



Universidad de Granada

decsai.ugr.es

Modelos de ciencia de datos no numéricos: Aplicaciones en redes sociales, web y gestión de procesos



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**



Universidad de Granada

decsai.ugr.es

Bloque II: Minería de Texto y de la Web



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**



Universidad de Granada

decsai.ugr.es

Minería Web



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

Minería Web

Descubrimiento de patrones en las relaciones entre documentos hipertexto y los links (*web structure mining – link mining*)

Descubrimiento de contenido en documentos web (*web content mining*)

Descubrimiento de patrones en los accesos a servidores y navegación en sitios web (*web usage mining*)

Minería Web de la Estructura

- Se analiza la estructura de un sitio web
- Se tienen en cuenta relaciones:
 - *intra-páginas*: estructura de la página en HTML o XML para extraer información
 - *inter-páginas*: se estudia la relación mediante links entre diversas páginas para identificar sus relaciones y popularidad

Ejemplo: El algoritmo PageRank de Google ayuda a determinar la relevancia y la importancia de las páginas web basándose en los enlaces que reciben

Herramienta DNS Checker

<https://dnschecker.org> › pagerank



Universidad de Granada

decsai.ugr.es

Minería Web del Contenido



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

Minería Web del Contenido

- Tiene los mismos principios que la recuperación de información(multimedia) basada en contenido.
P.ej. Dame todas las imágenes/documentos sobre coches
- Es básicamente igual al descubrimiento de asociaciones en text mining, sólo que con documentos de la web.
- Incluye la minería de opiniones y sentimientos

Preprocesamiento en Minería web del contenido

Web Scraping: Proceso de extraer datos de sitios web mediante el rastreo de sus páginas y la extracción de información relevante.

Ejemplo: Para analizar las opiniones sobre un producto de Amazon, mediante web Scraping podemos extraer los comentarios de los usuarios, el número de estrellas y la fecha de publicación.

Preprocesamiento en Minería web del contenido

Herramientas para web scraping:

- BeautifulSoup y Scrapy: Software en Python para hacer scraping
- API: Muchas plataformas web ofrecen APIs públicas que proporcionan acceso estructurado a sus datos.
- Herramientas online de web scraping:
 - Easy scraper (extensión de Chrome)
 - ParseHub
 - Octoparse



Universidad de Granada

decsai.ugr.es

Minería de Opiniones y Sentimientos



DECSAI

**Departamento de Ciencias de la
Computación e Inteligencia Artificial**

Análisis de Sentimientos

El análisis de sentimientos, también conocido como minería de opiniones o sentimientos, es un campo de estudio que analiza las opiniones, sentimientos, evaluaciones, actitudes y emociones de las personas respecto a entidades tales como productos, servicios, organizaciones, individuos, problemas, temas, eventos, y sus atributos.

La investigación se inicia alrededor del año 2000, pero es con el reciente auge de las redes sociales cuando ha conseguido adquirir realmente relevancia, dado el elevado volumen de datos de opinión presentes.

La investigación en análisis de sentimientos ha tenido un gran impacto en el procesamiento de lenguaje natural (NLP), pero también en otros ámbitos como:

- Ciencias políticas
- Ciencias económicas
- Ciencias sociales

En definitiva, en todos aquellos en los que se tiene en consideración la opinión de las personas

Niveles de Análisis

El análisis de sentimientos se ha investigado a tres niveles:

Nivel de Documento: Clasificar si la opinión expresada en un documento es positiva o negativa. Se aplica cuando el documento hace referencia a una entidad única, no a varias a lo largo del texto.

Nivel de Sentencia: Determina si una sentencia expresa una opinión positiva, negativa o neutral (sin opinión).

Entidades y Aspectos/Características: El análisis a nivel de sentencia o documento no determina exactamente que es lo que gusta o no gusta.

El nivel de Aspecto realiza un análisis más detallado a nivel de características.

Se analiza la opinión en sí misma, abstrayendo de las estructuras del lenguaje (documentos, párrafos, frases...).

Se basa en la idea de que una opinión consiste en un sentimiento (positivo o negativo) y un objetivo (de la opinión).

Se usa para descubrir sentimientos en entidades y/o sus aspectos características.

Tipos de Opiniones I

Existen dos tipos de opiniones:

Opiniones regulares (regular opinions):

Expresa un sentimiento sobre una entidad o un aspecto de ésta.

“Me gusta la Coca Cola”.

“Odio la Pepsi Cola”.

- **Opinión directa (direct opinion):** Se expresa directamente sobre una entidad o un aspecto de ésta.
- *“Esta pastilla tiene un sabor horrible”.*
- **Opinión indirecta (indirect opinion):** Es una opinión expresada de forma indirecta sobre una entidad o alguno de sus aspectos, basándose en los efectos sobre otra entidad.
- *“Tras tomar la pastilla, el dolor de cabeza empeoró”.*

La mayor parte de la investigación en este campo se centra en opiniones directas. Son más fáciles de manejar que las indirectas, que tienen mayor complejidad.

Tipos de Opiniones II

Opiniones comparativas (comparative opinions):

Comparan varias entidades en bases a aspectos que comparten.

“La Coca Cola tiene mejor sabor que la Pepsi Cola”.

Una opinión comparativa habitualmente se expresa usando la forma comparativa o superlativa de un adjetivo o adverbio, aunque no siempre.

Las opiniones son **subjetivas**, y por tanto es necesario analizar opiniones de un amplio grupo de personas para hacerse una idea adecuada de la opinión general. La gran cantidad de opiniones generadas hace necesario el uso de mecanismos para resumirlas.

Tipos de Opiniones III

Opinión explícita (explicit opinion): Expresa una opinión subjetiva regular o comparativa.

“La Coca Cola tiene un gran sabor”.

Opinión Implícita (implicit opinion): Es una sentencia objetiva que implica una opinión regular o comparativa.

“La batería de Apple dura más que la de los Nokia”.

Las opiniones explícitas son más fáciles de detectar y clasificar que las implícitas.

La investigación actual se centra en el uso de opiniones explícitas, aunque también hay trabajos sobre opiniones implícitas.

Palabras de Sentimiento/Opinión I

Son palabras que se usan habitualmente para expresar sentimientos positivos o negativos.

Positivo : Bueno, sorprendente, excelente,...

Negativo: Malo, pobre, terrible,...

Además de palabras individuales, también tenemos frases o dichos populares.

“Cuesta un riñón”.

“Esto es la caña”.

Las palabras de sentimiento y las frases, se organizan en un lexicon (sentiment lexicon / opinion lexicon).

Este Lexicon es **necesario pero no suficiente** para realizar análisis de sentimientos.

Palabras de Sentimiento/Opinión II

Problemas:

1. Una palabra de sentimiento puede tener una orientación positiva o negativa según el **dominio de aplicación**.

“La pantalla es demasiado brillante”.

pero también puede ser positivo

“Este libro es brillante”.

2. Una sentencia que contiene palabras de sentimiento puede que **no exprese ningún sentimiento**.

“Recomiéndame un buen refresco”.

3. El **sarcasmo** es difícil de procesar esté o no acompañando de palabras de sentimiento.

“Un móvil estupendo. A la semana dejó de funcionar”

4. Algunas sentencias **sin palabras de sentimiento** puede expresar opiniones.

“Este móvil consume mucha batería”.

Opinion Spammers

El anonimato en las redes sociales permite a cualquier persona en cualquier lugar expresar opiniones sin miedo a represalias.

Esto hace realmente valiosas esas opiniones, pero también puede ser usado de forma malintencionada para desacreditar un producto, servicio u organización.

La detección de spammers de opinión no es un problema sólo de NLP, ya que implica el análisis del comportamiento en la publicación de mensajes. Esto lo convierte en un problema de Data Mining.

Hay tres tipos de Spam:

- **Reseñas Falsas:** No son sinceras y se escriben con algún motivo oculto, como promocionar un producto, o dañar su reputación (bots).
- **Reseñas sólo sobre marcas:** No comentan productos específicos. Aunque pueden ser auténticas, se consideran spam al no centrarse en productos.
- **No-Reseñas:** No son reseñas en realidad, suelen ser o bien anuncios o textos irrelevantes que no contienen opinión.

Definición de Opinión

Una opinión (regular opinion) es una **quíntupla**:

$$(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$$

e_i : el nombre de una entidad.

a_{ij} : un aspecto de la entidad e_i

s_{ijkl} el sentimiento sobre el aspecto a_{ij} de la entidad e_i ,

h_k : el sujeto que emite la opinión (opinion holder).

t_l : tiempo en el que el sujeto h_k expresó la opinión.

El sentimiento s_{ijkl} puede ser positivo, negativo o neutral, y puede expresarse con diferentes niveles de intensidad (escala de puntuación decimal, estrellas en una reseña, pulgar arriba/abajo).

Cuando se habla de la opinión de la entidad en conjunto se usa un aspecto especial denominado GENERAL para indicarlo.

En este caso e_i y a_{ij} juntos representan el objetivo del opinión.

Definición de Entidad

Una entidad es un producto, servicio, tema, problema, persona, organización o evento.

Se describe como un par:

e: (T, W),

T : una jerarquía de partes y subpartes, etc...

W : un conjunto de atributos de e.

La entidad como jerarquía de un número indeterminado de niveles, necesita de una relación anidada para representarla. Para simplificar se representan dos niveles y se usa el término aspecto para denotar a ambos elementos, partes y atributos.

En el árbol simplificado, el nodo raíz es la entidad, pero el segundo nivel (hojas) representa distintos aspectos de la entidad.

Esta representación simplificada es la que se usa habitualmente en la implementación de sistemas de análisis de sentimientos

Definiciones de Entidad

Entidad Categórica (entity category): Representa una entidad única.

Expresión de Entidad (entity expression): Palabra o frase que aparece en el texto indicando una entidad categórica.

Cada entidad categórica (o simplemente entidad) debería de tener un ***nombre único*** en una aplicación.

El proceso de agrupar expresiones de entidades en entidades categóricas se denomina ***categorización de entidades***.

Definiciones de Expresiones de Aspectos

Aspecto Categórico (aspect category): Representa un único aspecto de la entidad.

Expresión de Aspecto (aspect expression): Palabra o frase que aparece en el texto indicando un aspecto categórico.

- **Explícita** (explicit aspect expression): Son nombres y frases nominales.
 - “Buen precio”
- **Implícita** (implicit aspect expression): No son nombres o frases nominales. Estas expresiones suelen ser adjetivos o adverbios que se usan para describir o cualificar aspectos específicos.
 - “Barato” → Precio

Modelos

Modelo de entidad:

Una entidad e_i es representada como un todo y como un conjunto finito de aspectos $\{a_{i1}, a_{i2}, \dots, a_{in}\}$.

e_i puede expresarse con cualquier elemento del conjunto finito de sus expresiones de entidad $\{ee_{i1}, ee_{i2}, \dots, ee_{is}\}$.

Cada aspecto a_{ij} de una entidad e_i puede expresarse con cualquier elemento de un conjunto finito de expresiones de aspecto $\{ae_{ij1}, ae_{ij2}, \dots, ae_{ijm}\}$

Modelo de un documento de opinión:

Un documento d contiene opiniones sobre un conjunto de entidades $\{e_1, e_2, \dots, e_r\}$ y un subconjunto de sus aspectos, dadas por un conjunto de opinadores $\{h_1, h_2, \dots, h_p\}$ en un tiempo específico.

Tareas I

Extracción todas las quintuplas de un conjunto de documentos de opinión D :

Tarea 1 (extracción de entidades y categorización): Extraer todas las expresiones de entidad en D , categorizar expresiones de entidad en grupos o categorías que representen una entidad e_i .

Tarea 2 (extracción de aspectos y categorización): Extraer todas las expresiones de aspectos de las entidades, categorizar expresiones de aspectos en grupos o categorías que representen un único aspecto a_{ij} .

Tarea 3 (extracción de opinadores y categorización): Extraer los opinadores del texto y categorizarlos.

Tarea 4 (extracción de tiempo y estandarización): Extraer el tiempo en el que se efectúa la opinión y estandarizar los diferentes formatos.

Tareas II

Tarea 5 (clasificación de sentimientos de aspectos): Determinar si la opinión sobre un aspecto a es positiva, negativa o neutra, o asignar un valor numérico al sentimiento calificando el aspecto.

Tarea 6 (generación de las quintuplas de opinión): Extraer las quintuplas de un documento d , en base a las tareas anteriores.
Aunque aparentemente esta tarea parece sencilla, es la más difícil en muchos casos

El análisis de sentimientos (o minería de opinión) basado en este método se conoce como análisis de sentimientos basados en aspectos (aspect-based sentiment analysis), o análisis de sentimientos basado en características (feature-based sentiment analysis).

Ejemplo de las Tareas I

Posted by: bigJohn; Date: Feb. 15, 2023

- (1) I bought a Samsung camera and my friends brought a Canon camera yesterday.
- (2) In the past week, we both used the cameras a lot.
- (3) The photos from my Samy are not that great, and the battery life is short too.
- (4) My friend was very happy with his camera and loves its picture quality.
- (5) I want a camera that can take good photos.
- (6) I am going to return it tomorrow.

Ejemplo de las Tareas II

[Liu, 2012]

Task 1 **Task 2** **Task 3** Task 4 Task 5

Posted by: **bigJohn**; Date: **Feb. 15, 2023**

- (1) I bought a **Samsung** camera and my friends brought a **Canon** camera yesterday.
- (2) In the past week, we both used the cameras a lot.
- (3) The **photos** from my **Samy** are not that great (-), and the battery life is short (-) too.
- (4) **My friend** was very happy with **his** camera (+) and loves **its picture** quality (+).
- (5) I want a camera that can take good **photos**.
- (6) I am going to return it tomorrow.

Task 6

- (Samsung, picture_quality, negative, bigJohn, Feb-15-2023)
- (Samsung, battery_life, negative, bigJohn, Feb-15-2023)
- (Canon, GENERAL, positive, bigJohn's_friend, Feb-15-2023)
- (Canon, picture_quality, positive, bigJohn's_friend, Feb-15-2023)

Resumen de Opiniones

Las opiniones son **subjetivas**, luego una opinión no es suficiente para tomar una decisión.

En muchas aplicaciones se debe de tener en cuenta la opinión de una **multitud de usuarios**. Por tanto, es necesario resumir las opiniones obtenidas.

El resumen puede ser **estructurado**, o un pequeño **texto**.

El resumen debe hacer referencia a las distintas entidades y sus aspectos e incluir una perspectiva **cuantitativa**. No es lo mismo tener un 20% de opiniones positivas que un 80%.

Las quintuplas de opinión son una buena fuente de información para hacer resúmenes cualitativos y cuantitativos.

Una forma de resumen habitual se basa en aspectos (aspect-based opinion summary)

Subjetividad y Opinión

Subjetividad en una sentencia (sentence subjectivity): Una sentencia **objetiva** presenta información sobre hechos del mundo real, mientras una sentencia **subjetiva** expresa sentimiento personales, puntos de vista o creencias.

Las expresiones subjetivas se presentan de diversas formas, opiniones, alegaciones, deseos, creencias, sospechas, especulaciones, etc...

Existe cierta confusión entre los investigadores a las hora de distinguir subjetividad y 'con opinión'.

'Con opinión', hace referencia a un documento o sentencia que expresa un sentimiento positivo o negativo. **Una sentencia subjetiva puede que no muestre ninguna opinión.**

Ambos conceptos no son equivalentes, aunque tienen una intersección amplia.

La tarea de determinar si una sentencia es subjetiva u objetiva se denomina clasificación de la subjetividad (subjectivity classification)

Subjetividad y Emoción I

Emoción: Sentimientos y pensamientos subjetivos.

No existe un conjunto de emociones básico consensuado entre los investigadores.

El psicólogo Parrott, clasifica seis emociones primarias: *amor, felicidad, sorpresa, rabia, tristeza, y miedo*, que pueden subdividirse en otras emociones secundarias y terciarias. Cada emoción puede tener diferentes intensidades.

Las emociones están muy relacionadas con los sentimientos. La fuerza de un sentimiento o una opinión se conectan con la intensidad de ciertas emociones como felicidad o rabia.

Las opiniones que se tratan en análisis de sentimientos, son principalmente evaluaciones (aunque no siempre).

Las evaluaciones se pueden categorizar en dos tipos : **racionales** y **emocionales**.

Subjetividad y Emoción II

Evaluación Racional: Son evaluaciones razonadas, sobre elementos tangibles.

“La voz del teléfono se escucha nítida”.

Evaluación Emocional: Son evaluaciones no tangibles y respuestas emocionales a entidades que afectan el estado mental de las personas

“Estoy cabreado con el servicio de atención al cliente.”

Para usar estas evaluaciones en la práctica, se determinan 5 calificaciones de sentimientos:

- Emocional negativo (-2), Racional Negativo (-1).
- Neutral (0).
- Racional Positivo (+1), Emocional Positivo (+2).

Los conceptos de emoción y opinión no son equivalentes. Las opiniones racionales no expresan emociones, y las sentencias con emoción pueden no expresar sentimientos.

Clasificación de Documentos

Dada un documento de opinión d evaluando una entidad, se debe determinar el sentimiento general del emisor de la opinión sobre la entidad.

Se trata de determinar s sobre el aspecto GENERAL en la quintupla:

$(_, \text{GENERAL}, s, _, _)$,

donde la entidad e , el emisor de la opinión h y el tiempo de opinión t se asumen conocidos o irrelevantes.

Si se emplean *valores categóricos* es un *problema de clasificación*.

Si se usan *valores numéricos u ordinales*, se convierte en un *problema de regresión*.

Para facilitar la tarea se asume:

El documento de opinión d expresa opiniones sobre una única entidad e y contienen opiniones de un único emisor de opinión h .

Clasificación de Sentimientos con Aprendizaje Supervisado I

La clasificación de sentimientos es un problema de clasificación de texto. En lugar de clasificarlos en temas, se clasifican en dos clases, positivo y negativo.

Los datos de entrenamiento y prueba empleados son reseñas de productos.

Entre los distintos métodos empleados se encuentran:

- Clasificación naïve Bayes
- Máquinas de vector de soporte (SVM).
- Deep Learning

Se usan distintas características para la clasificación.

Términos y su frecuencia: Se trabaja con palabras (unigramas) y n-gramas y sus frecuencias. En ocasiones puedes usarse la posición, o un esquema de pesos como TF-IDF. Los resultados obtenidos han sido muy buenos.

Clasificación de Sentimientos con Aprendizaje Supervisado II

Categoría Gramatical (POS): El POS es importante a la hora de detectar opiniones, se ha demostrado que *los adjetivos son buenos indicadores de opiniones*.

Palabras y Frases de Sentimiento: La mayor parte de palabras de sentimiento son adjetivos, pero también pueden serlo nombres (*basura, maravilla*) y verbos (*odiar, gustar*). Además de palabras individuales también pueden aparecer frases hechas.

Reglas de opinión: Existen expresiones y estructuras del lenguaje que permiten expresar opiniones y sentimientos.

Modificadores de Sentimiento (Sentiment shifters): Expresiones que pueden cambiar la orientación del sentimiento de positivo a negativo y viceversa.

Dependencias Sintácticas: Algunos investigadores han empleado también características basadas en la dependencia de las palabras.

Clasificación de Sentimientos con Aprendizaje No Supervisado

Se usan palabras de sentimiento, para aprendizaje no supervisado.

[Turney, 2002] Propone una clasificación usando patrones sintácticos que se suelen emplear para expresar opiniones, estos patrones están formados por las etiquetas POS de las palabras.

[Ding, Liu and Yu, 2008; Hu and Liu, 2004; Kim and Hovy, 2004] Emplean diccionarios de palabras y frases de sentimiento con sus orientaciones e intensidades. Trabajan la clasificación de sentimientos a nivel de frase y aspecto.

[Taboada et al., 2011] Usa la anterior aproximación incorporando intensificación y negación para calcular un valor de sentimiento para cada documento.

Clasificación de Sentimientos con Aprendizaje No Supervisado

El algoritmo de Turney, consta de tres fases:

- 1) Se obtienen pares de palabras cuyo POS se ajusta a alguno de los patrones propuestos de tres palabras (en la que la tercera no se extrae).
- 2) Se calcula la orientación del sentimiento con la medida pointwise mutual information (PMI), que determina el grado de independencia estadística entre dos términos. La orientación del sentimiento (SO) se determina al buscar el grado de asociación con las palabras de referencia “excellent” y “poor”. Las probabilidades se obtienen a través del número de resultados devueltos por una consulta a un buscador.
- 3) En una reseña, se calcula la orientación del sentimiento para todas las frases, y se calcula la media, si el valor de la media es positivo se considera positiva, de forma análoga para la media negativa.

Clasificación de Sentimientos

Predicción de la Calificación de Sentimientos

Además de la clasificación en positivo/negativo se puede establecer una puntuación (1-5 estrellas) para las reseñas.

En este caso el problema se formula como un problema de regresión ya que la puntuación es ordinal.

Clasificación Inter-dominios

La misma palabra en distintos dominios puede tener orientaciones diferentes.

Clasificación multilingüe

Se aplica a nivel de documento y de frase.

Se suele emplear traducción automática para usar las herramientas disponibles en inglés.

Sentimiento a Nivel de Sentencia

La clasificación por sentimiento a nivel de documento, puede ser bastante rudimentaria para algunas aplicaciones.

En ocasiones es más apropiado *trabajar a nivel de sentencia*.

No existe una gran diferencia entre clasificar el sentimiento de un documento o una sentencia, ya que podemos ver las sentencias como documentos breves.

Se suele hacer la suposición de que una sentencia contiene una única opinión (aunque no siempre es así) y que un documento contiene varias opiniones.

En el análisis a nivel de sentencia se asume que una sentencia contiene un único sentimiento de un único emisor de opinión.

Sentimiento a Nivel de Sentencia I

El análisis de sentimientos a nivel de sentencia se puede resolver como:

Problema de clasificación con tres clases:

Clasificamos en las clases positivo/neutro/negativo

Dos problemas de clasificación:

- ***Clasificar si una sentencia tiene opinión o no.***
- Clasificación de la subjetividad, clasifica las sentencias en dos clases objetivas/subjetivas.
- La mayoría de aproximaciones empleadas se basan en aprendizaje supervisado.
- ***Clasificar sentencias con opinión como positivas o negativas.***

Sentimiento a Nivel de Sentencia II

La subjetividad se puede clasificar en cuatro clases:

- **S:** Subjetiva y con opinión.
- **SN:** Subjetiva, pero sin opinión.
- **O:** Sentencia objetiva sin opinión.
- **OO:** Sentencia objetiva, con sentimiento negativo.

Una sentencia puede contener tanto partes subjetivas como objetivas.

Para una sentencia clasificada como subjetiva, podemos determinar si expresa una opinión positiva o negativa.

Se puede aplicar aprendizaje supervisado como en la clasificación a nivel de documento, y también se pueden aplicar métodos basados en lexicon.

Otras Sentencias

Sentencias condicionales: Describen implicaciones o situaciones hipotéticas y sus consecuencias. Suelen contener dos cláusulas, la condición y el consecuente, que son independientes entre sí. La relación entre ambas va a determinar si la sentencia expresa un sentimiento positivo o negativo.

Se necesitan métodos específicos para sentencias condicionales.

Un método para resolver este problema, es usar características lingüísticas (palabras/frases de sentimiento, POS, patrones, conectivas condicionales) y aplicar aprendizaje supervisado.

Sarcasmo: A través de una sentencia, se transmite lo contrario de lo que se quiere decir.

El sarcasmo no es habitual en reseñas de productos, pero es muy habitual en discusiones on-line y comentarios sobre política.

Clasificación de Sentimientos Multilingüe

Existen multitud de trabajos sobre clasificación multilingüe de sentimientos y subjetividad a nivel de sentencia.

Estos trabajos abordan el problema utilizando recursos y herramientas disponibles en **Inglés**.

Se trabaja en tres líneas:

- 1) Traducir las **sentencias de test** del lenguaje **destino** al lenguaje de **origen** y clasificarlas utilizando un **clasificador** en el **lenguaje de origen**.
- 2) Traducir un **conjunto de entrenamiento** de un lenguaje de **origen**, al lenguaje de **destino** y construir un **clasificador basado en corpus** en el lenguaje **objetivo**.
- 3) Traducir un **lexicon** de sentimientos o subjetividad en el lenguaje de **origen** al lenguaje **destino** y construir un **clasificador basado en lexicon** en el lenguaje **destino**.

Clasificación de Sentimientos de Aspectos

Existen dos aproximaciones, aprendizaje supervisado y basado en lexicon.

Método Básico basado en lexicon

Asumimos que las entidades y aspectos son conocidos.

1) Marcar palabras y frases de sentimiento:

2) Por cada sentencia que contiene uno o más aspectos...

3) Cada palabra positiva tiene un valor de +1, negativas -1

4) Aplicar modificadores de sentimiento:

5) Palabras que cambian la orientación del sentimiento.

6) Tratar las cláusulas con 'pero-but':

7) Palabras o frases que indican lo contrario pueden cambiar la orientación del sentimiento.

8) Agregar opiniones:

- Se calcula la puntuación final agregada para determinar la orientación final del sentimiento en cada aspecto de la sentencia.

Extracción de Aspectos

Puede verse con una tarea de extracción de información.

La principal característica es que cada opinión siempre tiene un objetivo.

Algunas expresiones pueden desempeñar dos roles, por ejemplo, indicando un sentimiento positivo o negativo e implicando (implícito) un aspecto (objetivo).

“Este coche es muy caro”.

Caro implica sentimiento, y a su vez designa el aspecto precio.

Existen cuatro aproximaciones a la extracción de aspectos:

- 1) Extracción basada en nombre frecuentes y frases nominales
- 2) Extracción de relaciones entre opiniones y objetivos.
- 3) Extracción mediante aprendizaje no supervisado.
- 4) Extracción usando *topic modeling*.

Recursos

Lexicons:

SentiWordNet 3.0

Palabras de sentimiento en Inglés

<https://github.com/aesuli/SentiWordNet>

SentiWS

Palabras de sentimiento en Alemán

<https://wortschatz.uni-leipzig.de/en/download>

Datasets en inglés :

Pang & LEE movie reviews

<http://www.cs.cornell.edu/people/pabo/movie-review-data/>

SNAP - Web data: Amazon reviews

<http://snap.stanford.edu/data/web-Amazon.html>