

Práctica 12 de Sistemas de Información para la Web

Ángel García Menéndez

Universidad de Oviedo

Introducción

A continuación se exponen los métodos y sistemas explorados a la hora de obtener un predictor de las fluctuaciones del índice *Down Jones* de la bolsa estadounidense mediante las noticias más relevantes del día.

Aproximación

Existen numerosos ejemplos de la forma de abordar el problema. El ejemplo canónico suele ser el de la clasificación de cuentos clásicos, y pese a las similitudes entre los problemas, existen notables diferencias a considerar.

Por un lado, en el caso de los cuentos lo que se explora es una propiedad inherente a los mismos, como es la temática. Se puede esperar una relación casi directa. Sin embargo, el caso que nos atañe es más complejo, pues la bolsa es un dominio ya de por sí errante y ya de por sí difícilmente predecible. De aquí se puede derivar que la correspondencia entre el contenido de las noticias y las variaciones del índice tendrán una relación menos clara, y por ende será necesario aplicar algo más de creatividad al problema.

Si se intenta elaborar el clasificador siguiendo el sistema de la navaja de Okham, agrupando los textos en *bags of words* y proceder a elaborar el clasificador con el resultado. Empero, el desempeño de esta técnica es bastante mejorable, siendo primero necesaria la eliminación de palabras vacías propias del dominio, y finalmente ninguno de los sistemas de clasificación ofrecidos por la herramienta consigue un resultado aceptable. El método alternativo empleado para el presente caso sigue la siguiente reflexión: la bolsa, como regla general, suele responder de buena forma a noticias positivas, mientras que ante las que se pudieran considerar negativas suele actuar de forma más recelosa.

Dentro del abanico de herramientas de *Orange* existe una que, dado un corpus, es capaz de dar una puntuación a los sentimientos que produce el texto, agrupados en:

- Positivo
- Negativo
- Neutro
- Conjunto

Así, es posible conseguir un valor numérico (el compuesto) que nos de una idea de los sentimientos producidos por las noticias, y de esta forma asignar un valor numérico que emplear a la hora de clasificar. Finalmente, se experimentará con diferentes técnicas y parámetros hasta encontrar una que de una respuesta aceptable.

Resultados