



TESINA DE LICENCIATURA

Título: Clasificación de Subjetividad utilizando Técnicas de Aprendizaje Automático

Autores: Juan Manuel Coria

Director: Dra. Claudia Pons

Codirector: Dr. Waldo Hasperué

Carrera: Licenciatura en Informática

Resumen

La clasificación de subjetividad es un ámbito de la minería de texto poco estudiado en el idioma español, y sin embargo sus aplicaciones son extensas. Su estudio permite comprender mejor la semántica de un texto y la intención de su autor, sin mencionar las implicaciones de su uso en la inteligencia de negocios, para identificar las necesidades de los clientes y obtener métricas valiosas a partir de sus críticas. En este trabajo se intenta aplicar técnicas conocidas de análisis de subjetividad en inglés, adaptadas al español, construyendo en el proceso una base de datos y un sistema clasificador de oraciones.

Palabras Claves

Aprendizaje Automático, Máquinas de Vectores de Soporte, Redes Neuronales, Subjetividad, Objetividad, Minería de Texto.

Conclusiones

Se obtuvieron resultados satisfactorios en ambos modelos, pero la máquina de vectores de soporte alcanzó un mayor rendimiento en general. Sin embargo, el sistema en producción probó ser dependiente de un estimador de subjetividad base, lo que indica que el método planteado puede resultar útil para incrementar la performance de otros clasificadores.

Trabajos Realizados

Se construye una base de datos y un sistema de software capaz de clasificar oraciones como subjetivas u objetivas utilizando un clasificador de máquinas de vectores de soporte y un multiperceptrón. Se compara el rendimiento de los modelos entrenados y se simula la puesta en producción del sistema, utilizando como entrada la misma base de datos.

Trabajos Futuros

Se propone como posibles trabajos futuros desarrollar una base de datos con mayor cantidad de oraciones y con textos más actuales; entrenar, evaluar y comparar otros tipos de clasificadores; explorar otras características extraídas del texto; aplicar algún mecanismo de desambiguación semántica.