Authors:

Anuj Sharma

Shubhamoy Dey

Abstract

Sentiment analysis or opinion mining has become an open research domain after proliferation of Internet and Web 2.0 social media. People express their attitudes and opinions on social media including blogs, discussion forums, tweets, etc. and, sentiment analysis concerns about detecting and extracting sentiment or opinion from online text. Sentiment based text classification is different from topical text classification since it involves discrimination based on expressed opinion on a topic. Feature selection is significant for sentiment analysis as the opinionated text may have high dimensions, which can adversely affect the performance of sentiment analysis classifier. This paper explores applicability of feature selection methods for
sentiment analysis and investigates their performance for classification in term of recall, precision and accuracy. Five feature selection methods (Document Frequency, Information Gain, Gain Ratio, Chi Squared, and Relief-F) and three popular sentiment feature lexicons (HM, GI and Opinion Lexicon) are investigated on movie reviews corpus with a size of 2000 documents. The experimental results show that Information Gain gave consistent results and Gain Ratio performs overall best for sentimental feature selection while sentiment lexicons gave poor performance. Furthermore, we found that performance of the classifier depends on appropriate number of representative feature selected from text.

References

of the Association for Computational Linguistics (ACL), July 21–26, 2004 (pp. 271–278). Barcelona, Spain.


Performance Investigation of Feature Selection Methods and Sentiment Lexicons for Sentiment Analysis categorization. ICML, 412–420.

Index Terms
Computer Science
Hpc Applications

Keywords
Sentiment Analysis  Feature Selection  Sentiment Lexicon  Classification