ISSN 1870-4069

Una propuesta de clasificación automática de polaridad para notas periodísticas

José A. Baez, Orlando Ramos, María J. Somodevilla, Ivo H. Pineda, Darnes Vilariño, Concepción Pérez de Celis

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, México

icc_bagatella@outlook.com, {orlandxrf, mariajsomodevilla, ivopinedatorres, dvilariñoayala, mcpcelish}@gmail.com

Resumen. En este trabajo una técnica para determinar la subjetividad escrita en lenguaje natural es presentada, en particular la polaridad de en conjunto de noticias periodísticas. Un algoritmo cuantitativo para la clasificación automática depolaridad basado en un árbol de búsqueda binaria es propuesto. En dicho algoritmo se utiliza TF*IDF para obtener las palabras con mayor peso en las noticias,ademas de un lexicón de palabras positivas y negativas como soporte para crear el modelo de clasificación. Se realizaron experimentos utilizando y sin utilizar el algoritmo propuesto como preprocesamiento. Los resultados de la clasificación de polaridad con preprocesamiento fueron satisfactorios independientemente del número de noticias del corpus de prueba.

Palabras clave: Clasificación, aprendizaje automático, análisis de sentimientos, minería de opinión, polaridad, noticias.

A Proposal for Automatic Classification of News Reports Polarity

Abstract. This paper presents a technique for determining the subjectivity written in natural language, in particular the polarity of news reports. A quantitative algorithm for automatic polarity classification, based on a binary search tree, is proposed. TF*IDF is applied in order to determine the words more weight in the news; besides a lexicon of positive and negative words, as a support for building the classification model, is also consulted. Experiments were performed using and without using the proposed algorithm as preprocessing. The classification results of preprocessing polarity were satisfactory regardless of the number of news in the test corpus.

Keywords. Classification, machine learning, sentiment analysis, opinion mining, polarity, news.

1. Introducción

Algunos años atrás, cuando se deseaba conocer la opinión escrita en lenguaje natural de las personas sobre algún tema o servicio específico, las empresas especializadas en esta tarea, se dedicaban a la ardua labor de realizar encuestas, claramente definidas y orientadas a los temas de los cuales se deseaba conocer la opinión de las personas (usuarios), esto claro de forma manual. El paso siguiente está a cargo de un equipo especializado de personas, encargado de analizar de forma manual el contenido de los datos recabados, para así determinar la opinión de los usuarios.

Actualmente el Análisis de Sentimientos (AS) o Minería de Opinión (MO) es uno de los tópicos recientes y más estudiados del Procesamiento de Lenguaje Natural (PLN), esto para determinar de manera automática o semi-automática, la polaridad de un texto en lenguaje natural de un determinado autor (usuario) sobre alguna temática específica. Las aplicaciones sobre esta tarea son diversas, por ejemplo conocer la opinión de los usuarios de Twitter sobre alguna persona (político, artista), empresa, gobierno, campañas sobre productos, noticias, servicios, etc.

Dada la problemática planteada en los párrafos anteriores, en este trabajo se presenta un método para clasificar noticias de manera automática, con ayuda de técnicas, algoritmos de PLN y recursos externos, como lexicones de palabras positivas y negativas, para asignar la polaridad correspondiente a cada noticia.

La estructura del artículo es como sigue, en la segunda sección se describe el estado del arte relacionado con el AS y MO, la tercera sección presenta los conocimientos fundamentales referentes al aprendizaje automático y clasificación. El planteamiento del problema y el conjunto de datos usado es presentado en la cuarta sección, en la quinta sección presentamos nuestra propuesta, los resultados obtenidos se describen en la sexta sección y en la séptima sección las conclusiones.

2. Trabajos relacionados

En esta sección abordaremos los trabajos relacionados con el análisis de sentimientos y la minería de opinión. El trabajo [10] del SemEval 2015 en la tarea 12, utilizó un clasificador de aprendizaje supervisado automático para predecir cada polaridad de opinión (positiva, negativa y neutra). El clasificador se usa en combianción con un proceso de selección basado en la probabilidad para entidades y la detección de atributos de la categoría, teniendo una bolsa de palabras, lemas, bigramas después de verbos y un lexicón basado en características, alcanzado buenos resultados en el dominio de laptops y restaurantes.

Técnicas sobre análisis de sentimientos y minería de opinión se presentan en [5], los enfoques más populares son utilizar lexicón subjetivo, modelos de ngramas y aprendizaje automático. En cambio las técnicas que proponen en el proceso del análisis de sentimientos para textos son: generar un lexicón

(para extraer el conocimiento de los sentimientos), detectar la subjetividad (clasificar el texto en nivel de su naturaleza subjetiva y objetiva), detección de polaridad de sentimientos (la clasificación de los sentimientos en clases semánticas), estructuración de sentimientos (basada en las 5W: why, where, when, what, who) y resumen , visualización y seguimiento de sentimientos (la visualización es generada de manera prudente en un grafo de acuerdo a una dimensión o combinación de dimensiones).

En [1] identificaron tres subtareas: definición del objetivo, separación de las buenas y malas noticias sobre el contenido de sentimientos buenos y malos expresados en el objetivo, y el análisis de opinión claramente marcado que se expresa de forma explícita, sin necesidad de interpretación. El conjunto de datos que utilizan proviene de las aplicaciones NewsBrief[1] y MedISys[2] de EMM[3] . En sus experimentos utilizaron WordNet [11], SentiWordNet [3], MicroWNOp [2].

Los experimentos realizados en los trabajos anteriores utilizaron diferentes ventanas alrededor del objetivo, mediante el cálculo de una puntuación de las palabras de opinión identificadas y la eliminación de las palabras que estaban en las mismas palabras de opinión, de tiempo y palabras categoría. Además, se utilizó un recurso incorporado de las palabras de opinión con la polaridad asociada, que denotaron como Tonalidad JRC. Cada uno de los recursos empleados fue mapeado en cuatro categorías, que fueron dadas con puntuaciones diferentes: positivo (1), negativo (-1), positivo alto (4) y alta negativo (-4). Los mejores resultados fueron obtenidos con la combinación de la Tonalidad JRC y MicroWN, en una ventana de 6 palabras.

La tarea para determinar la polaridad de un texto, es una tarea complicada, más aún cuando el texto contiene palabras positivas y negativas en la misma oración. El problema ya no resulta trivial, es decir; se tiene que analizar el contexto que quiere expresar el autor. Por ejemplo, una palabra negativa seguida de una positiva: *el alcalde rechazó darles audiencia*, si solamente se analiza un conteo de palabras positivas y negativas, en el ejemplo sería: una positiva y una negativa. Sin embargo se debe categorizar su polaridad como negativa teniendo en cuenta el contexto. El método propuesto en este trabajo incorpora una mejora en este sentido.

3. Aprendizaje automático y clasificación

En esta sección se describe el aprendizaje automático y clasificación en el contexto de la predicción de polaridad en textos.

[1] Noticias de última hora y noticias en vivo de los últimos minutos/horas. Noticias clasificadas de acuerdo a los sujetos.

[2] Análisis en Tiempo Real de Noticias de Medicina y temas relacionados con la salud, alertas tempranas por categoría y país.

[3] Europe Media Monitor

3.1. Aprendizaje automático

El aprendizaje automático es una rama o subdisciplina de las ciencias de la computación fundamentada en el cómputo suave y el cómputo granular. Dicha disciplina estudia la construcción de algoritmos que puedan aprender de datos de análisis y posteriormente hacer predicciones con otros conjuntos de datos. Para la labor del aprendizaje, el algoritmo debe ser capaz de construir un modelo basado en un conjunto de entrenamiento con el fin de realizar predicciones en el conjunto de datos de prueba.

Los conjuntos de datos utilizados en el aprendizaje automático deben poseer características específicas, y deben corresponder a uno de los siguientes tipos:

- Continuos. Es decir, pueden ser cadenas de texto plano.
- Categorizados. Lo que se puede entender como una discretización de valores ya sea numéricos o alfanuméricos.
- Binarios. Todos los de tipo "verdadero" o "falso".

De manera similar a la minería de datos, la función de los algoritmos es encontrar patrones y tomar decisiones ajustadas correctamente. El aprendizaje se divide en 2 tipos: supervisado y no supervisado. Si en el conjunto de datos las instancias son marcadas con la respuesta correcta entonces el aprendizaje es de tipo supervisado. En contraste con el aprendizaje no supervisado las instancias no están marcadas por lo cual los investigadores lo utilizan para descubrir datos útiles [6]. Algunos autores describen el tipo semisupervisado como un conjunto parcialmente marcado para un aprendizaje corto y predicción casi desde cero.

El aprendizaje automático tiene varias aplicaciones más allá de desarrollar inteligencia artificial, por ejemplo, según Forbes.com [7], el sitio Amazon planea crear un algoritmo para automatizar el control de acceso a empleados sin la necesidad de la intervención humana que garantice y revoque permisos. Otra aplicación es la identificación de fallas cardíacas que desarrolla IBM lo que supone que una computadora pueda aprender de la información de un paciente y determinar si tiene, tendrá o tuvo una falla cardíaca, sin la necesidad de un médico que lo determine. Y también se ha utilizado en Singapur para desarrollar una aplicación de teléfono móvil que predice ataques y convulsiones, el sistema aprende la diferencia de movimiento común del usuario y cuenta con el patrón de movimientos que ocurren durante una convulsión o un ataque. Entre otras aplicaciones.

3.2. Clasificación

La clasificación automática de textos es un gran reto hoy en día debido a la gran cantidad de información que se puede encontrar en la Web, y una manera de enfrentar este reto es el aprendizaje automático. Sin embargo para que se pueda entrenar a una computadora para clasificar estos textos es necesaria la intervención humana en el preparativo de los datos de entrenamiento, es decir,

una o muchas personas expertas en el área deberían ser las que clasifiquen la información, lo cual consume una enorme cantidad de tiempo y esfuerzo.

La clasificación automática puede ser dividida en 2 tareas bien definidas: la clasificación supervisada que provee de un conjunto de datos de entrenamiento que han sido etiquetados manualmente por un ser humano, y la clasificación no supervisada en la cual no posee ningún esfuerzo humano, la computadora es encargada de inferir alguna función para poder aprender de los datos y etiquetarlos. Hay una tercera forma llamada clasificación semi-supervisada o híbrida, en la cual solo algunos datos son etiquetados con intervención humana.

Entre los métodos más utilizados para la clasificación automática se encuentran: Expectativa Máxima (EM), Clasificador de Naïve Bayes, Máquinas de soporte vectorial, Algoritmo de los vecinos más cercanos, árboles de decisión y Redes neuronales artificiales. Se necesitan grandes cantidades de información para obtener una alta precisión, y la dificultad de obtener datos etiquetados lleva a una importante pregunta: ¿Qué otras fuentes de información pueden reducir la necesidad de datos etiquetados? [9].

Esta investigación se enfoca en proponer un modelo algorítmico con el fin de lograr una clasificación automática híbrida de los datos de entrenamiento, intentando eliminar la necesidad de la intervención humana en el etiquetado de los datos con los que la computadora aprenderá.

4. Planteamiento del problema

Se cuenta con una gran colección de noticias en Español recuperadas y digitalizadas de distintos diarios digitales del estado de Puebla, los cuales abarcan una gran variedad de temas. Es necesario obtener la polaridad de estas noticias dependiendo del reportaje debe clasificarse positiva o negativa.

Debido a la gran cantidad de texto que estas noticias contienen e igualmente el alto número de noticias con las que se cuentan causaría un gran esfuerzo humano el clasificar cada una de ellas como buena o mala noticia, ¿Será posible etiquetar estas noticias con algún método algorítmico?

4.1. Conjunto de datos de aprendizaje

Para esta investigación se usarán diversos recursos, principalmente se cuenta con un corpus de noticias que se clasificarán mediante el algoritmo 1. Para el entrenamiento se utilizó un corpus conteniendo 2904 noticias. En la etapa de pruebas se utilizaron cuatro corpus los cuales contenían 100, 500, 1000 y 1762 noticias cada uno.

Los corpus de entrenamiento serán reclasificados con el algoritmo de clasificación automática propuesto. Se cuenta con dos lexicones obtenidos de Minqing Hu y Bing Liu [4,8], estos lexicones contienen 4783 palabras con polarización negativa y 2005 palabras con polarización positiva, los cuales originalmente están escritos en el idioma inglés y fueron traducidos de manera automática mediante el traductor de Google al idioma Español. Por otra parte se utiliza un diccionario que contiene las conjunciones del idioma Español.

5. Propuesta de solución

Se propone un algoritmo cuantitativo para la clasificación automática de textos basados en un árbol de búsqueda binaria el cual generará además un diccionario pesado mediante la medida TF*IDF para obtener la importancia de las palabras en las noticias y su polaridad. El resultado de este algoritmo es un conjunto de datos de prueba. Se utilizará además el algoritmo de Naïve Bayes para el entrenamiento y prueba, esta última será apoyada por el diccionario generado mediante el algoritmo 1.

El modelo que se propone es un modelo cuantitativo que utiliza una ventana máxima de 3 palabras para analizar frases y determinar la polaridad de la frase analizada. Dicho modelo se enfoca en revisar el texto redactado por una persona de una noticia que ha aparecido en un diario y se ha digitalizado. Para la evaluación de las frases se utilizan los dos lexicones mencionados en la sección 4.1, los cuales fueron traducidos automáticamente al idioma español y enlistan 2005 palabras que tienen una polaridad positiva y 4783 palabras con una polaridad negativa, ninguna palabra esta repetida y si una palabra aparece en un lexicón, significa que no aparecerá en el otro, es decir, son lexicones disjuntos. La revisión del texto de la noticia no se hace palabra por palabra pues esto causa una pérdida de contexto, sino que se analiza un conjunto de tres palabras continuas reconocidas por los lexicones. El análisis del texto obedece al algoritmo 1.

Algoritmo 1. Clasificación cuantitativa de polaridad.

```
N: Noticia
cp: Contador Positivo
cn: Contador Negativo
D: Documento
w: palabra
LP: Lexic\'on Positivo
LN: Lexic\'on Negativo
For each N do:
  cp := 0
  cn := 0
  For i:=0 to len(D) do:
    If w[i] pertenece LP then:
      If w[i+1] pertenece LP then:
        If w[i+2] pertenece LP then:
          cp:+= 5
        Else If w[i+2] pertenece LN then:
            cn:+= 3
        Else:
            cp:+= 2
      Else If w[i+1] pertenece LN then:
        If w[i+2] pertenece LP then:
          cn:+= 3
        Else If w[i+2] pertenece LN then:
          cp:+= 3
```

```
      Else:
        cn:+= 2
    Else If w[i+2] pertenece LP then:
      cp:+= 2
    Else:
      cp:+= 1
  Else If w[i] pertenece LN then:
    If w[i+1] pertenece LP then:
      If w[i+2] pertenece LP then:
        cp:+= 1
      Else If w[i+2] pertenece LN then:
        cn:+= 3
      Else:
        cn:+= 1
  Else If w[i+1] pertenece LN then:
    If w[i+2] pertenece LP then:
      cn:+= 3
    Else If w[i+2] pertenece LN then:
      cn:+= 5
    Else:
      cn:+= 2
  Else If w[i+2] pertenece LN then:
    cn:+= 1
  Else:
    cn:+= 1
If cp > cn then: "positiva"
Else If cp < cn then: "negativa"
Else: "positiva"
```

Un ejemplo del cálculo de la polaridad se presenta en la Figura 1, la cual representa un árbol de búsqueda binaria. El primer nivel del árbol representa la polaridad de la primera palabra a analizar, el segundo nivel es similar al primer nivel pero con la palabra que le sigue y el tercer nivel es la palabra que está a 2 palabras de distancia de la primera.

Una palabra positiva es aquella que se encuentra en el lexicón positivo, de igual manera para las palabras negativas, mientras que una palabra neutra es aquella que no aparece en ningún lexicón o que sea una conjunción. Cuando la frase de 3 palabras seguidas es terminada de evaluar, un contador toma el valor expresado en la última hoja del árbol en la que se haya movido. Si solo quedan 2 palabras finalmente por analizar entonces se analizará una frase de 2 palabras y el contador se incrementará dependiendo del nodo al cual se haya movido, de igual manera se realiza si solo queda una última palabra por analizar.

6. Experimentos

Un primer experimento utilizó como entrenamiento el corpus con la polaridad original. Posteriormente el corpus de entrenamiento es modificado en su

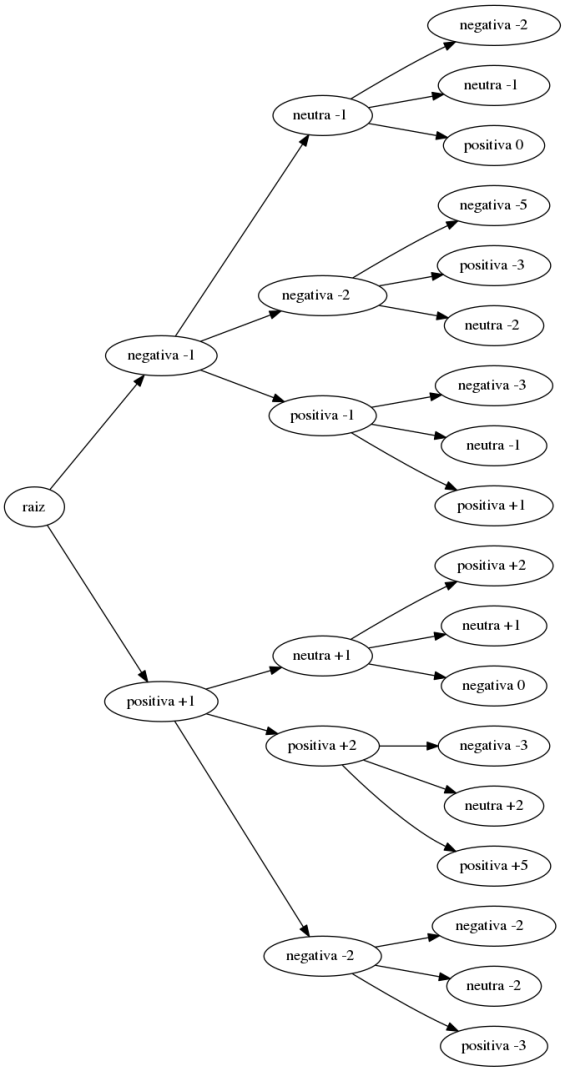

Fig. 1. Árbol de búsqueda binaria que expresa como es evaluada una frase de 3 palabras dentro de un texto.

atributo de polaridad, el cual es calculado utilizando el árbol de decisión binario propuesto. Además se genera un diccionario de apoyo para la prueba del algoritmo de Naïve Bayes. Este diccionario contiene palabras clasificadas como positivas y negativas y la relevancia de la palabra en las noticias del entrenamiento utilizando la medida TF*IDF. Para los experimentos se presentan dos ejemplos de clasificación, los cuales analizan noticias que no son un caso trivial de clasificación.

6.1. Clasificación negativa de noticia

La siguiente noticia es un ejemplo de la complejidad al determinar una polaridad utilizando clasificación automática. Texto original:

El gobernador rechazó darles audiencia para abordar la represión y los presos políticos en Puebla. El gobernador de Puebla Rafael Moreno Valle no recibió a diputados y senadores izquierdistas quienes buscaron una audiencia con él Los legisladores de izquierda Aida Valencia Ricardo Monreal José Arturo López Cándido Manuel Huerta Ladrón de Guevara Alfonso Durazo Juan Luis Martínez Rodrigo Chávez Gerardo Villanueva Loreta Ortiz Luisa María Alcalde María Fernanda Romero Jaime Bonilla y los senadores Mario Delgado David Monreal Martha Palaxof Adán Augusto López Manuel Bartlett y Rabindranath Salazar acudieron a Casa Puebla para reunirse con el gobernador de Puebla Sin embargo personal de las instalaciones les informó que no serían recibidos por Moreno Valle. Luego de esperar algunos minutos afuera de Casa Puebla los legisladores ingresaron a las instalaciones de Gobierno donde aguardaron la reunión con Moreno Valle la cual fue solicitada con antelación por Fernando Jara de la dirigencia estatal de Morena en Puebla. La diputada Luisa María Alcalde escribió "Nos dan aviso en las oficinas del gobernador Moreno Valle de Puebla Rafa Gobernador que rechaza audiencia" La administración estatal había condicionado la reunión de legisladores de Morena para que se llevara a cabo en Casa Aguayo con el secretario de Gobierno Luis Maldonado. La renuencia de Moreno Valle para reunirse con los diputados izquierdistas contrasta con las reuniones que el propio gobernador ha sostenido con diputados locales del PRI PAN Panal PRD y con los líderes magisteriales del SNTE a quienes ha recibido personalmente

El analizador es capaz de juzgar las frases:

-1 | Frase: rechazo darles audiencia
-1 | Frase: abordar la represion
-1 | Frase: no recibió a
-1 | Frase: no serian recibidos
-1 | Frase: rechaza audiencia la administración
-1 | Frase: renuencia de moreno

Esta noticia alcanza un puntaje negativo de 6 unidades en el algoritmo, por lo tanto será clasificada como negativa.

6.2. Clasificación positiva de noticia

Un segundo ejemplo para tratar de determinar una polaridad utilizando clasificación automática. Texto original:

Ayuntamiento de Quecholac presenta su bando de policía y gobierno. El ayuntamiento de Quecholac encabezado por el alcalde Néstor Camarillo medina presentó mediante una reunión general de la sindicatura la Regiduría de Gobernación la dirección de seguridad pública la Regiduría de Hacienda y el juzgado calificador el bando de policía y gobierno para el municipio este ordenamiento municipal fue reformado y adecuado a la actualidad ya que el

pasado bando era obsoleto por lo que ahora estará a la par de las reformas constitucionales y se pretende con el mismo regular las acciones conductas y movimientos mercantiles y de servicios que los ciudadanos reciben a través del ayuntamiento lo anterior fue dado a conocer por la síndico municipal Araceli Campos Jiménez quien explicó que el bando de policía y gobierno fue aprobado en sesión de cabildo el pasado de Abril y publicado en el diario oficial del gobierno del estado en Noviembre del 2015 en la reunión se dijo que no solamente fue una decisión del ayuntamiento el actualizar el reglamento municipal de la administración sino que también es en cumplimiento a la constitución federal local y la Ley orgánica municipal el bando de policía y gobierno se encuentra publicado en el portal de transparencia del Ayuntamiento de Quecholac.

El analizador es capaz de distinguir las siguientes frases:

+1 | Frase: dirección de seguridad
+1 | Frase: reformado y adecuado
-1. | Frase: pasado bando era
-1. | Frase: obsoleto por lo
+1 | Frase: reformas constitucionales y
-1. | Frase: dado a conocer
-1. | Frase: no solamente fue
+1 | Frase: decisión del ayuntamiento
+1 | Frase: cumplimiento a la

Esta noticia contiene frases que denotan tanto positividad como negatividad, y siguiendo el algoritmo propuesto da como resultado que el contador positivo alcanza 5 unidades y el contador negativo alcanza 4 unidades. Por lo que la noticia será clasificada como positiva.

7. Resultados

El algoritmo propuesto tiene una complejidad O(n), donde "n" es el número de palabras contenidas en el corpus a preprocesar. Después de la reclasificación de la polaridad, se ejecuta Naïve Bayes. Los resultados se muestran en la Tabla 2. Se aplicó el clasificador Naïve Bayes usando como entrenamiento noticias que se desconoce su criterio de clasificación, el cual contiene 2904 noticias. Las pruebas se harán 4 diferentes corpus y los resultados se especifican en la Tabla 1.

Tabla 1. Clasificador Naïve Bayes sin preprocesamiento.

Corpus	Numero de noticias	Presicion	Recall
1	100	34%	66%
2	500	37.6%	62.4%
3	1000	35.2%	64.8%
4	1762	53.97%	46.03%

Los resultados muestran precisiones con un promedio de 40.19%. Se observa que el numero de noticias afecta significativamente en la precisión. El corpus 4 duplica al corpus 3 y supera a este en aproximadamente 18% en precisión.

Tabla 2. Clasificador Naïve Bayes con preprocesamiento.

Corpus	Numero de noticias	Presicion	Recall
1	100	61%	39%
2	500	66%	34%
3	1000	64.9%	35.1%
4	1762	72.99%	27.01%

Estos resultados tienen precisiones con un promedio de 66.22%. Se puede observar una mejora de 26.03% como promedio en la precisión. Con respecto a la Tabla 1 a diferencia de los resultados de Naïve Bayes sin procesamiento, la precisión no incrementó significativamente con respecto al tamaño del corpus. Por lo tanto se puede concluir que el algoritmo propuesto exhibe una buena precisión en clasificación en corpus con pocas noticias.

8. Conclusiones

La propuesta reporta resultados satisfactorios en esta primera etapa de desarrollo. Aunque se enfrenta un reto de mejora, el algoritmo es capaz de reducir el esfuerzo humano en la clasificación de noticias. Sin embargo, se cree que se puede incrementar su eficacia utilizando diversas métricas tales como, técnicas de similitud semántica para determinar la carga emocional del texto en el diccionario de apoyo al algoritmo de clasificación Naïve Bayes.

No se presentan comparativas con los trabajos de SemEval, porque los corpus utilizados están en idioma Inglés, y además son para dominios específicos como laptops y restaurantes. Por otra parte solo se han encontrado reportados en la bibliografía predicción de polaridad de noticias en Twitter.

Se encuentra en progreso la utilización de otros clasificadores como los árboles de decisión J48 y las máquinas de soporte vectorial con el objetivo de incrementar la precisión en los resultados.

Referencias

1. Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., Pouliquen, B., Belyaeva, J.: Sentiment analysis in the news. arXiv preprint arXiv:1309.6202 (2013)
2. Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., Gandini, G.: Language resources and linguistic theory: Typology, second language acquisition,

english linguistics. chapter micro-wnop: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining. Milano, IT 23 (2007)
3. Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: Proceedings of LREC. vol. 6, pp. 417–422. Citeseer (2006)
4. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 168–177. ACM (2004)
5. Kaur, A., Gupta, V.: A survey on sentiment analysis and opinion mining techniques. Journal of Emerging Technologies in Web Intelligence 5(4), 367–371 (2013)
6. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques (2007)
7. Laura, H.: Six novel machine learning applications. In: Forbes (Janury 2014), http://www.forbes.com/sites/85broads/2014/01/06/six-novel-machine-learning-applications/#5d6091f967bf
8. Liu, B., Hu, M., Cheng, J.: Opinion observer: analyzing and comparing opinions on the web. In: Proceedings of the 14th international conference on World Wide Web. pp. 342–351. ACM (2005)
9. Nigam, K., Andrew, K.M., Thrun, S.: Text classification from labeled and unlabeled documents using em. Kluwer Academic Publishers 39, 103–134 (2000)
10. Saias, J.: Sentiue: Target and aspect based sentiment analysis in semeval-2015 task 12. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 767–771. Association for Computational Linguistics, Denver, Colorado (June 2015), http://www.aclweb.org/anthology/S15-2130
11. Strapparava, C., Valitutti, A., et al.: Wordnet affect: an affective extension of wordnet. In: LREC. vol. 4, pp. 1083–1086 (2004)