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.

categorization with respect to rating scales. In Proceedings of the 43rd annual meeting of the
Association for Computational Linguistics (ACL), June 25–30, 2005 (pp. 115–124). University
of Michigan, USA.


- Goldberg, A. B., & Zhu, X. (2006). Seeing stars when there aren't many stars:
graph-based semi-supervised learning for sentiment categorization. In Proceedings of
TextGraphs workshop on graph based methods for natural language processing Association for
Computational Linguistics.

conflicting opinions. In: IEEE Symposium on Visual Analytics Science and Technology, pp
59–66.

analysis of on-line text. In Proceedings of the 11th International Conference on Electronic
Publishing. Vienna, Austria.


the hidden dimension for unsupervised text classification. In: Proceedings of the 2009
conference on empirical methods in natural language processing, Association for Computational

- Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel
destinations by supervised machine learning approaches. Expert Systems with Applications,
36(3), 6527–6535.

- Paltoglou, G., & Thelwall, M. (2010). A study of information retrieval weighting
schemes for sentiment analysis. In Proceedings of the 48th Annual Meeting of the Association

- Bai, X. (2011). Predicting consumer sentiments from online text. Decision Support

algorithms for sentiment analysis of restaurant reviews. Expert Systems with Applications

SIGKDD (pp. 168–177).

semantic orientation from association. ACM Transactions on Information Systems, 21(4),
315–346.

through gloss classification. In Proceedings of CIKM-05, the ACM SIGIR conference on
information and knowledge management, Bremen, DE.

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