pesado: a) booleano, asignando 1 cuando el término t_k aparece en el documento d_i y utilizando 0 en otro caso; b) TF (frecuencia del término), asignando el número de veces que el término t_k aparece en el documento d_i ; y TF-IDF (frecuencia de término - frecuencia inversa del documento), $TF(t_k) \times IDF(t_k)$, donde $IDF(t_k) = \log \frac{|D|}{|\{d_i \in D: t_k \in d_i\}|}$.

⁴ Disponible en: <http://sentic.net>

Para medir el desempeño de cada configuración se usaron las medidas precisión, recuerdo y la medida F. La precisión (P) es la proporción de instancias clasificadas correctamente en una clase c_i con respecto a la cantidad de instancias clasificadas en esa misma clase. El recuerdo (R), la proporción de instancias clasificadas correctamente en una clase c_i con respecto a la cantidad de instancias que realmente pertenecen a esa clase. Usando estas medidas es posible tener evaluación global del sistema de clasificación mediante el cálculo de la medida F, definida como lo indica la Ecuación 1:

$$F = \frac{(1 + \beta^2)P * R}{\beta^2(P + R)}. \quad (1)$$

Adicionalmente, los algoritmos de clasificación usados en los experimentos fueron *Naive Bayes* y el algoritmo *Optimización Mínima Secuencial* (SMO) con la implementación de Weka [8] usando los parámetros por defecto. Finalmente, para todos los experimentos se usó un esquema de validación cruzada a 10 pliegues

En las siguientes sub-secciones se describen las colecciones de datos usados, el pre-procesamiento realizado a estas colecciones de datos y se describen tres experimentos: i) evaluación del uso de los recursos léxicos en el esquema descrito previamente; ii) evaluación de ensambles de recursos léxicos; y iii) evaluación de la utilización del valor afectivo asociado a cada término en algún recurso léxico.

5.1. Colección de datos

Para los experimentos se usaron colecciones de datos en español e inglés. El corpus utilizado para el idioma inglés fue construido por John Blitzer, et al [2] y está compuesto por opiniones vertidas en Amazon de cuatro tipos de productos (dominios): Libros, DVDs, Electrónicos, y Artículos de Hogar. Cada dominio contiene una colección de 2000 opiniones, 1000 por cada clase. En este corpus, las opiniones están asociadas a una puntuación de 0 a 5 estrellas al producto. A partir de esta puntuación se etiquetaron las opiniones en dos clases: las opiniones con una calificación mayor a 3 fueron etiquetadas como positivas y las opiniones con calificación menor a 3 se les etiquetó como negativas.

Por otro lado, el corpus en español cuenta con un total de 3,553 críticas de cine extraídas del sitio *www.muchocine.net* [4]. Cada crítica tiene asociada una calificación que va de 1 a 5. Para separar las críticas, los autores, consideraron como positivas a todas aquellas que contaran con una calificación mayor a 3, y en caso contrario a todas las críticas que tuviesen una calificación menor de 3, se les consideró negativas, el resto fueron ignoradas. Tras la separación de las críticas, según su calificación, el corpus fue reducido a 2,277 críticas, de las cuales 1,147 resultaron ser positivas, y 1130 negativas.

5.2. Pre-procesamiento

Antes de realizar los experimentos se requirió un pre-procesamiento de los documentos de ambas colecciones de datos. De manera general, se eliminaron

signos de puntuación, se convirtieron todas las letras a minúscula, y se hizo un parseo a codificación UTF-8 para evitar pérdida de información.

Dado que los términos utilizados para la representación de los documentos deben estar en un recurso léxico, no se requirió hacer otro tipo de pre-procesamiento a estos documentos.

5.3. Evaluando recursos léxicos en un enfoque supervisado

En este primer experimento se busca determinar el desempeño del modelo de aprendizaje usando una representación basada únicamente en un recurso léxico (de los descritos en la sección 4). Las tablas 2 y 3 muestran los resultados obtenidos para las colecciones en inglés y español, respectivamente.

De la tabla 2 podemos observar que los mejores resultados para tres dominios se obtienen cuando se usa el recurso léxico SentiWordNet usando el algoritmo de aprendizaje SMO. Para el dominio Libros, la representación que usa el léxico SenticNet obtuvo el mejor resultado. Detrás se encuentra SenticNet y por último ML-SentiCon. En todos los casos sobresalió el uso del pesado booleano en la representación.

Para la colección en español, se usaron los recursos léxicos ANEW y ML-SentiCon, los resultados en medida F se muestran en la tabla 3. Respecto a los resultados obtenidos de los recursos en español, el léxico SentiCon fue con el que se obtuvieron los mejores resultados. Note que el clasificador que funcionó mejor fue el construido bajo el algoritmo SMO con pesado booleano con un medida F de 0.707.

De forma general, de este primer experimento podemos concluir que el tamaño del recurso léxico no importa de forma relevante en el desempeño de clasificación. Por ejemplo, para el caso del corpus en español, la diferencia de la cantidad de palabras que contiene ANEW contra ML-SentiCon es 10 veces menor (ver tabla 1) pero el mejor resultado obtenido con ANEW es 0.68 de medida F mientras que con ML-SentiCon es 0.707 de medida F.

5.4. Evaluando la complementariedad de los recursos léxicos

La idea detrás de este segundo experimento es determinar si combinando información proveniente de los recursos léxicos disponibles es posible mejorar el desempeño de clasificación. Para probar la idea se realizaron dos *ensambles* de recursos léxicos. El primer ensamble, llamado intersección, se realizó para tener un léxico conformado por palabras con una alta probabilidad de estar asociadas a un sentimiento, pues al estar presentes en todos los léxicos, nuestra hipótesis es que esas son las palabras polarizadas.

El segundo ensamble, llamado unión, se construyó bajo la idea de que al contener todas las palabras a las que se les ha asociado una carga positiva o negativa, puede apoyar a la clasificación; i.e., se busca tener mayor cobertura de vocabulario. La tabla 4 muestra los resultados en medida F de la clasificación usando los dos ensambles descritos anteriormente. En ambos casos, el valor

Tabla 2. Resultados en medida F de dos algoritmos de clasificación con tres diferentes esquemas de pesado para la colección en Inglés. Se usaron tres diferentes recursos léxicos: SentiWordNet, SenticNet y SentiCon.

SentiWordNet					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive Bayes	BOOL	0.710	0.786	0.778	0.780
	TF	0.680	0.696	0.711	0.710
	TF-IDF	0.683	0.696	0.711	0.710
SMO	BOOL	0.735	0.783	0.782	0.802
	TF	0.728	0.749	0.746	0.770
	TF-IDF	0.728	0.750	0.746	0.770
SenticNet					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive Bayes	BOOL	0.736	0.756	0.770	0.776
	TF	0.766	0.681	0.71	0.691
	TF-IDF	0.685	0.682	0.711	0.691
SMO	BOOL	0.744	0.786	0.769	0.784
	TF	0.722	0.727	0.739	76.7
	TF-IDF	0.722	0.728	0.74	0.767
SentiCon					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive Bayes	BOOL	0.699	0.763	0.735	0.780
	TF	0.676	0.656	0.667	0.667
	TF-IDF	0.676	0.659	0.665	0.668
SMO	BOOL	0.737	0.739	0.755	0.757
	TF	0.685	0.705	0.735	0.726
	TF-IDF	0.685	0.705	0.73.5	0.726

afectivo fue modificado al promedio de la suma de los valores en los léxicos involucrados. Nótese que este valor de sentimiento asociado a las palabras en los ensambles no se utiliza en este experimento.

De acuerdo con los resultados mostrados en la tabla 4, los mejores resultados fueron ligeramente menores a los que se obtuvieron en el primer experimento, usando únicamente el recurso léxico SentiWordNet. Se puede observar que el mejor desempeño fue de 0.79 de medida F y se obtuvo bajo el dominio Artículos de hogar con el algoritmo SMO en ambos casos.

Para el idioma español, se hicieron ambos ensambles: intersección y unión, con los dos recursos disponibles para este idioma. La tabla 5 muestra los resultados obtenidos para la colección en español. Se puede observar que con el ensamble unión se llegaron a resultados similares, aunque ligeramente mejores, que el mejor resultado del primer experimento. Se observa que, en el mejor caso, aquí se obtuvo 0.712 contra 0.707 del primer experimento, ambos bajo el algoritmo Naive Bayes con un esquema de pesado booleano. Por otro lado, con

Tabla 3. Resultados en medida F de dos algoritmos de clasificación con tres diferentes esquemas de pesado para la colección en Español. Se usaron tres diferentes recursos léxicos: ANEW y ML-SentiCon

		Anew	ML-SentiCon
Naive Bayes	BOOL	0.628	0.685
	TF	0.682	0.656
	TF-IDF	0.580	0.658
SMO	BOOL	0.643	0.707
	TF	0.632	0.694
	TF-IDF	0.633	0.694

Tabla 4. Resultados en medida F de dos algoritmos de clasificación con tres diferentes esquemas de pesado para la colección en Inglés. Se usaron dos ensambles de recursos léxicos: intersección y unión.

Intersección					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive Bayes	BOOL	0.709	0.764	0.719	0.750
	TF	0.670	0.643	0.645	0.644
	TF IDF	0.670	0.643	0.644	0.644
SMO	BOOL	0.733	0.748	0.749	0.775
	TF	0.688	0.710	0.737	0.739
	TF IDF	0.681	0.710	0.737	0.739
Unión					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive Bayes	BOOL	0.727	0.792	0.781	0.791
	TF	0.690	0.711	0.722	0.717
	TF IDF	0.689	0.711	0.721	0.717
SMO	BOOL	0.756	0.774	0.787	0.791
	TF	0.741	0.757	0.759	0.772
	TF IDF	0.741	0.757	0.759	0.772

el ensamble procedente de la intersección se llegó a, básicamente, los mismos resultados que en el experimento uno.

5.5. Evaluando la importancia de la carga afectiva de los recursos léxicos

Hasta el momento sólo se han usado los recursos léxicos para obtener el vocabulario a utilizar en la representación de los documentos. La idea detrás de este tercer experimento es determinar si la incorporación del valor afectivo que está asociada a cada palabra de cada recurso léxico es útil para mejorar el desempeño de la clasificación.

Para usar este valor afectivo se hicieron dos propuestas, la primera, multiplicar el valor booleano por el valor afectivo; la segunda, multiplicar la frecuencia de

Tabla 5. Resultados en medida F de dos algoritmos de clasificación con tres diferentes esquemas de pesado para la colección en Español. Se usaron dos ensambles de recursos léxicos: intersección y unión.

		Intersección Unión	
Naive Bayes	BOOL	0.685	0.694
	TF	0.656	0.673
	TF-IDF	0.658	0.672
SMO	BOOL	0.707	0.712
	TF	0.694	0.697
	TF-IDF	0.694	0.697

término por el valor en el léxico. En este experimento sólo se usaron los léxicos que obtuvieron los mejores resultados en los experimentos previos, i.e., SentiWordNet para inglés y el ensamble intersección para el español. Los resultados del desempeño del método de clasificación se muestran en la tabla 6.

Tabla 6. Resultados en medida F de dos algoritmos de clasificación con dos diferentes esquemas de pesado para la colección en Inglés. Se usó el recurso léxico SentiWordNet.

Sentiwordnet (Valor Afectivo)					
		Libros	DVD	Electrónicos	Artículos de hogar
Naive	BOOL	0.717	0.783	0.780	0.779
Bayes	TF	0.679	0.703	0.726	0.720
SMO	BOOL	0.738	0.770	0.769	0.796
	TF	0.73	0.748	0.749	0.777

Para la colección en inglés el mayor medida F es 0.796 contra 0.800 de la evaluación de SentiWordNet (ver Tabla 2), también bajo el dominio artículos de hogar con el algoritmo SMO y con el esquema de pesado que resultó de multiplicar el valor afectivo por el valor booleano. Para el caso de la colección en español se obtuvieron básicamente los mismos resultados que en experimentos previos (ver Tabla 7).

Tabla 7. Resultados en medida F de dos algoritmos de clasificación con dos diferentes esquemas de pesado para la colección en español. Se usó el recurso léxico ensamble intersección.

Ensamble (Valor Afectivo)		
Cine		
Naive	BOOL	0.685
Bayes	TF	0.664
SMO	BOOL	0.707
	TF	0.695

En general, los resultados producidos en los experimentos realizados se pueden considerar convenientes para el análisis de sentimientos. Entre los algoritmos utilizados se pueden apreciar mejores resultados de SMO sobre Naive Bayes, y por otro lado, de los esquemas de pesado propuestos, destaca booleano, incluido cuando se involucra el valor afectivo en los léxicos. Aunque, en la mayoría de los experimentos se llegaron a resultados muy parecidos. Por lo tanto, de estos experimentos se puede concluir que no importa ni el valor afectivo de las palabras ni el pesado diferente al booleano, es decir, es suficiente con indicar que una palabra del recurso léxico está o no presente en el documento. En la figura 2 y 3 se muestra un resumen de los mejores resultados obtenidos por los tres experimentos, indicando además valores de medida F por cada clase. Como puede observarse, los resultados de clasificación para ambas clases son similares por lo tanto las diferencias entre precisión, recuerdo y medida F son consistentes entre colecciones de datos.

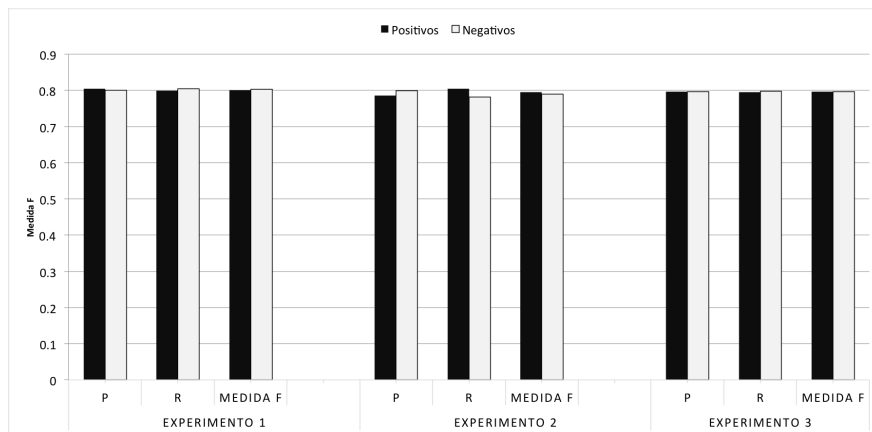


Fig. 2. Resumen de los mejores resultados de clasificación obtenidos para la colección en Inglés.

Otra conclusión importante es que el recurso léxico SentiWordNet es el más adecuado para diferentes dominios de revisiones de productos. Por otro lado, el ensamble intersección para español resultó ser mejor la clasificación de revisiones de películas. Es necesario realizar más experimentos con otras colecciones de datos para poder llegar a conclusiones más generales.

6. Conclusiones y trabajo a futuro

En este artículo se desarrolló un método de clasificación supervisado usando recursos léxicos para la tarea del análisis de sentimientos. La idea general de nuestro método es utilizar como representación el vocabulario de los documentos etiquetados que aparecen en ciertos recursos léxicos.

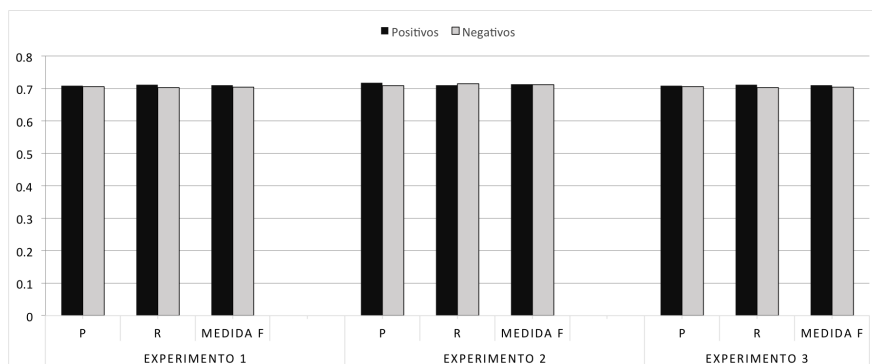


Fig. 3. Resumen de los mejores resultados de clasificación obtenidos para la colección en Español.

Además de evaluar diferentes recursos léxicos en diferentes idiomas, se trataron dos nuevas propuestas de recursos léxicos. La primera en la que se plantea dos ensambles que resultaron de la unión y la intersección de los léxicos disponibles. La segunda propuesta es utilizar el valor afectivo de los recursos léxicos como parte del pesado en la representación vectorial.

Con la primera propuesta de los ensambles de recursos léxicos se tenía la idea de que al tener mayor cobertura en los documentos se podrían obtener resultados más precisos, sin embargo, no resultó ser así en todos los casos, incluso se manifestaron resultados inferiores. De manera general, los resultados conseguidos no mostraron una diferencia significativa a los obtenidos de los clasificadores construidos con los léxicos. Se observa que se pueden llegar al mismo o, ligeramente, mejor rendimiento si se usan sólo los léxicos con un vocabulario de menor extensión.

Por otro lado, en la segunda propuesta se intentó comprobar cuanto afecta el valor afectivo de las palabras. Se llegó a resultados muy parecidos a los mostrados en el resto de experimentos. A partir de esto, se puede concluir que la información en los léxicos no afectan de forma importante.

Finalmente, se percata que las diferentes propuestas tienen un grado de fiabilidad bastante similar, también puede verse que en todos los experimentos se llegó a tener un rendimiento efectivo.

Como trabajo futuro se pretende primero hacer una evaluación de significancia estadística con los resultados obtenidos en este artículo. Además se piensan replicar el experimentos con otros recursos léxicos y otras colecciones de datos.

Agradecimientos. El trabajo de los dos últimos autores fue parcialmente financiado por el proyecto CONACyT CB-2015 No. 258588. También agradecemos al Departamento de Tecnologías de la Información de la Universidad Autónoma Metropolitana Unidad Cuajimalpa por el apoyo otorgado para la realización de este trabajo.

Referencias

1. Baccianella, S., Esuli, A., Sebastiani, F.: Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: LREC. vol. 10, pp. 2200–2204 (2010)
2. Blitzer, J., Dredze, M., Pereira, F., et al.: Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In: ACL. vol. 7, pp. 440–447 (2007)
3. Bradley, M.M., Lang, P.J.: Affective norms for english words (anew): Instruction manual and affective ratings. Tech. rep. (1999)
4. Cruz, F., Troyano, J., Enriquez, F., Ortega, J.: Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento del lenguaje Natural* 41, 73–80 (2008)
5. Cruz Mata, F., Troyano Jiménez, J.A., Pontes Balanza, B., Ortega Rodríguez, F.J.: MI-senticon: un lexicón multilingüe de polaridades semánticas a nivel de lemas (2014-09)
6. Devaraj, M., Piriyani, R., Singh, V.K.: Lexicon ensemble and lexicon pooling for sentiment polarity detection. *IETE Technical Review* 33(3), 332–340 (2016), <https://doi.org/10.1080/02564602.2015.1073572>
7. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: Proceedings of the 2008 international conference on web search and data mining. pp. 231–240. ACM (2008)
8. Garner, S.R.: Weka: The waikato environment for knowledge analysis. In: In Proc. of the New Zealand Computer Science Research Students Conference. pp. 57–64 (1995)
9. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. pp. 79–86. Association for Computational Linguistics (2002)
10. Real-Flores, G., García-Mendoza, B., Calderón-Casanova, E., de-la Rosa, G.R., Villatoro-Tello, E.: Generación y enriquecimiento automático de recursos léxicos para el análisis de sentimientos. In: *Research in Computing Science*. vol. 134, pp. 163–176 (2017)