# Método no supervisado para la clasificación de polaridad en Twitter

## Unsupervised method for polarity classification in Twitter

J.M. Yero Moreno[1] and R. Ortega Bueno[2]

[1]Departamento de Computación, Universidad de Oriente, Santiago de Cuba, Cuba, jose.yero@csd.uo.edu.cu
[2]Centro de Estudios de Reconocimiento de Patrones y Minería de Datos, Santiago de Cuba, Cuba, reynier.ortega@cerpamid.co.cu

*Abstract*— **This paper describes the specifications and results of an unsupervised system for polarity classification of Twitter´s messages. The proposal system includes three phases: data preprocessing, word polarity detection and message classification. The preprocessing phase comprises treatment of emoticon, slang, lemmatization and POS-tagging. The words polarity is detected by the use of a sentiment lexicon, combining four sentiment lexicons to increase the performance of the new one. Finally the overall polarity is determined using a rule based classifier. The results obtained in a Twitter data set, are good considering that our proposal are unsupervised and outperform the results of many supervised method in the literature.**

*Keywords*— **sentiment analysis, opinion mining, polarity classification, lexicon generation, Twitter**

### I. Introducción

La explosión de la Web 2.0 ha marcado una nueva era para la humanidad. El creciente uso de las redes sociales como Facebook, MySpace, LinkedIn y Twitter, ofrecen un espacio de intercambio de información en tiempo real para las personas. Twitter es una de las redes sociales más populares, por lo que ha estado creciendo a pasos agigantados. El número de usuarios activos excede los 645 millones y el promedio de tweets escritos diariamente supera los 58 millones (hasta el 31 de marzo del 2014). Con el uso de las aplicaciones de Twitter, los usuarios intercambian opiniones acerca de política, personalidades, productos, compañías, eventos, etc. Esto ha despertado la atención de diferentes comunidades de científicas, interesadas en el análisis del contenido de estos textos, motivando el desarrollo de muchas tareas de Procesamiento de Lenguaje Natural, como son el Análisis de Sentimientos, la Detección de Emociones, la Recuperación de Opiniones y el Resumen de Opiniones.

Una de las tareas más populares del Análisis de Sentimientos es la clasificación de polaridad, cuya principal función es la clasificación los textos de opinión en positivos, negativos o neutros [1,2,3]. Aproximaciones no supervisadas en la literatura científica [4,5], en los últimos años, muestran el empleo efectivo de recursos lingüísticos y lexicones afectivos con polaridades anotadas como SentiWordNet [6]. Los trabajos desarrollados en su mayoría, realizan un preprocesamiento de los mensajes, donde se tratan las inconsistencias del texto [7]. Otros métodos determinan la polaridad de las palabras, según el contexto en que son empleadas [8]. En la clasificación frecuentemente se emplean clasificadores basados en reglas, donde estas determinan la polaridad global del mensaje de Twitter. Actualmente la mayoría de las aproximaciones presentadas en la literatura científica son supervisadas, dependientes de los datos de entrenamiento, difíciles de adaptar a otros dominios y dependientes del idioma. Por lo que resulta necesario poder desarrollar métodos no supervisados (basados en conocimiento) para la clasificación de la polaridad, que sean independientes del dominio y puedan ser fácilmente extendidos a otros idiomas (sobre todo inglés y español que son los idiomas más populares en Twitter).

### II. Materiales y métodos

En este trabajo se presenta una estrategia no supervisada, para determinar la polaridad de los términos del mensaje a partir de la polaridad anotada para el término en el lexicón sentimental. Este contiene los términos clasificados en muy positivos, muy negativos, positivos y negativos, según en sentimiento que expresan. En la Fig. 1 se muestra la arquitectura general de nuestro clasificador sentimental.

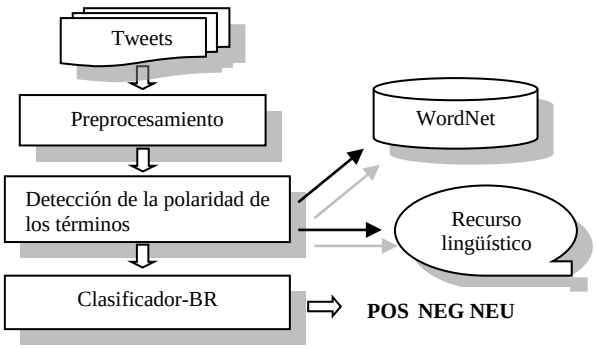


Fig. 1 Arquitectura general del clasificador sentimental.

Como primer paso en la metodología es realizado un preprocesamiento de los mensajes de Twitter, eliminando toda inconsistencia y sustituyendo términos en lenguaje informal por sus respectivos términos formales dentro del lenguaje, se lematizan los términos, se etiquetan según su parte del discurso y se eliminan las palabras sin carga semántica (stopwords). Se realiza la detección de la polaridad de las palabras mediante el empleo del lexicón de polaridad y finalmente se procede a la clasificación del mensaje por un clasificador basado en reglas. Se crearon para el tratamiento de los mensajes diferentes recursos lingüísticos, como lexicones de términos de negaciones, diccionarios de emoticones y de lenguaje informal empleado en internet. Se desarrolló un nuevo y efectivo recurso de polaridad, para los idiomas español e inglés, a partir de otros recursos desarrollados por la comunidad científica internacional y el Centro de Reconocimiento de Patrones y Minería de Datos (CERPAMID). Se diseñó además un clasificador basado en reglas, donde se determinó emplear los parámetros de polaridad con mejores resultados dentro del estado del arte para este tipo de clasificador.

*A. Preprocesamiento de los datos*

Los mensajes de los tweets difieren de los textos de libros, artículos, etc. Con un límite de 140 caracteres, los mensajes muchas veces incluyen en su texto diversas informalidades introducidas por el autor como emoticones, lenguaje informal o jerga de internet, errores ortográficos, URLs, "RT" de re-tweets, menciones de usuarios (@usuario) , hashtags (#hashtag) y repeticiones de caracteres. Por ello es necesario el preprocesamiento del texto, en vista de retirar la mayor cantidad de información inconsistente del texto.

La fase de preprocesamiento comprende los siguientes pasos. El texto es tokenizado, las URLs, re-tweets y menciones a usuarios son removidas. Los tokens hashtags usualmente contienen información relevante acerca del tema sobre el que trata el tweet, por ello no son removidos. Se remplaza los emoticones por una palabra predefinida que expresa su polaridad mediante un diccionario de emoticones, obtenido de Wikipedia. Cada emoticón es manualmente anotado con una palabra sentimental asociada según el sentimiento que expresa, ej. Los emoticones que sugieren emociones con polaridad positiva ":)", ":D", son anotados con la palabra emocional "*happy*" o "*feliz*" y los emoticones que expresan una polaridad negativa, ":( ", ":,(", son anotados con la palabra emocional *"sad"* o *"triste"*. La presencia de abreviaturas es detectada y sustituidas por su significado (ej. Tqm-te quiero mucho), empleando un diccionario. Finalmente el texto es etiquetado morfológicamente y las palabras sin carga semántica se descartan del texto.

*B. Detección de la polaridad de las palabras*

La polaridad de las palabras del mensaje es detectada mediante el uso de un lexicón de polaridad, el cual incluye términos clasificados en muy positivos (HP), muy negativos (HN), positivos (P), negativos (N). El recurso de polaridad empleado fue construido a partir de la unión de varios recursos de polaridad, los cuales se corrigieron y expandieron detectando los términos incorrectamente clasificados y añadiendo además 35 nuevos términos. En la creación del lexicón de polaridad se emplearon los lexicones de polaridad SentiWordNet 3.0 [6], Micro-WordNetOp [9] y JRC_Tonality [10], los dos primeros creados a partir de WordNet [11], lo cual permite la extensión del recurso a otros idiomas de EuroWordNet [12] mediante el índice interlingua. Las versiones empleadas en idioma español de los lexicones WordNetOp y JRC_Tonality fueron traducidas manualmente por el Centro de Estudios de Reconocimiento de Patrones y Minería de Datos (CERPAMID).

En la Tabla 1 se muestra la cantidad de términos por polaridad para el lexicón creado, en sus versiones en español e inglés.

Tabla 1 Cantidad de términos del lexicón de polaridad.

| Idioma | HP | HN | P | N |
|---|---|---|---|---|
| Español | 867 | 2781 | 3126 | 5645 |
| Inglés | 1316 | 2008 | 4095 | 6040 |

*C. Clasificador basado en reglas*

Se emplea un clasificador basado en reglas para clasificar los tweets en positivos, negativos y neutrales. Para ello se asocia un valor de polaridad a cada palabra, basada en la ecuación 1 y a partir de la clase de polaridad determinada en el lexicón de polaridad. Es importante señalar que la polaridad de una palabra puede ser negada mediante un modificador de polaridad (obtenidos de la categoría de negaciones en el General Inquirer [13]).

$$polaridad(t) = \begin{cases} 4 \text{ si } t \text{ es clasificada como } HP \\ -4 \text{ si } t \text{ es clasificada como } HN \\ 2 \text{ si } t \text{ es clasificada como } P \\ -2 \text{ si } t \text{ es clasificada como } N \\ 0 \text{ si } t \text{ es clasificada como } O \end{cases} \quad (1)$$

La polaridad del tweet es obtenida a partir de las polaridades de las palabras que contiene. Para ello se obtiene los valores de polaridad globales, mediante la suma de los valores asociados a las polaridades de cada palabra clasificada. Los valores totales de polaridad positiva *PosScore(t)* y negativa *NegScore(t)* son calculadas de la siguiente forma:

$$PosScore(t) = \sum_{w_i \in W_p} polaridad(w_i) \quad (2)$$

$W_p$: Palabras clasificadas como HP y P en el tweet *t*

$$NegScore(t) = \sum_{w_i \in W_n} polaridad(w_i) \quad (3)$$

$W_n$: Palabras clasificadas como HN y N en el tweet *t*

Si el valor de *PosScore(t)* es mayor que el valor absoluto de *NegScore(t)*, entonces el mensaje es clasificado como positivo. Si el valor de *PosScore(t)* es menor que el valor absoluto de *NegScore(t)* entonces el mensaje es clasificado como negativo. Finalmente si el valor *PostScore(t)* es igual al valor absoluto de *NegScore(t)* es considerado el tweet como neutral.

*D. Ejemplo de la clasificación de un tweet*

Consideremos el siguiente tweet:

*@joseym90: Ibiza island :-), I'm going to an EXTREME party in that place, cooool.*

Aplicando la fase de preprocesamiento el tweet queda normalizado de la siguiente forma:

*Ibiza island "happy", I am going to an extreme party in that place, cool.*

Cuando el texto es lematizado, las stopwords son removidas y los signos de puntuación retirados, obtenemos es siguiente grupo de términos (para cada palabra mostramos: lema y parte del discurso).

*Ibiza#n island#n happy#a go#v extreme#a party#n place#n cool#a*

Obtenemos mediante la ecuación 1 las polaridades para cada una de ellas. Las polaridades asignadas para cada palabra son las siguientes:

*Ibiza#O#0 island#O#0 "happy"#HP#4 go#O#0 extreme#O#0 party#P#2 place#O#0 cool#P#2*

A continuación calculamos los valores totales de las polaridades positivas y negativas:

$$PosScore(t)=4+2+2=8$$

$$NegScore(t)=0$$

Luego, el tweet es clasificado como positivo.

## III. Resultados y discusión

Para el análisis de los resultados de nuestro método se decide evaluar su desempeño en dos colecciones de referencia internacional, destinadas al análisis de sentimientos en Twitter. La primera colección de prueba fue provista para el taller TASS 2013, perteneciente al congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) [4], este corpus contiene 60798 tweets escritos en idioma español. La segunda colección de prueba pertenece a la tarea de Análisis de Sentimientos en Twitter (tarea 2) del Congreso Internacional de Evaluación Semántica (SemEval 2013) [5]. Este corpus contiene 3813 tweets escritos en idioma inglés. En la Tabla 2 se presentan los resultados obtenidos en la clasificación de los mensajes de la colección de prueba del TASS, donde se puede comparar el efecto en la clasificación de ignorar las características que usualmente aparecen en el texto de un tweet con el tratamiento de todas.

Tabla 2 Resultados de la clasificación en el tratamiento o no de las características del tweet.

| Característica (ignorada) | F1 positivos | F1 negativos | F1 neutrales | Promedio F1 |
|---|---|---|---|---|
| Ninguna | 0.6405 | 0.4723 | 0.3713 | 0.5564 |
| Caracteres repetidos | 0.6395 | 0.4732 | 0.3695 | 0.5563 |
| Lenguaje vulgar | 0.6376 | 0.4829 | 0.3834 | 0.5602 |
| Emoticones | 0.6356 | 0.4653 | 0.3767 | 0.5505 |

Como se puede apreciar, al ignorarse el tratamiento de las repeticiones de caracteres y los emoticones en la fase de preprocesamiento, el desempeño en la clasificación disminuye. En el caso del lenguaje vulgar, al ignorarse su tratamiento, se incrementa la precisión; esto es debido al ruido introducido al expandir estos términos por un conjunto de términos que expresen su significado formal. En numerosas ocasiones el término o la frase en la que se expande el término del lenguaje vulgar contienen una polaridad asociada. En la mayoría de los casos estudiados, esta polaridad inducida no es la deseada por el usuario al escribir el mensaje, lo que conlleva a un error en la clasificación. Ejemplo de ello sería la abreviatura "*omg"* correspondiente a la frase "*Oh my God"* en el lenguaje informal de internet. Esta abreviatura es empleada mayormente en la expresión de una emoción de asombro. Hemos determinado que, esta frase contiene un alto nivel de ambigüedad en términos de polaridad. En los diferentes contextos analizados, su empleo estuvo orientado tanto a transmitir una polaridad negativa como positiva. En el análisis desde la perspectiva del método, la polaridad de la frase es determinada como positiva, ya que el término *God* está definido en el lexicón como positivo.

Mediante los resultados obtenidos sobre las colecciones de datos, se ha podido comparar nuestro método con otros métodos de clasificación de polaridad, participantes en los congresos antes mencionados. En la Tabla 3 se muestran los resultados obtenidos sobre la colección de prueba del SemEval 2013 y TASS 2013 de los métodos participantes en estas competiciones.

Tabla 3 Resultados del método para la colección de prueba del SemEval 2013 y TASS 2013.

| Sistema de SemEval 2013 | F1-macro | Sistema de TASS 2013 | F1-micro |
|---|---|---|---|
| SAIL | 60.14 | UPV | 0.674 |
| UT-DB | 59.87 | CITIUS-Cilenis | 0.668 |
| FBK-irst | 59.76 | DLSI-UA | 0.663 |
| nlp.cs.aueb.gr | 58.91 | JRC | 0.612 |
| UNITOR | 58.27 | **Método Propuesto** | **0.593** |
| Umigon | 57.14 | ITA | 0.543 |
| NILC USP | 56.31 | TECNALIA-UNED | 0.496 |
| **Método Propuesto** | **55.64** | UNED-JRM | 0.496 |
| DataMining | 55.52 | UNED-LSI | 0.479 |
| ASVUniOfLeipzig | 54.56 | **ETH-Zurich** | **0.466** |
| **SSA-UO** | **50.17** | **SINAI-EMML** | **0.409** |

Como se puede apreciar en la Tabla 3 los resultados del método propuesto superan los resultados de todos los métodos no supervisados participantes en las competiciones (marcados en negrita) con puntuación macro-F1 sobre el corpus de SemEval 2013 de 55.64 y sobre el corpus del TASS 2013 se obtuvo una puntuación micro-F1 de 0.593. Los resultados alcanzados superan además muchos de los sistemas supervisados participantes.

## IV. Conclusiones

Con el desarrollo de este trabajo hemos definido las especificaciones de un método no supervisado para la clasificación de polaridad en mensajes de Twitter. Se ha logrado diseñar un método capaz de adaptarse a cualquier contexto y analizar mensajes de Twitter en los idiomas español e inglés. Se construyó un nuevo recurso de polaridad que permite el análisis en los idiomas español e inglés, a partir de otros recursos disponibles, mejorando significativamente los resultados en la clasificación. Los resultados obtenidos superan los métodos no supervisados de los que se tiene referencia y muchos de los métodos supervisados presentes en el estado del arte. Consideramos nuestros resultados como alentadores, tomando en consideración la dificultad de tratar con información subjetiva, la naturaleza inconsistente de los mensajes del Twitter y la estrategia desarrollada considerada como no supervisada.

## Referencias


1. Pang, B., & Lee, L. (2002). Thumbs up? sentiment classification using machine learning techniques. (pp. 79-86). In Proceeding of Empirical Methods in Natural Language Processing.
2. Turney, P. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. 417–424.
3. Willson, T., & Wiebe, J. (2006). Recognizing strong and weak opinion clauses. *22*, pp. 73-99. Computational Intelligence.
4. Esteban, A., Alegría, I., & Villena Román, J. (Eds.). (2013). Proceedings of the TASS workshop at SEPLN 2013. Actas del XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural. IV Congreso Español de Informática. Madrid, Spain.
5. Nakov, P., Rosenthal, S., Kozareva, Z., Stoyanov, V., Ritter, A., & Wilson, T. (2013). SemEval-2013 Task 2: Sentiment Analysis in Twitter. *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (págs. 312-320). Association for Computational Linguistics.
6. Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, et al. (Ed.), *LREC*. European Language Resources Association.
7. Ortega-Bueno, R., Fonseca-Bruzón, A., Gutiérrez, Y., & Montoyo, A. (2013). SSA-UO: Unsupervised Twitter Sentiment Analysis. Second Joint Conference on Lexical and Computational Semantics (*SEM). Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 501-507. Atlanta, Georgia, USA: Association for Computational Linguistics.
8. Martínez-Cámara, E., García-Cumbreras, M. A., Martín-Valdivia, M., & Ureña López, L. A. (2013). SINAI-EMML: Combinación de Recursos Lingüísticos para el Análisis de la Opinión en Twitter. In Proc. of the TASS workshop at SEPLN 2013 (págs. 187-194). 17-20 de Septiembre 2013: IV Congreso Español de Informática.
9. Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., & Gandini, G. (2007). Language resources and linguistic theory: Typology, second language acquisition, English linguistics (Forthcoming), chapter Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. Milano, IT: Franco Angeli Editore.
10. Balahur, A., Steinberger, R., der Goot, E. V., Pouliquen, B. & Kabadjov, M. A. (2009). Opinion Mining on Newspaper Quotations. *Web Intelligence/IAT Workshops* (p./pp. 523-526), : IEEE.
11. Fellbaum, C. (1998). WordNet: an electronic lexical database. MIT Press.
12. Vossen , P. (1998). EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers.
13. Stone, P. J. (1966). The General Inquirer: A Computer Approach to Content Analysis. The MIT Press.