

Técnicas para el análisis del sentimiento en Twitter: Aprendizaje Automático Supervisado y SentiStrength

Resumen: Paula Monge - B85066

El tema tratado en inglés ha sido bastante explorado, sin embargo el desarrollo de estas técnicas aplicadas al castellano se encuentra todavía en fase de maduración. El artículo tiene la siguiente estructura:

- Primero se introducirá el análisis de sentimiento.
- Segundo se presentarán las fases de la técnica basada en los algoritmos de aprendizaje automático supervisado. También se hablará de una iniciativa de la Sociedad Española para el Procesamiento del Lenguaje Natural

ANÁLISIS DE SENTIMIENTO EN TWITTER

El análisis de sentimientos en Twitter supone asignarle a cada tweet publicado un valor relacionado con la carga emocional que transmite.

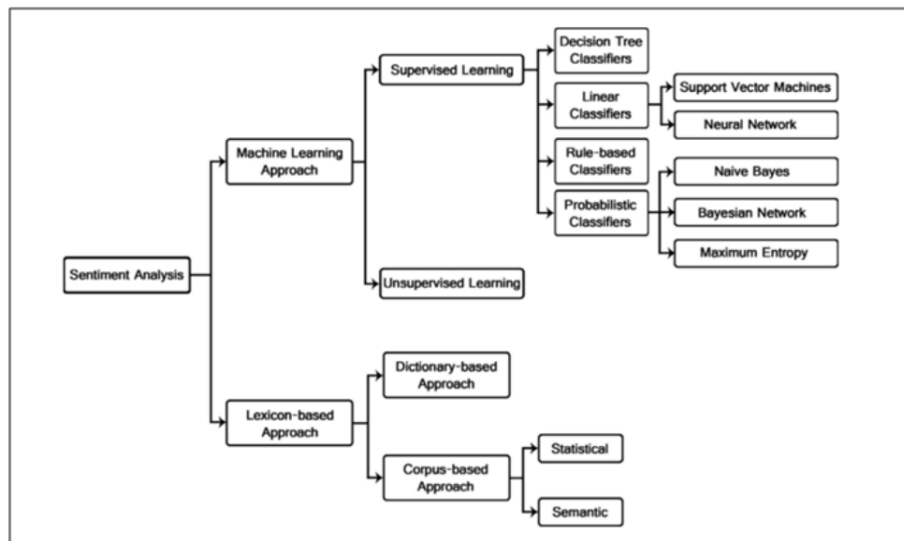
Tipos de cargas emocionales:

- Polaridad: indica si el mensaje tiene un sentimiento positivo o negativo. En algunos análisis se introduce una tercera categoría para clasificar los mensajes neutros.
- Intensidad: proporciona un valor numérico en relación con la intensidad del sentimiento. Se puede distinguir entre una intensidad positiva y una intensidad negativa.
- Emoción: clasifica el texto según los distintos tipos de emociones, como puede ser la alegría, la tristeza o la ira.

Este tipo de análisis de sentimientos de cualquier texto no siempre resulta algo unívoco, ya que este puede variar si se realiza de forma manual o con el tipo de codificador que se use.

Según Medhat et al. (2014), las principales técnicas de análisis de sentimiento se dividen en dos grandes grupos: las que se basan en aprendizaje automático (machine learning approach) y las que se basan en diccionarios (lexicon-based approach).

Figura 1: Tipología de las técnicas de Análisis de Sentimiento en Twitter



Fuente: Medhat et al. (2014).

Cuando se evalúa el sentimiento de un documento, se tiene en cuenta el conjunto completo de información.

ANÁLISIS DE SENTIMIENTO BASADO EN APRENDIZAJE AUTOMÁTICO: Máquina de soporte Vectorial

Cuando se habla de aprendizaje automático la máquina no se programa para que responda de una determinada forma según las entradas recibidas, sino más bien para que extraiga patrones de comportamiento a partir de las entradas recibidas, y en base a dicha información aprendida o asimilada, realice la evaluación de nuevas entradas. Hay dos tipos:

- **Supervisado:** el algoritmo usa un corpus con el cual lleva a cabo dos procesos: encontrar los mejores parámetros para el algoritmo, y evaluar el nivel de fiabilidad con esos parámetros. A esta fase se le llama aprendizaje o entrenamiento.
Un punto importante de este proceso es la elaboración del **corpus** (de test y otro de development) para el entrenamiento. Conviene que sea lo más voluminoso posible y que sea representativo del conjunto de datos que se quiere analizar, este tiene un doble objetivo: ajustar el algoritmo y evaluar su fiabilidad.
- **No supervisado:** realizan el procesamiento en base únicamente a las entradas.

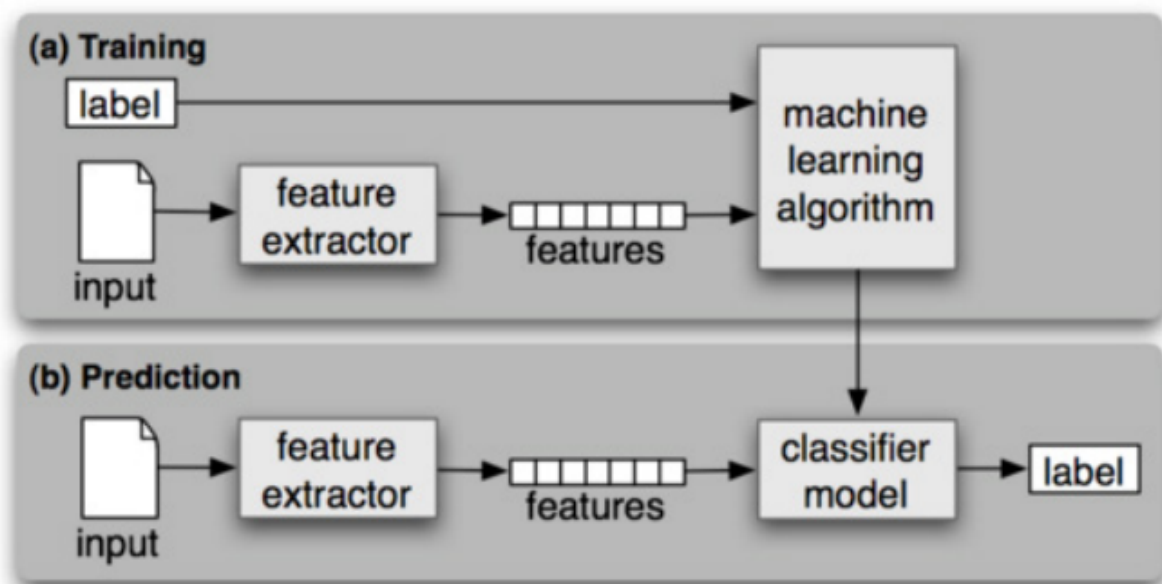
Para hacer análisis de sentimientos es mejor hacer uso de aprendizaje supervisado ya que se obtiene más fiabilidad del resultado, estos pueden ser:

- **Regresión:** proporcionan una salida numérica.
- **Clasificación:** asignan a la entrada una categoría proveniente de un conjunto limitado de elementos.

Cada clasificador que se crea requiere de un contexto, por lo que un algoritmo que define la polaridad de los tweets en cuestión de política requeriría ser ajustado y entrenado para poder funcionar con tweets de deporte, es por esto que es tan importante tener un corpus bien realizado.

Fases del algoritmo.

Figura 2: Fases del AAS aplicado a la clasificación



Fuente: Bird et al. (2009).

Un elemento en común entre la fase de aprendizaje y de clasificación es el extractor de atributos (evalúa una serie de características)

Una vez extraídas, se le pasan al algoritmo, el cual hará distintas funciones dependiendo de la fase en que se encuentre: los atributos extraídos de la entrada servirán para entrenar el algoritmo en la fase de aprendizaje, mientras que en la fase de clasificación serán utilizados para categorizar la entrada recibida.

En esta fase también se incluye el preprocesamiento del tweet ya que los atributos no suelen extraerse directamente del texto del mensaje ya que es necesario corregir las expresiones coloquiales de los textos.

Un ejemplo de este preprocesamiento puede verse en el trabajo de Yu y Wang (2015). Algunos de las acciones que llevaron a cabo fueron: eliminar las URL del mensaje, pues no aportan información de sentimiento; tokenizar o extraer las palabras del tweet; pasar a minúsculas, y quitar las stopwords o palabras vacías, esto es, aquellas palabras que no aportan información de sentimiento como los artículos y las preposiciones.

Lo más habitual en el preprocesamiento es quitar las stopwords y lematizar (determinar la forma de una palabra que se constituye en lema.). La unificación de las formas léxicas supone identificar términos con el mismo significado semántico, como sería el caso de una palabra en singular y otra en plural, o las diferentes formas verbales. Por ejemplo: la palabra “comerían” pasaría a ser “comer”, de tal forma que se identifica un mismo sentido semántico en todas las formas verbales de “comer” que aparezcan en el corpus. Todas estas herramientas pertenecen al ámbito del Procesamiento de Lenguajes Naturales (PLN; en inglés Natural Language Processing, NLP).

Existen diversidad formas de extraer los atributos, puede ser palabra por palabra o un conjunto de 2 o 3 palabras juntas, haciendo que la cantidad de atributos disminuya, esta forma de extraer atributos condiciona el tipo de algoritmo a utilizar. Como entrada al algoritmo, tendremos un listado de pocos atributos por cada tweet en relación a todos los atributos posibles, es decir, la extracción conformará una matriz de datos en la que abundarán los valores nulos. Pues bien, un algoritmo muy apropiado para tratar este tipo de datos es el algoritmo de **Máquinas de soporte Vectorial**.

Básicamente, el algoritmo construye un enorme conjunto de hiperplanos de tal forma que maximiza los puntos de separación entre las muestras y facilita la clasificación de las nuevas entradas.

La forma de medir la precisión es accuracy, que evalúa el porcentaje de entradas del corpus de prueba que han sido clasificadas correctamente por el algoritmo.

Proceso de extracción de atributos (separación por polaridad) :

1. Se consideraron solo unigramas de lemas con una frecuencia mínima preestablecida. Un unigrama equivale a una palabra del mensaje del tweet.
2. Los hashtag, las menciones a usuarios, los números, las fechas y los signos de puntuación fueron unificados respectivamente en una sola característica.
3. Se sustituyeron los emoticonos por su correspondiente categoría: happy, sad, tongue, wink y otros.
4. Se excluyeron los términos pertenecientes a ciertas categorías morfosintácticas poco significativas para el análisis de sentimiento.

5. Se utilizó como recurso externo varios diccionarios de polaridad.

Se obtuvo un accuracy de 72% con el uso de MSV en castellano, por lo que se concluye que este algoritmo no fue tan óptimo como se pensó.

ANÁLISIS DE SENTIMIENTO BASADO EN DICCIONARIOS: Sentistrength

En este contexto, un diccionario consiste en un listado de términos (pueden ser palabras o multipalabras) que tienen asociados una determinada orientación de sentimiento, este detecta coincidencias con los diccionarios y articula un modo de evaluar el sentimiento en base al número de concurrencias encontradas.

Este elemento depende fuertemente del idioma en que se escriba el tweet. En inglés están disponibles varios diccionarios con un alto nivel de elaboración que proporcionan información muy refinada sobre la orientación del sentimiento de los términos listados. En cambio, los diccionarios en castellano están más bien orientados hacia la polarización de las palabras.

SentiStrength proporciona dos valores por cada texto analizado: uno mide la intensidad del sentimiento positivo y el otro la intensidad del sentimiento negativo. Ambos índices varían de 1 a 5. Esta doble medida se corresponde con la presencia de emociones diferentes combinadas en una misma expresión.

SentiStrength tiene como base de trabajo una serie de archivos en formato de diccionarios:

- EmotionLookUpTable.txt: lista de palabras en la que cada una tiene asociada un número que refleja su intensidad de sentimiento. El valor numérico varía entre -5 y +5.
- EmoticonLookUpTable.txt: lo mismo que el anterior pero en lugar de palabras tiene emoticonos.
- IdiomLookUpTable.txt: lista de expresiones idiomáticas con su índice de sentimiento, p.ej. a duras penas.
- NegatingWordList.txt: lista de palabras de negación, que servirán para invertir el sentimiento de la palabra siguiente, p.ej. jamás, nadie.
- BoosterWordList.txt: lista de modificadores de la intensidad de sentimiento, p.ej. muy, extremadamente.

SentiStrength ofrece la posibilidad de ajustar los pesos de los términos para que se adapte mejor al contexto. Para ello, se requiere de un corpus codificado manualmente que haga la función de regla de oro. El manual de la herramienta recomienda que el volumen de este corpus sea superior a 1.000 textos y que esté validada por tres codificadores. Este proceso resulta un tanto sofisticada pues hay

que asignar a cada texto dos índices de intensidad de sentimiento, uno para el positivo y el otro para el negativo.

La evaluación de la herramienta con estos diccionarios fue llevada a cabo con un corpus de 1.600 tweets, el ajuste a los índices de sentimiento positivo fue de un 79,9% y el ajuste del sentimiento negativo fue de un 86,2%

Link a un diccionario en español:
<http://sentistrength.wlv.ac.uk/SpanishSentiDataDavidVilares.zip>.

DISCUSIÓN

. Los procesos de AAS pueden servirse de diccionarios de polaridad para extraer atributos de los tweets; mientras que el SentiStrength, que está basado en el uso de diccionarios, utiliza técnicas de AAS para ajustarse al contexto. Una primera consecuencia emerge ya del presente estudio: no conviene obviar uno a favor de otro, puesto que las técnicas básicas se utilizan en ambos enfoques.

El corpus tiene un papel clave

Una diferencia entre ambos planteamientos radica en la relación con el contexto. Los métodos que utilizan el AAS dependen del contexto, dado que necesitan un corpus de entrenamiento, en cambio, los diccionarios tienen una polaridad asociada a las palabras que permite usar estos métodos en diversos contextos.

Ahora bien, la evaluación del sentimiento puede ser un poco rígida en los métodos basados en diccionarios, puesto que depende exclusivamente del contenido semántico de los términos. En cambio, el AAS permite una mayor adaptación, dado que el valor añadido que aporta este tipo de procesos es la detección de patrones de palabras que transmiten un sentimiento particular. Lo que sí que conviene señalar en ambos métodos es la limitación para detectar correctamente expresiones irónicas.

Tabla 1: Comparación entre los métodos de análisis de sentimiento en Twitter

	PLN + AAS-MVS	SentiStrength
Base de trabajo	Procesos	Diccionarios
Habilidad del Investigador	Analista de Datos	Lingüística
Codificación manual	Corpus de Entrenamiento	Regla de Oro para el ajuste de pesos
Evaluación	Accuracy de la Clasificación	Precisión de la Validación
Adaptación	Flexible por el Entrenamiento	Rígida por los Diccionarios
Contexto	Dependiente	Independiente
Valor aportado	Detecta patrones no evidentes	Sencillez de uso
Dificultad	Alta	Moderada - Alta

A grandes rasgos, no se puede evaluar el uno contra el otro ya que ambos pueden estar mezclados, ambos tienen aciertos y dificultades a la hora de su aplicación, lo que se puede concluir es que ambos son efectivos, sin embargo no hay mucho desarrollo del análisis de sentimientos de tweets en español.