

## Introducción

→ Tema del artículo

**El tema del artículo es realizar un sistema de análisis de sentimientos para el reconocimiento automático de las emociones representadas en un texto, utilizando un conjunto de clasificadores.**

**Esto es debido a que, aunque existen muchas formas de expresar emociones, ya sea mediante texto, pinturas, etc, el método más utilizado para expresar las emociones es mediante textos escritos.**

→ Esquema de clasificador del conjunto

**Se basa en la noción de combinar métodos de clasificación de aprendizaje automático basados en conocimientos y estadísticos, con el objetivo de beneficiarse de sus méritos y minimizar sus inconvenientes.**

**Si observamos la imagen, vemos que el funcionamiento del clasificador del conjunto es el siguiente:**

- **Dado un texto (input text), se divide el texto en oraciones y se realiza un análisis de cada oración.**
- **Una vez realizado el análisis, se realiza bien una extracción de propiedades (feature extraction) o se aplica una herramienta basada en el conocimiento.**
  - **Si aplicamos la extracción de propiedades, se utilizan 2 clasificadores de enfoque estadístico. Para este sistema de análisis, se han utilizado tanto el Clasificador de Naive Bayes como el clasificador de Máxima Entropía.**
    - **El clasificador de Naive Bayes es un enfoque de clasificación basado en la probabilidad que asume que las palabras documentadas se generan a través de un mecanismo de probabilidad. Las unidades léxicas se etiquetan con una categoría particular o un conjunto de categorías y se procesan computacionalmente. Asume que las palabras son mutuamente independientes y, por lo tanto, cada palabra individual se supone que es una indicación de la emoción asignada. Se utiliza para calcular la probabilidad de una clase definida, basada en las características del documento.**
    - **El clasificador de Máxima Entropía es un modelo basado en característica que prefiere los modelos más uniformes que satisfacen una restricción dada, de modo que los datos etiquetados en la fase de entrenamiento se utilizan para derivar las restricciones para el modelo que caracterizan la clase. A diferencia del clasificador de Bayes, el clasificador de máxima entropía no asume la independencia por sus características.**

**Ambos clasificadores se alimentan del gran cuerpo de entrenamiento de textos sentimentales anotados, con el fin de ser entrenados con el objetivo de aprender el estado emocional y la fuerza de las palabras emocionales, así como tener en cuenta las fortalezas de otras palabras arbitrarias.**

- **La herramienta basada en conocimiento realiza un análisis profundo de las oraciones en lenguaje natural. A diferencia de los enfoques estadísticos, intenta analizar y extraer conocimiento de cada oración para especificar su estado emocional.**

**Como finalización, el sistema adopta un enfoque de votación, el cuál se verá más adelante.**

## HERRAMIENTA DE CLASIFICACIÓN BASADO EN CONOCIMIENTO

A continuación, veremos el funcionamiento de la herramienta de clasificación basada en conocimiento. Si vemos la imagen, observamos que esta herramienta realiza un análisis de las

oraciones y, aplicando una base de conocimiento, realiza una extracción de las emociones de las oraciones del texto para, de este modo, obtener las emociones del texto.

A continuación, se muestra un ejemplo de su funcionamiento para la oración representada en la imagen. Como podemos observar, se genera un árbol jerárquico donde el nodo raíz es el verbo de la oración y cada nodo representa una palabra de la oración. De este modo, se especifica las dependencias de las palabras y se determina la conexión de las palabras para la transmisión de emociones.

## SISTEMA DE RECONOCIMIENTO DE EMOCIONES

→ Reconocimiento de emociones:

**Se presenta el clasificador conjunto desarrollado para analizar el lenguaje natural y reconocer el contenido emocional del texto, ilustrándose su funcionalidad.**

**El análisis del lenguaje natural se lleva a cabo a nivel de oración, de modo que un documento se divide en oraciones.**

- **Los clasificadores de enfoque estadístico son entrenados para reconocer sentimientos en datos textuales.**
- **La herramienta basada en el conocimiento realiza un análisis profundo de la estructura del lenguaje natural.**

→ Representación de características

**Para la representación de un texto de lenguaje natural utilizamos la técnica de representación de bolsas de palabras (BOW).**

**En este enfoque, un documento se considera una colección desordenada de palabras y su posición no tiene importancia.**

**En el sistema:**

- **Una nueva oración originalmente se tokeniza y se divide en palabras.**
- **Luego, cada palabra se lematiza y se especifica su forma base.**
- **Las palabras de frenado de la oración se filtran y las características de la oración se reenvían a los clasificadores.**

→ Votación de clasificador

**El clasificador de conjuntos adopta un enfoque de votación por mayoría para tomar una decisión de clasificación basada en los productos de cada clasificador base.**

**De este modo, cada clasificador tiene un voto que, para cada oración, es una clase determinada por el clasificador de conjunto.**

**En general, el voto mayoritario cuenta los votos de cada clase sobre los clasificadores de entrada y selecciona la clase mayoritaria.**

**Teóricamente, si los clasificadores de base seleccionados pueden cometer errores independientes, se demuestra que el voto mayoritario es adecuado y puede superar al mejor clasificador.**

## ESTUDIO DE LA EVALUACIÓN

1. **Recopilación de datos: Se crea un cuerpo de datos textuales de diferentes fuentes y se utiliza un experto humano para anotar manualmente cada uno. Finalmente, se evalúa el contenido emocional de la oración en un rango del 0 al 100.**
2. **Resultados de la evaluación: Se lleva a cabo un estudio de la evaluación para evaluar el mecanismo desarrollado y proporcionar una visión de desempeño.**
  1. **Se evalúa el desempeño del sistema en el reconocimiento de las emociones**
  2. **Se evalúa el desempeño en el reconocimiento de la polaridad emocional de las oraciones.**

## CONCLUSIONES

- Rendimiento satisfactorio en cuanto a la capacidad de reconocer la presencia de emociones y para identificar la polaridad de las emociones.
- La técnica de ensamble es una manera efectiva de combinar diferentes algoritmos de clasificación para una mejor clasificación emocional, funcionando mejor en ambas tareas el esquema de conjunto que la única clasificación.

## POSIBLES TRABAJOS FUTUROS

- Aplicar una evaluación a mayor escala
- Ampliar la fase de formación de los clasificadores y las reglas de dependencia de la herramienta basada en conocimiento para enfrentar oraciones que expresan emociones en función de la persona.