

Evaluación del Módulo 6

Laura Rodríguez Navas
rodrigueznabas@posgrado.uimp.es

Marzo 2020

1. ¿Qué es la Categorización de Textos? (1 puntos).

La categorización de textos es un sistema que consiste en la asignación de una o más categorías preexistentes a cada uno de los documentos de una colección. El objetivo de un sistema de CT es decidir si un documento concreto pertenece o no a una categoría particular.

Métodos utilizados:

- Sistemas con Categorización de Textos Manual.
 - Muy precisos cuando el trabajo es realizado por expertos.
 - Viables cuando la colección es pequeña.
 - Difícilmente escalables.
- Sistemas con Categorización de Textos Automática.
 - Sistemas basados en reglas codificadas manualmente. Construir y mantener las reglas es muy costoso, pero se consigue una alta precisión si las reglas se refinan por un experto.
 - Sistemas basados en aprendizaje. Requieren de una colección de entrenamiento y otra de evaluación para probar la eficiencia del sistema. Es fácil de escalar y mantener usando los algoritmos: k-Nearest Neighbors, Naive Bayes, Support Vector Machines y Neural Networks.

2. Razone la diferencia entre clasificación multiclase y multietiqueta. Lectura recomendada: Sección Classification with more than two classes del libro Introduction to Information Retrieval (2 puntos).

La clasificación multiclase es una tarea de clasificación con más de dos clases; que parte del supuesto de que cada clase está asignada a una sola etiqueta. Y la clasificación multietiqueta también es una tarea de clasificación de más de dos clases pero con la diferencia de que asigna a cada clase un conjunto de etiquetas, es decir, una clase puede tener más de una etiqueta.

3. Defina la tarea Análisis de Opiniones (1 puntos).

La tarea de Análisis de Opiniones se centra en el desarrollo de metodologías destinadas al estudio de la opinión, a la identificación de los distintos elementos que la componen, y al descubrimiento de su orientación.

4. ¿Qué diferencia hay entre clasificación de la subjetividad y clasificación de la opinión? (2 puntos).

Las dos clasificaciones se basan en la clasificación binaria. Pero mientras, la clasificación de la opinión o de polaridad, determina si las opiniones son positivas o negativas (a favor o en contra), dado un documento, una oración, un rasgo o una característica, la clasificación de la subjetividad determina si las opiniones son objetivas o subjetivas.

Normalmente la clasificación de la subjetividad puede ser una subtask del análisis de opiniones más difícil de determinar que la clasificación de la opinión o clasificación de la polaridad.

5. ¿Cuáles son los elementos de una opinión? (2 puntos).

Si una opinión es un punto de vista, apreciación o actitud sobre un objeto u entidad expresado por un usuario u organización (opinion holder), los elementos de una opinión son:

- Objeto u entidad: sobre la que la opinión es expresada.
- Opinion holder: usuario u organización que expresa una opinión sobre un objeto u entidad.

6. ¿Cuáles son los niveles de análisis que se pueden desarrollar en el Análisis de Opiniones? (2 puntos).

Existen tres tipos de niveles de análisis que se pueden desarrollar en el Análisis de Opiniones:

- Análisis de opinión a nivel de aspectos y características.
- Análisis de opinión a nivel de frase.
- Análisis de opinión a nivel de documento.