Abstract

This paper is a case study on the different algorithms and techniques used to build a robust sentiment analysis system. In this case study we take the problem of classifying movie reviews based on their sentiment. For this, machine learning algorithms based on naive Bayes and support vector machine were used. Naive Bayes with minimum cuts algorithm [1] to generate context sensitive summaries before classifying the document was also used. There are mainly four phases in this process. A tokenizer is used to normalize words and to split the text into sentences. A feature extraction model is used to remove unwanted words and to take into account negation. Then the Naive Bayes with minimum cuts algorithm is used to filter the subjective sentences. Finally, the standard SVM/Naive Bayes algorithm was used to arrive at a proper sentiment.

References

1. Bo Pang and Lilian Lee: A Sentimental Education: Sentiment Analysis Using Subjectivity
Summarization Based on Minimum Cuts, Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume.


Index Terms

Computer Science

Algorithms

Keywords

Tokenizer, Feature Extraction, Classifier, Training, Minimum cuts