



TRABAJO FIN DE GRADO
DOBLE GRADO EN INGENIERÍA INFORMÁTICA Y
MATEMÁTICAS

Análisis de sentimientos

**Aplicación de Deep Learning para la extracción de
características.**

Autor

Nuria Rodríguez Barroso

Directores

Francisco Herrera Triguero

M^a Victoria Luzón García



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍAS INFORMÁTICA Y DE
TELECOMUNICACIÓN

Granada, septiembre de 2018

Agradecimientos

AGRADECIMIENTOS

Resumen

RESUMEN EN ESPAÑOL

Abstract

RESUMEN EN INGLES

Consentimientos

Índice general

1	Introducción	11
1.1	Contexto	11
1.2	Motivación	12
1.3	Objetivos	15
1.4	Requerimientos	15
2	Análisis de sentimientos	17
2.1	Procesamiento del Lenguaje Natural	17
2.2	Concepto de Análisis de sentimientos	18
2.2.1	Concepto de Opinión	19
2.2.2	Niveles de clasificación	20

Capítulo 1

Introducción

1.1 Contexto

La principal característica del ser humano y la que le diferencia del resto de seres vivos es la racionalidad. El ser humano utiliza la razón para la toma de decisiones, contemplando los diferentes escenarios y evaluando la mejor manera de alcanzar sus objetivos. Sin embargo, el ser humano consta de una racionalidad limitada, por lo que el abanico de posibilidades que baraja en cada momento está limitado por su visión de la realidad.

Es por esto que el proceso de toma de decisiones constituye un gran reto. Para facilitar el proceso de toma de decisiones y con el fin de ampliar el abanico de consecuencias contempladas, es muy común la acción de “pedir opinión” a terceros.

La ayuda en la toma de decisiones no es la única función de las opiniones. Las personas también utilizamos las opiniones para expresar nuestro juicio acerca de variados temas. Cuando se produce un intercambio de opiniones ambos participantes se enriquecen con el punto de vista del contrario.

El uso de las opiniones ha evolucionado con el ser humano a lo largo de la historia. Comenzando en los inicios del lenguaje, pasando a estar por escrito con la llegada de la escritura y disparándose con el auge de la difusión de opinión en la prensa con la invención del a imprenta.

Uno de los acontecimientos más influyentes en la sociedad fue la invención de Internet en la segunda mitad del s. XX. Trajo consigo una gran fuente de información de fácil acceso. A principios del s.XXI enfatizó el cambio social que ya estaba ocurriendo un nuevo concepto de Web, la Web 2.0. Este innovador concepto ofrecía a todo usuario de ella la posibilidad de compartir todo tipo de información en Internet. Así fue como se preparó el camino para la llegada de las redes sociales, los blogs, o los foros. Plataformas donde los usuarios podían compartir todo tipo de pensamientos u opiniones sin ningún filtro. También se sumaron al carro de la Web 2.0 las empresas. Muchas de ellas incorporaron la opción de compra *online* suponiendo una nueva fuente de ingresos e incluso surgieron muchas nuevas empresas que llevan a cabo toda su funcionalidad a través de Internet.

1.2 Motivación

En el contexto de la Sección 1.1, tanto el auge de las redes sociales como la incorporación de las empresas al negocio *online* no trajeron consigo solo nuevas fuentes de beneficios. La Web 2.0 supuso una nueva (y enorme) fuente de información. Esta generación exponencial de datos de naturaleza desestructurada en Internet propició el nacimiento de lo que hoy conocemos como *Big Data*.

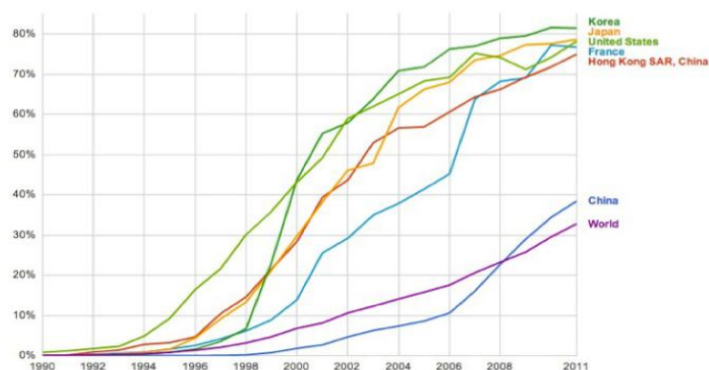


Figura 1.1: Gráfica del crecimiento del *Big Data* en algunos países y en el mundo.

1.2. MOTIVACIÓN

Para hacernos una idea del impacto del nacimiento de la Web 2.0, observamos la Figura 1.1 en la que se representa el crecimiento de *Big Data* en algunos países desarrollados y en el mundo en general. Notamos que en el año 1998 se produce un aumento en el ritmo de crecimiento, coincidiendo con el nacimiento de la Web 2.0. El siguiente cambio de crecimiento más brusco lo encontramos en el año 2006, año de mayor impacto de las redes sociales.

Como ya veníamos advirtiéndolo, está en la naturaleza del ser humano dar su opinión sobre los temas que le rodean. Esta caracterización del ser humano junto con la cantidad de datos depositados por usuarios en los diferentes portales de Internet nos da una idea de la cantidad de opiniones disponibles. Esta información será de gran utilidad para las empresas que quieran conocer la opinión de los usuarios sobre un determinado producto o, simplemente, para conocer la opinión de la población sobre un determinado tema. Sin embargo, la inmensurable cantidad de información disponible hace que contratar a personal especializado para que se dedique a estudiar las opiniones de un determinado producto o tema de actualidad sería inviable tanto desde el punto de vista económico como desde el temporal. Debido a la necesidad de monitorizar este procedimiento surge el concepto de Análisis de Sentimientos o Minería de Opinión, del que hablaremos en los siguientes capítulos.

Por su estructura de *microblogging* (red social con un número de caracteres limitados para resaltar la opinión), Twitter es un perfecto generador de opiniones para analizar mediante Análisis de Sentimientos. Ya se han llevado a cabo estudios sobre esta plataforma obteniendo muy buenos resultados. Por ejemplo, en las últimas elecciones presidenciales de EEUU se realizó un estudio sobre la opinión de los usuarios de Twitter sobre ambos candidatos. Para este estudio se utilizaron los tweets publicados hasta 43 días antes de las elecciones comenzando el primer día de candidatura. Utilizando expertos para el etiquetado de los datos como positivos o negativos referentes a cada uno de los candidatos y entrenando con un algoritmo de Naïve Bayes consiguieron un 94 % de correlación.

El término de Análisis de Sentimientos es un término relativamente reciente. Aunque ya haya tenido éxito en algunos ámbitos, es un campo aún sin explotar pues presenta dificultades que aún no han sido solventadas. Entre estas dificultades se encuentran la subjetividad de las opiniones, las particularidades del lenguaje, los contextos, los sarcasmos ... entre una interminable lista.



Figura 1.2: Gráfica que representa el interés del análisis de sentimientos en todo el mundo según Google Trends.

Esto hace que se presente como una campo de estudio muy llamativo, con muchas mejoras por hacer y con un futuro muy prometedor. Cada día más gente centra su interés en esta materia (Figura 1.2). De ahí que a lo largo de esta memoria nos dediquemos al estudio de una parte de este amplio campo.

Como ya podemos imaginar, el término Análisis de Sentimientos abarca un gran conjunto de tareas. El trabajo se va a centrar en una de esas tareas en concreto: la extracción de entidades o aspectos, que sería el primer paso del Análisis de Opiniones a Nivel de Aspecto (ABSA). Esta tarea tiene como fin identificar todas las entidades y sus aspectos y clasificar la opinión que sobre ellos se está expresando en un fragmento de texto.

Para explicar en qué consiste y su importancia, trabajaremos sobre un ejemplo concreto. Imaginemos que queremos analizar la siguiente opinión:

[1] *El ordenador tiene muy buen procesador, sin embargo la pantalla tiene poca resolución y el precio es muy elevado.*

Si tuviéramos que establecer una polaridad a dicha frase, sería una tarea complicada incluso para una persona. En la frase se nombran tres componentes de un mismo ordenador y se da una opinión sobre cada una de ellas. Así, *buen procesador* tendría connotaciones positivas mientras que *la pantalla tiene poca resolución* y *el precio es muy elevado* negativas. La polaridad general del producto dependerá de la importancia que se le dé a cada una de las partes. Como no podemos saber a priori un orden de prioridad establecido, sería muy útil poder dar una polaridad a cada una de las componentes.

Ahora bien, para poder realizar este proceso de forma automatizada nos

1.3. OBJETIVOS

encontraríamos con dos tareas: en primer lugar la identificación del aspecto y, en segundo lugar, la detección de la polaridad de esta. En este trabajo nos centraremos en la primera de estas tareas.

1.3 Objetivos

El objetivo del proyecto será comparar el uso de redes neuronales de convolución (CNN) y redes neuronales recurrentes basadas en puertas, en concreto la arquitectura Long-Term Short Memory (LSTM) en la tarea de la extracción de aspectos y entidades.

De este modo, se desarrollarán dos modelos neuronales distintos, y se comparará tanto el rendimiento de los mismos como la exactitud en los resultados obtenidos.

Desde un punto de vista más teórico se estudiarán las bases del deep learning, centrándonos en la base matemática de las redes neuronales contempladas. En concreto, nos centraremos en el estudio matemático de las redes neuronales recurrentes pues son de gran interés para nuestro problema.

Así, la estructura del trabajo constaría de los siguientes tomos:

1. Introducción teórica al Análisis de sentimientos.
2. Fundamentos matemáticos del Deep Learning.
3. Implementación del modelo basado en CNN.
4. Implementación del modelo basado en LSTM.
5. Comparativa entre los modelos desarrollados.

1.4 Requerimientos

Los requerimientos del proyecto que desarrollaremos se resumen en:

1. Clasificar/extraer entidades o aspectos mediante una arquitectura basada en CNN.

2. Clasificar/extraer entidades o aspectos mediante una arquitectura basada en LSTM.
3. Comparar los resultados obtenidos por las arquitecturas implementadas.

Capítulo 2

Análisis de sentimientos

Para definir el concepto de Análisis de sentimientos es necesaria una breve introducción al Procesamiento del Lenguaje Natural.

2.1 Procesamiento del Lenguaje Natural

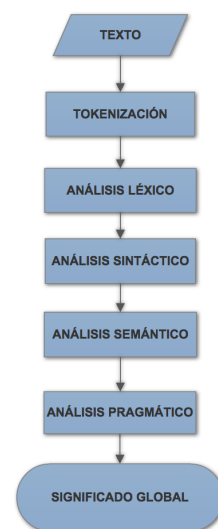
El Procesamiento del Lenguaje Natural (PLN) es un campo de la Inteligencia Artificial que se encarga de estudiar la comunicación que surge entre personas y máquinas. El objeto de estudio del PLN es, por tanto, el lenguaje. Para que las máquinas puedan conseguir capacidad de procesamiento y de generación del lenguaje, es necesario que se tenga comprensión total de la lengua. De esta misión se encarga la Lingüística, que trata de caracterizar y explicar los procedimientos del lenguaje.

El procedimiento usual de un sistema PLN se divide en tres fases principales: análisis sintáctico, semántico y pragmático. En la práctica, estas fases no son suficientes. Esto se debe a que no es viable realizar el análisis sintáctico de un texto sin haber identificado las palabras y las oraciones además de la información morfológica de las palabras que componen la frase. Así, a las tres fases anteriores habría que añadir dos nuevas fases: *tokenización* y análisis léxico.

A continuación observamos el flujo de trabajo que seguiría un sistema PLN. La funcionalidad de la cada una de las fases sería:

2.2. CONCEPTO DE ANÁLISIS DE SENTIMIENTOS

1. *Tokenización*: Identificar las unidades en las que se divide el mensaje, *tokens*.
2. *Análisis léxico*: Obtiene información de los *tokens* a partir de la función que desempeñan en la oración.
3. *Análisis sintáctico*: Obtiene las relaciones y dependencias existentes entre los elementos que componen la oración.
4. *Análisis Semántico*: Obtención de la intención del autor a partir de la información obtenida en la fase anterior. Busca descubrir el significado de la frase.
5. *Análisis Pragmático*: Busca las conexiones entre oraciones para dar un sentido global al discurso.



La búsqueda del entendimiento del lenguaje no es la única misión del PLN. También persigue la generación automática de lenguaje, aunque no va a ser en lo que centremos esta memoria.

2.2 Concepto de Análisis de sentimientos

El Análisis de sentimientos (AS) recibe muchos nombres dado a la evolución histórica que ha tenido: Análisis de Opiniones, Minería de Opiniones, etc. A partir de ahora y durante todo el proyecto nos referiremos a él como Análisis de sentimientos.

Definimos el Análisis de sentimientos como el estudio computacional de opiniones, sentimientos y emociones expresadas en un texto. Puede ser considerado como un proceso de clasificación cuyo fin es decidir la polaridad del sentimiento expresada en el texto. La dificultad de esta ciencia recaerá en el objeto de estudio pues se trata de información desestructurada, teniendo que convertirlos en datos estructurados.

2.2. CONCEPTO DE ANÁLISIS DE SENTIMIENTOS

2.2.1 Concepto de Opinión

Como hemos comentado en la sección anterior, el objetivo del AS consiste en la extracción de información a partir de textos. Fundamental para definir ese concepto será la definición de su objeto de estudio, el concepto de opinión.

Una primera aproximación para la definición de opinión la encontramos en el diccionario de la RAE, una opinión es un “juicio o valoración que se forma una persona respecto de algo o de alguien.” De esta definición podemos destacar varias cosas. Claramente una opinión es subjetiva, juicio o valor, pero también vemos que se hace énfasis en quién crea la opinión y sobre qué. Empezamos así a vislumbrar que la opinión estará formada por varios elementos.

Para hacer más énfasis en este concepto de opinión formada por varios elementos, veamos un ejemplo concreto:

[1] El pasado jueves fui con un grupo de amigos a cenar a un restaurante. [2] Pedí unos macarrones a la carbonara que estaban deliciosos. [3] Sin embargo, el postre no me gustó nada, pedí una tarta de queso. [4] Creo que en este sentido acertó más mi amiga María que pidió la tarta de tres chocolates, ¡le encantó! [5] En cuanto al precio, no estaba nada mal para la cantidad que servían. [6] También hay que tener en cuenta que se encontraba a las afueras, solo se puede acceder en coche.

Si observamos el texto anterior con la intención de extraer una opinión global sobre el restaurante, encontramos que se da información “contradictoria” pues hay oraciones que expresan sentimientos negativos, positivos e incluso neutros. La oración [1] expresa un hecho, sin ningún juicio de valor. Por su parte, las oraciones [2], [4] y [5] expresan sentimiento positivo mientras que las oraciones [3] y [6] expresan alguna opinión negativa. Vemos que cada una de estos juicios se realizan a ciertos elementos del restaurante, por lo que será algo a tener en cuenta dependiendo de si queremos determinar opinión de la entidad (restaurante) o de cada uno de sus componentes. Otro detalle presente en el ejemplo es que no siempre la persona que juzga algo es el autor del texto. En este caso, en la frase [4] el autor expresa que a su amiga le encantó su postre. Destacar también la importancia del tiempo a la hora de realizar un juicio de valor pues nuestra opinión puede variar a lo largo del tiempo.

En este punto, podemos considerar la opinión como una cuádrupla $(o, p, h,$

t) donde o representa el objeto sobre el que recae la opinión, p la polaridad de la misma, h el autor y t el tiempo en el que se produjo.

Si nos fijamos en la frase [3] que expresa un sentimiento negativo, vemos que la opinión negativa no es sobre el restaurante en su conjunto si no que es hacia un postre concreto. Según la definición de opinión como una cuádrupla (o, p, h, t) su polaridad recaería sobre el objeto en cuestión, el restaurante. Así, surge la necesidad de considerar que el objeto pueda ser dividido en diferentes componentes. Definimos como entidad e a un producto, servicio, tema, persona, organización o ente. Esta entidad se representa mediante (T, W) donde T es una jerarquía de componentes y W el conjunto de atributos de e .

Con esta nueva consideración llegamos a nuestra definición de opinión, introducida en [Liu, 2012]. Una opinión estaría formada por una quintupla $(e_i, a_{ij}, p_{ijkl}, h_k, t_l)$ donde e_i representa la entidad, a_{ij} sería el atributo de e_i sobre el que recae la opinión, p_{ijkl} se correspondería con la polaridad de la opinión que se hace sobre el aspecto a_{ij} en el momento de tiempo t_l por la persona h_k y, por último, h_k y t_l la persona y el tiempo en que se realiza el juicio de valor respectivamente.

2.2.2 Niveles de clasificación

Como ya hemos introducido, el objetivo final del Análisis de sentimientos es la clasificación de un texto según su polaridad. Ahora bien, a la hora de clasificar la polaridad de un texto podemos considerar tres niveles:

1. **Nivel de texto:** Analiza el documento completo y asigna una polaridad global al documento en su conjunto. Es el nivel más general y, por tanto, menos preciso.
2. **Nivel de sentencias:** Detecta el sentimiento en frases. Está muy relacionado con la subjetividad ya que esta se expresa en frases y no en documentos largos. Obtendríamos una clasificación más precisa del sentimiento.
3. **Nivel de aspectos o entidades:** Detecta el sentimiento u opinión y la entidad a la que se refiere. Es el nivel más preciso de análisis, pues no produce generalizaciones si no que se identifica una polaridad con un aspecto concreto.

2.2. CONCEPTO DE ANÁLISIS DE SENTIMIENTOS

Como ya hemos comentado, nuestro proyecto se centrará en el último de estos niveles de clasificación. La importancia de esta especificación queda reflejada en el ejemplo que pusimos en la sección 2.2.1. Además, una vez tengamos clasificada la polaridad de los diferentes aspectos o entidades, es fácilmente extrapolable al resto de niveles de clasificación. Podemos obtener la clasificación a nivel de sentencias simplemente con la combinación de la polaridad de cada uno de los aspectos que aparecen en la sentencia y análogamente con la clasificación a nivel de texto.

Bibliografía

- [Daniel Bestard Delgado] ¿Cómo se explica el crecimiento del BIG DATA en la última década? ¿Qué retos nos ha planteado esta ciencia?
- [Liu, 2012] Sentiment analysis and opinion mining. Synthesis lectures on human language technologies. Liu, B. (2012).