

Sentiment Classification using Word Sub-Sequences and Dependency Sub-Trees

Shotaro Matsumoto, Hiroya Takamura, and Manabu Okumura

Tokyo Institute of Technology, Precision and Intelligence Laboratory, 4259
Nagatsuta-cho Midori-ku Yokohama, Japan

`shotaro@lr.pi.titech.ac.jp`, `{takamura, oku}@pi.titech.ac.jp`

Abstract. Document sentiment classification is a task to classify a document according to the positive or negative polarity of its opinion (favorable or unfavorable). We propose using syntactic relations between words in sentences for document sentiment classification. Specifically, we use text mining techniques to extract frequent word sub-sequences and dependency sub-trees from sentences in a document dataset and use them as features of support vector machines. In experiments on movie review datasets, our classifiers obtained the best results yet published using these data.