

# SUPERVISIÓN REMOTA EN EL ENTRENAMIENTO DE UN CLASIFICADOR DE SENTIMIENTOS EN COMENTARIOS TURÍSTICOS

Martín, C.

Dpto. Ingeniería Informática y de Sistemas. Universidad de La Laguna, cmartin@ull.edu.es

Aguilar, R.M.

Dpto. Ingeniería Informática y de Sistemas. Universidad de La Laguna, raguilar@ull.edu.es

Torres, J.M.

Dpto. Ingeniería Informática y de Sistemas. Universidad de La Laguna, jmtorres@ull.edu.es

Díaz S.

Dpto. Ingeniería Informática y de Sistemas. Universidad de La Laguna, sdiazgon@ull.edu.es

## Resumen

*Este artículo describe el uso de algoritmos para automatizar la detección de sentimientos en los comentarios dejado por los turistas en plataformas eWOM (Electronic Word of Mouth – Boca a boca electrónico). Para ello se diseña y entrena un clasificador utilizando Redes Neuronales LSTM (Long short-term memory), con el que determinar el sentimiento de los comentarios publicados por los turistas en las redes sociales después de su estancia en los hoteles. Se demuestra la mejora del proceso de entrenamiento con el uso de supervisión remota. El incremento en el número de muestras en el conjunto de entrenamiento, aunque éstas presenten ruido, mejora los resultados del clasificador. En este trabajo se presenta un caso de uso con los comentarios de un conjunto de hoteles de la isla de Tenerife (Islas Canarias).*

**Palabras clave:** análisis de sentimientos, redes sociales, aprendizaje automático, supervisión remota, turismo

## 1. INTRODUCCIÓN

Las compañías cada vez tienen más presente la importancia de las opiniones, acerca de sus productos y servicios, que los usuarios publican en las redes sociales (RRSS). Concretamente, en el sector turístico la aparición de las redes sociales ha proporcionado una nueva y valiosa fuente de información para el estudio de la imagen de los distintos destinos. Esta forma de conocer las opiniones de los turistas conlleva ventajas, ya que la información que proporciona es espontánea, natural, y no está condicionada por el diseño de una encuesta. Así, podemos saber los factores realmente importantes en el destino desde el punto de vista del turista y su opinión sobre ellos. No obstante, debido a la diversidad y cantidad de información, hacer un seguimiento en tiempo real de las redes sociales es un reto enorme, que sólo se puede abordar de forma automatizada. Sin embargo incluso así se trata de un problema que presenta series dificultades, dada la heterogeneidad de la información y la dificultad añadida del procesamiento del lenguaje natural. En este sentido la aplicación de técnicas de aprendizaje automático es necesaria para salvar estas barreras. Los sentimientos expresados por los turistas, son las culpables de que éstos repitan experiencia o simplemente, compartan públicamente su estado de ánimo mientras disfrutaban de sus vacaciones. De ahí viene la importancia que existe de que los empresarios turísticos puedan medir el nivel de satisfacción de sus clientes. La idea es conocer qué tipo de comodidades desatan emociones y ser capaces de identificarlas y ponerles nombre. Pero para ello debemos ser capaces de medir estas

emociones. Se dice que lo que no se puede medir no se puede controlar, y por lo tanto, no se puede mejorar.

No es tan sencillo comprender la afirmación anterior. Estamos hablando de medir emociones y eso nunca es tarea fácil. Por ejemplo, cómo registramos esa sensación de oler a café recién hecho mientras uno responde la primera llamada del día. Este tipo de sensaciones ocurren cada día pero no solemos registrarlas como tal y, aun así, nos generan una emoción. Esta emoción, más adelante es la que acompaña a los recuerdos que vamos almacenando en nuestra mente. Y así es como vamos dando forma a la estancia vivida.

Se sabe que el producto turístico, al estar compuesto por servicios, se caracteriza entre otras cosas por la caducidad del mismo. Es decir, el producto turístico no es almacenable: se consume en el momento programado o se pierde. Una vez consumido, el envoltorio, que es lo único que queda de él, es ese recuerdo que recoge las emociones sentidas durante la estancia. Y este recuerdo es uno de los componentes de la reputación que consigue una empresa turística. El conjunto de emociones sentidas por el cliente durante su estancia son las que ayudarán a crear la imagen de la cadena. Y a la vez estas emociones, convertidas después en recuerdos, son las que impulsarán al turista a compartir su experiencia con el entorno, que actualmente son las Redes Sociales. No hay que olvidar que las emociones son el concepto más viral en el mundo de las redes sociales. Hay que conseguir hacer sentir para lograr que el turista voluntariamente hable bien de la cadena. Para medir estas emociones en las redes sociales, se propone en este trabajo el uso de algoritmos de aprendizaje automático para extraer automáticamente los sentimientos expresados por los turistas en su visita a hoteles de Tenerife (Islas Canarias).

El aprendizaje automático (Machine Learning) es una rama dentro del campo de la Inteligencia Artificial. El término se viene manejando desde los años 50, pero ha sido en los últimos años cuando ha tomado gran relevancia gracias al enorme aumento de la capacidad de cómputo y al gran volumen de datos que las empresas empiezan a manejar. Estos algoritmos tienen como objetivo dotar a los computadores de la capacidad de aprender sin la necesidad de ser explícitamente programados. Tradicionalmente los algoritmos se han venido clasificando en dos grandes grupos:

- Aprendizaje supervisado: Donde disponemos de un conocimiento anterior que nos puede ayudar a entender los nuevos datos que llegan.

- Aprendizaje no supervisado: En este caso no tenemos experiencia previa para analizar los nuevos datos, está más orientado a la búsqueda de patrones.

En el caso que nos ocupa, tenemos muchos comentarios en redes sociales que sabemos clasificar como positivos o negativos. Por lo tanto, cabe diseñar algún algoritmo de aprendizaje supervisado que extraiga ese conocimiento. Los algoritmos de aprendizaje supervisado se basan en que disponemos de datos históricos de los que podemos aprender el comportamiento del sistema. Por ejemplo en un sistema de control de SPAM, donde el usuario indica los correos electrónicos que son SPAM y el sistema aprende de ello para ser capaz de identificar nuevos correos de SPAM en el futuro. Otro ejemplo podría ser un sistema de clasificación de enfermedades donde podemos basarnos en el conocimiento previo de los síntomas y características de ciertas enfermedades para que nuestro sistema aprenda y reconozca esas enfermedades en el futuro.

Para que estos algoritmos puedan funcionar correctamente debemos ser capaces de alimentarlos con un buen conjunto de datos de entrenamiento hasta conseguir el nivel de acierto que buscamos. Los datos de entrenamiento deben estar correctamente etiquetados para que el algoritmo infiera las características para que un elemento pertenezca a un grupo u otro.

El artículo describe en la sección 2 el conjunto de entrenamiento utilizado. En el apartado 3 se describe el diseño del algoritmo de aprendizaje automático para la determinación de los sentimientos en los comentarios turísticos. En el epígrafe 4 se presentan los resultados en la implementación de dicho algoritmo. Finalmente se termina con unas conclusiones.

## 2. CONJUNTO DE DATOS

La preparación de datos de entrenamiento es diferente para cada problema. La preparación comienza con pasos simples, como cargar datos, pero rápidamente se vuelve difícil con las tareas de limpieza que son muy específicas de los datos con los que está trabajando.

Para el caso de monitorización automática de sentimientos de turistas de Tenerife, se extraen muestras de comentarios con sus valoraciones de turistas para hoteles de Arona, Costa Adeje y Puerto de la Cruz. El objetivo posterior es realizar el clasificador que indiquen el sentimiento reflejado en cada comentario. Para construir el clasificador es necesario usar una muestra de textos ya clasi-

ficados que sirvan de entrenamiento. Asimismo es necesario tener un conjunto de datos que no se hayan utilizado en el proceso de entrenamiento para las tareas de verificación.

### 2.1. Conjunto de entrenamiento y verificación

Para realizar el entrenamiento hemos adquirido muestras de comentarios positivos y negativos. Se obtienen las muestras de entrenamiento del portal británico de Booking.com. Los comentarios de los hoteles de booking presentan las siguientes características:

- Tiene una valoración numérica global entre 0 y 10 (admite decimales).
- En algunas ocasiones se puede obtener información sobre la procedencia del comentario y la fecha en la que se realizó.
- La persona que realiza el comentario tiene la posibilidad de poner en el mismo comentario un título, una parte positiva de su comentario y una parte negativa de su comentario.

Con estas características, y después de hacer una extracción masiva podremos configurar una muestra para entrenar de la siguiente manera:

- Extraemos aquellas revisiones con una puntuación mayor de 6 y las etiquetaremos como “buenas”, tomando como texto para entrenar la concatenación del título y parte positiva del comentario.
- Extraemos aquellas revisiones con una puntuación menor de 5 y las etiquetaremos como “malas”, tomando como texto para entrenar la concatenación del título y parte negativa del comentario.

De las muestras obtenidas haremos una selección de buenas y malas distribuidas uniformemente por el rango de puntuación. Durante el trabajo se observa que el número de comentarios negativos es muy inferior al de positivos. También varía mucho el número de comentarios según el destino turístico. Para obtener una muestra variada de comentarios para entrenar y no limitar las muestras de entrenamiento a comentarios realizados sobre los destinos que después se pretenden analizar se toman muestras de los siguientes emplazamientos: Puerto de La Cruz en Tenerife, Arona en Tenerife, Costa Adeje en Tenerife, Costa del Silencio en Tenerife, Maspalomas en Gran Canarias, Playa del Inglés en Gran Canaria, Puerto Rico en Gran Canaria.

Se desarrolla un programa que convierte el fichero integrado de muestras en un fichero formato CSV (valor separado por comas) con dos columnas. La primera es la concatenación del título con el comentario (parte positiva si la valoración es mayor de cinco y parte negativa si la valoración es menor que cinco), y la segunda que es una clasificación ‘Good’ o ‘Bad’ en función de si la puntuación es mayor o menor que cinco.

Para poder equilibrar y tener más muestras se extraen comentarios de otro portal de turismo que contiene comentarios (diferentes a las muestras obtenidas en el paso anterior): TripAdvisor.com. De este portal obtenemos la siguiente información de cada comentario:

- Título del comentario.
- Fecha del comentario.
- Contenido del comentario.
- Procedencia del turistas.
- Valoración del comentario. Se usa un sistema de estrellas en el que el turista podrá valorar entre 1 y 5.

En este portal Tripadvisor no se hace distinción entre la parte positiva del comentario y negativa del mismo. La valoración se entiende como una nota global del mismo. Se realiza una extracción de comentarios, obteniendo los siguientes datos: Puerto de la Cruz (13541 muestras), Costa Adeje (20017 muestras) y Arona (19614 muestras).

Para realizar la clasificación de estos ficheros de muestras se implementa un programa que añade las columnas ‘class’ (que podrá tener los valores ‘Good’ o ‘Bad’ en función de lo que determine el clasificador entrenado), y ‘probability’ la probabilidad de acierto (valor real entre 0 y 1) de la clasificación realizada.

El número total de muestras con el que se entrena es de 9640 muestras, mitad buenas y mitad malas.

### 2.2. Supervisión a distancia

En el aprendizaje automático se requiere de datos de entrenamiento etiquetados a mano para el entrenamiento de un clasificador. Con la amplia gama de temas discutidos por los turistas en las RRSS, sería muy difícil recopilar manualmente datos suficientes para implementar un clasificador de sentimientos para los comentarios en las plataformas web. La solución propuesta es usar supervisión a distancia, en la que los datos de entrenamiento consisten en tweets con emoticones. Este enfoque fue introducido por [1]. Los emoticones actúan como etiquetas ruidosas. Por ejemplo, :) en

un tweet indica que el tweet contiene sentimiento positivo y :( indica que el tweet contiene sentimiento negativo. Con la ayuda de la API de Twitter, es fácil extraer grandes cantidades de tweets con emoticones en ellos.

Esta es una mejora significativa ya que se pueden extraer un extenso grupo de muestras (1.5 Millones de muestras en el ejemplo realizado [2]) y reducimos las muchas horas que de lo contrario podría llevar etiquetar a mano dicho conjunto de entrenamiento. Las desventajas que a priori presenta el uso de tweet con emoticonos para entrenar el algoritmo de detección de sentimientos en las plataformas turísticas son:

- El tipo y longitud de comentarios no coinciden
- El conocimiento de dominio no es turismo

En el apartado 4 se muestran los resultados de los dos clasificadores implementados, uno utilizando el conjunto más reducido de muestras del dominio y otro utilizando los tweets con emoticonos. Se comprueba que el entrenamiento con un mayor número de datos (*big data*), aunque no sean del dominio de turismo y con una longitud menor cada comentario, mejora el rendimiento del clasificador.

### 3. MÉTODO

La clasificación de texto describe una clase general de problemas, como predecir el sentimiento de reseñas de turistas en RRSS, o como clasificar el correo electrónico como spam o no. Los métodos de deep learning están demostrando ser buenos en este tipo de clasificaciones. En este epígrafe se describirá la el algoritmo de deep learning LSTM- Long short-term memory para clasificación de texto. El modus operandi para la clasificación de texto implica el uso de Word embedding para representar palabras y un clasificador para aprender a discriminar documentos sobre problemas de clasificación. [3] comenta que las redes neuronales en general ofrecen mejor rendimiento que los clasificadores lineales clásicos, especialmente cuando se usan con Word embeddings. En este trabajo queremos demostrar que las redes neuronales LSTM son efectivas en la clasificación de documentos. La arquitectura de este método se compone de los siguientes 3 sistemas:

- *Word Embedding*: Una representación distribuida de palabras donde las diferentes palabras que tienen un significado similar (basado en su uso) también tienen una representación similar.

- *Modelo LSTM*: modelo de clasificación basado en redes neuronales recurrentes. LSTM son efectivas en la clasificación de documentos, concretamente porque son capaces de seleccionar características destacadas (por ejemplo, tokens o secuencias de tokens) de manera invariante a su posición dentro de la secuencia de entrada.
- *Perceptrón Multicapa*: Red neuronal para la interpretación de las características extraídas en términos de un resultado predictivo.

La combinación de estos tres elementos muestra buenos resultados en la clasificación de sentimientos de los comentarios de los turistas.

### 4. RED NEURONAL LSTM

Long Short-Term Memory (LSTM) son un tipo de redes neuronales recurrentes. Se han mostrado útiles en problemas complejos de diferentes dominios de aplicación como traducción de idiomas, subtítulos automáticos de imágenes y generación de texto.

Los LSTM son muy diferentes a otras técnicas de aprendizaje profundo, como los Perceptrones Multicapa (MLP) y las Redes Neuronales Convolucionales (CNN), ya que están diseñados específicamente para problemas de predicción de secuencia.

Como cualquier red neuronal, la red LSTM está compuesta de capas de neuronas. Los datos de entrada se propagan a través de la red para hacer una predicción. Igual que las RNN (redes neuronales recurrentes), los LSTM tienen conexiones recurrentes de modo que el estado de la red en el instante anterior se utiliza para formular la salida en el instante actual. Pero a diferencia de otras RNN, el LSTM tiene una formulación única que le permite evitar los problemas de entrenamiento y escalado de la red.

El gran problema en las RNN es determinar el algoritmo de entrenamiento que evite la degradación de gradiente que hace los pesos de las conexiones se vuelvan tan pequeños que no permitan converger. El desafío histórico técnico clave que enfrentan los RNN es cómo entrenarlos de manera efectiva. Los LSTM superan este handicap por diseño.

Una capa LSTM consiste en un conjunto de bloques conectados recurrentemente, conocidos como bloques de memoria. Cada uno tiene conectada una o más celdas de memoria y tres unidades multiplicativas - las puertas de entrada, salida y reseteo - que proporcionan operaciones de escritura, lectura y reinicio de las celdas. La red solo puede

interactuar con otras células a través de las puertas.

Una capa LSTM consiste en un conjunto de bloques conectados recurrentemente, conocidos como bloques de memoria. Cada celda tiene 3 entradas (estado anterior de la celda, salida anterior y entrada actual) y dos salidas (salida y estado actual).

## 5. RESULTADOS

El objetivo de este trabajo fue el diseño e implementación de un clasificador automático que determine el sentimiento expresado por los turistas de Tenerife (Canarias) cuando se han alojado en distintos centros hoteleros de la isla. Los comentarios se extraen a partir de las publicaciones que han compartido los turistas en las redes sociales a través de plataformas eWOM como booking.com y tripadvisor.com. [4] La primera fase consistió en la preparación del conjunto de datos para la aplicación de modelos de aprendizaje profundo. Cada comentario que se usa del conjunto de entrenamiento es sometido a un preprocesado antes de que pueda ser usado. Los tratamientos realizados sobre los comentarios consistieron en: separarlos en palabras, eliminar los signos de puntuación y aquellos términos que no eran alfabéticos y, por último, eliminar las palabras identificadas como *stopwords*. Estas palabras no aportan información a la hora de terminar si un comentario es positivo o negativo.

Los algoritmos utilizados requieren como entrada un vector de dimensión fija, en el que cada componente es un número. En la técnica de codificación de textos a vectores numéricos conocida como *Bag of Words* (BoW), se creará un diccionario con las palabras más frecuentes en todos los comentarios de entrenamiento. Cada comentario se codificará entonces en un vector de dimensión fija que corresponde con el número de palabras del diccionario creado. En BoW un comentario se codifica en un vector en el que cada componente es un contador de apariciones de cada palabra del diccionario en todo el comentario. Descartamos esta forma de codificación porque aunque esté representada la frecuencia de palabras en el comentario, se pierde la información relativa al orden en el que se encuentran dentro del mismo.

La técnica *Word Embedding* es actualmente una de las técnicas que producen mejores resultados para representar textos como vectores numéricos. Se trata de una representación aprendida en la que, palabras con significado similar tendrán una representación similar. Cada palabra del vocabulario se representará con un vector, y su representación es aprendida en función del uso que se

le ha dado en los comentarios de entrenamiento, de manera que palabras que hayan sido usadas de forma similar tendrán una representación similar. El proceso de aprendizaje para el embedding se realiza en el marco de este trabajo añadiendo una capa al frente de la red neuronal.

La salida de la capa de embedding es una matriz de 582x300 en el que cada comentario se transforma en un vector de 582 componentes, y cada una de las componentes (que representaban una palabra) estaba codificada en un vector de dimensión 300.

Finalmente se diseña la arquitectura de todo el algoritmo de aprendizaje automático como el que se describe en la Fig. 1 para el caso en que se entrena con datos turísticos y en la Fig. 2 para el entreno con tweets con emoticonos. Tienen la misma estructura pero varía el número de entradas y salidas por la diferencia en el número de datos de entrenamiento. A continuación se describe indicando  $a/b$ , donde en  $a$  son los datos para el primer caso y en el  $b$  para la supervisión a distancia. Está formado por una capa de *embedding* donde cada palabra del comentario de entrada (tienen como máximo 582/634 palabras) se codifica en un vector con 32/300 valores.

La LSTM se definió formada por 100/30 celdas. Como entrada reciben cada comentario (582/634 palabras codificadas con 32/300 valores) y genera una salida por cada celda. Finalmente el perceptron multicapa formado por una capa de entrada de 100/30 neuronas y una capa de salida de 1 neurona, determina si el comentario es positivo o negativo

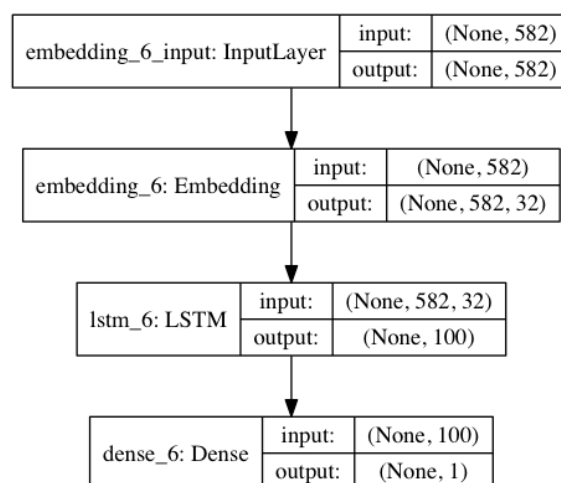


Figura 1: Estructura del algoritmo de clasificación de sentimientos. Conjunto de entrenamiento pequeño

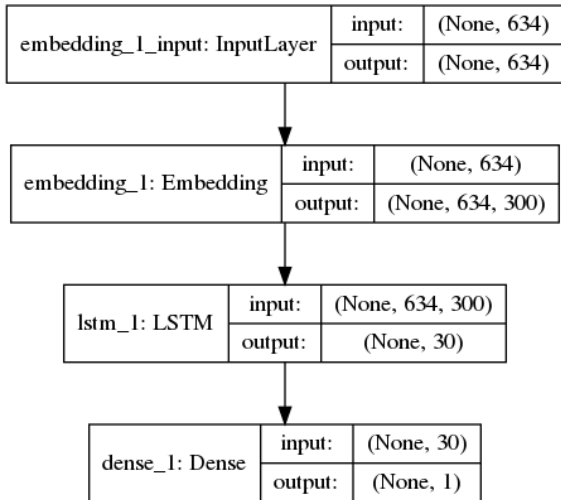


Figura 2: Estructura del algoritmo de clasificación de sentimientos. Conjunto de entrenamiento de big data

Los resultados del proceso de entrenamiento se puede observar en las Fig. 3 y Fig. 4. En 10 épocas obtenemos un error de 0 y se consigue un rendimiento de 0.99. Los resultados de convergencia resultan similares para ambos conjuntos de entrenamiento.

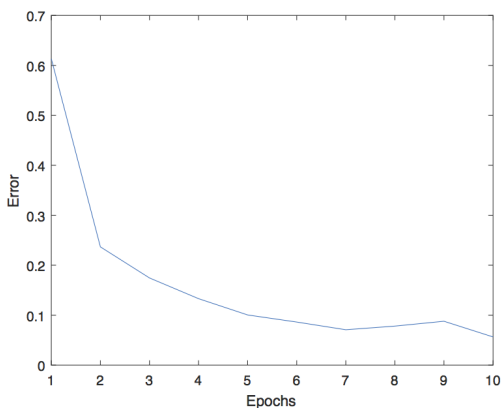


Figura 3: Error en el proceso de entrenamiento

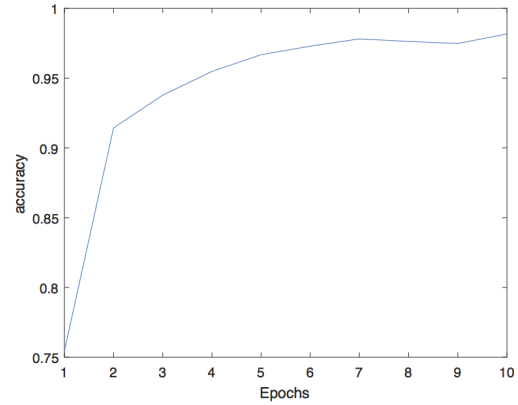


Figura 4: Aciertos en el proceso de entrenamiento.

La matriz de confusión, para el entrenamiento con datos turísticos, demuestra que el total predicciones erróneas fueron 310, el total de predicción acertadas 2475, obteniendo por lo tanto un porcentaje de error de 11.13%. El sesgo producido en el clasificador hace que haya más falso negativos que falsos positivos. Sin embargo, cuando entrenamos con un conjunto de datos mucho más extenso, aunque con un etiquetado con errores, se reduce el porcentaje de error a 8.04% ya que disminuye el sesgo que obtenemos con el primer clasificador.

		Valor Predicho	
		Positivo	Negativo
Valor Real	Positivo	1408	178
	Negativo	132	1377

Figura 5: Matriz de confusión del clasificador entrenado con datos reducidos del dominio.

		Valor Predicho	
		Positivo	Negativo
Valor Real	Positivo	1288	104
	Negativo	120	1273

Figura 6: Matriz de confusión del clasificador entrenado con datos masivos con emoticonos.

## 6. CONCLUSIONES

En el artículo se ha demostrado la necesidad de monitorizar la percepción del turista para mantener una ventaja competitiva, y cómo en el caso de la empresa turística, esto se hace especialmente importante. Sin embargo, determinar la percepción de millones de usuarios que dejan sus comentarios publicados por las redes sociales se convierte en una tarea engorrosa. Es por ello que nos planteamos la monitorización automática de las redes sociales y la clasificación de esos comentarios según si el sentimiento de la experiencia vivida es positivo o negativo. Como resultado del trabajo, proponemos utilizar supervisión remota para entrenar el clasificador. Esto es, el uso de millones de tweets etiquetados a través de los emoticonos que presentan. Aunque supone un conjunto con ruido, el hecho de tener gran cantidad (*big data*) de datos de entrenamiento mejora considerablemente la clasificación. Como algoritmo de clasificación hacemos uso de las redes neuronales LSTM (Long Short-Term Memory) que han mostrado buenos resultados en problemas de clasificación. Y dado que el impulsor principal de la economía Canarias es el turismo, se ha realizado un caso de uso sobre los comentarios realizados sobre los hoteles de la isla de Tenerife. En trabajos futuros se pretende crear clasificadores que puedan identificar los elementos sobre los que los turistas discuten (transporte, limpieza, belleza naturales, deportes, ...) para monitorizar automáticamente qué aspectos hay que promocionar, y cuáles reciben comentarios negativos y por lo tanto se deben mejorar.

### Agradecimientos

Este trabajo ha sido financiado por la Fundación CajaCanarias a través del proyecto titulado "VI-TUIN: Vigilancia Turística Inteligente de Tenerife en Redes Sociales" n°2016TUR15.



### English summary

## DISTANT SUPERVISION IN THE TRAINING OF A SENTIMENT CLASSIFIER IN TOURIST REVIEW

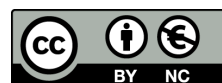
### Abstract

*This paper describes an algorithm for automatically identifying the sentiments expressed by tourists on eWOM (Electronic Word of Mouth) platforms. Based on reviews published by tourists after staying in hotels, a classifier is trained using an LSTM (Long short-term memory supervised learning algorithm). The improvement of the training process with the use of remote supervision is demonstrated. The increase in the number of samples in the training set, although these present noise, improves the results of the classifier. We present a use case for this method involving a group of hotels located on the island of Tenerife (Canary Islands).*

**Keywords:** sentiment analysis, social net, machine learning, distant supervision, tourism

### Referencias

- [1] J. Read, "Using emoticons to reduce dependency in machine learning techniques for sentiment classification," *Proceedings of ACL-05, 43rd Meeting of the Association for Computational Linguistics*, 2005.
- [2] G. Alec, B. Richa, and H. Lei, "Twitter sentiment classification using distant supervision," *Technical Report CS224N*, pp. 49–53, 2009.
- [3] Y. Goldberg, "A primer on neural network models for natural language processing," *Journal of Artificial Intelligence Research*, vol. 57, pp. 345–420, 2016.
- [4] C. M. L. M. García-Pablos A., Lo Duca A. and M. A., "Correlating languages and sentiment analysis on the basis of text-based reviews," *Information and Communication Technologies in Tourism*, pp. 565–577, 2016.



© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution CC-BY-NC 3.0 license (<http://creativecommons.org/licenses/by-nc/3.0/>).