



Máster de Tecnologías de Telecomunicación

Trabajo Fin de Máster

Clasificador Basado en Análisis de Sentimientos

Leandro Aarón López Rodríguez

José María Quinteiro González - Pablo Hernández Moreda

Septiembre 2011

Resumen:

El análisis de sentimiento trata de obtener la polaridad subyacente en texto libre, entendiendo por polaridad subyacente lo positivo o negativo que puede ser un texto. Este proceso también es conocido como minería de opinión ya que puede considerarse una rama de la minería de datos por la utilización que hace de sus herramientas y técnicas. En este Trabajo de Fin de Master vamos a considerar la polaridad de un texto para clasificarlo como positivo o negativo

Diseño del Clasificador

La arquitectura se basa principalmente en el patrón plantilla (Template Method). Este patrón de diseño define una estructura de herencia de manera que la superclase sirve de plantilla para los métodos en las subclases, permitiendo a las subclases redefinir ciertos pasos de un algoritmo sin cambiar su estructura.

Este proceso se ha dividido en 3 etapas, las cuales coinciden con las etapas clásicas del aprendizaje automático: carga de datos, procesamiento de datos y entrenamiento

Aproximaciones

Extracción de Atributos

Se extraen 21 atributos del texto a partir de un largo preprocesado de datos, estos atributos son:

- Frecuencia de nombre, verbos, adjetivos y adverbios (4 atributos)
- Polaridad de nombres, verbos, adjetivos y adverbios (12 atributos)
- Frecuencia de palabras positivas y negativas (2 atributos)
- Número de frases, interrogantes, exclamaciones y negaciones (4 atributos)

Bolsas de Palabras

Se crean secuencias de hasta cuatro conjuntos de palabras sobre los que se aplica lematización y eliminación de términos

Resultados

El mejor clasificador consiste en:

- Bolsas de Palabras de tamaño 2
- Usando Lematización
- Usando Stopwords

Conclusiones

Este Trabajo de Fin de Master ha cumplido con el objetivo de ser una iniciación a una nueva vía de investigación, aportando las bases y herramientas. Se ha comprobado que un número elevado de N-Gramas no mejora el clasificador, y que las técnicas clásicas de clasificación de datos, stemmer y stopwords sí lo mejoran. Ha permitido crear una arquitectura que puede ser reutilizada en el futuro, y que incluye procesos comunes al análisis de sentimientos. Por último se ha creado un clasificador prototipo.

