Abstract

The text data can be of two types: facts and opinions. With the introduction of UTF-8 standards and development of Web 2.0, we are in abundance of opinionated text data available in many languages on the web. Subjectivity analysis aims at dividing those opinionated data into subjective and objective sentences and automatic extraction of subjective information from it. Many subjectivity resources as well as subjectivity analysis works are available in English language. In this paper, we examine different methods of generating subjectivity resources in Hindi language and other Indian languages using resources and tools available in English language. Two methods are proposed using word-level subjectivity annotations. These methods use English language OpinionFinder subjectivity lexicon and a small seed word list of Hindi language which can be expanded to generate subjectivity lexicon, respectively. Four methods are proposed using sentence-level subjectivity annotations. These methods use subjectivity annotated corpora and tools available in English language. Different evaluation strategies are used to validate the generated lexicon and corpora in Hindi language. The simulations
conducted confirm that these methods are effective in rapidly creating subjectivity resources in Hindi language and other Indian languages.

References


stopword removal algorithm. In Microelectronics, Computing and Communications (MicroCom),

17. Vandana Jha, N Manjunath, P Deepa Shenoy, K R Venugopal, and L MPatnaik. Homs:
Hindi opinion mining system. In Recent Trends in Information Systems (ReTIS), 2015 IEEE 2nd

18. Vandana Jha, R Savitha, P Deepa Shenoy, and K R Venugopal. Reputation system:
Evaluating reputation among all good sellers. In Proceedings of NAACL-HLT, pages 115–121,
2016.

Hmdsad: Hindi multidomain sentiment aware dictionary. In 2015 International Conference on

20. Thorsten Joachims. Text categorization with support vector machines: Learning with

21. Aditya Joshi, AR Balamurali, and Pushpak Bhattacharyya. A fall-back strategy for

22. Jaap Kamps, Maarten Marx, Robert J Mokken, and Maarten De Rijke. Using wordnet to
measure semantic orientations of adjectives. In LREC, volume 4, pages 1115–1118. Citeseer,
2004.

23. Arun Karthikeyan Karra, Prabhatkar Pande, Rohan Railkar, Aditya Sharma, and Pushpak

of the 20th international conference on Computational Linguistics, page 1367. Association for

25. Soo-Min Kim and Eduard Hovy. Identifying and analyzing judgment opinions. In
Proceedings of the main conference on Human Language Technology Conference of the North
American Chapter of the Association of Computational Linguistics, pages 200–207. Association

26. Bing Liu. Sentiment analysis and subjectivity. Handbook of natural language processing,

Sentiment analysis of hindi review based on negation and discourse relation. In Sixth
International Joint Conference on Natural Language Processing, page 45, 2013.

An experience in building the indo wordnet-a wordnet for hindi. In First International Conference
on Global WordNet, Mysore, India, 2002.

29. Bo Pang and Lillian Lee. 4.1.2 subjectivity detection and opinion identification. Opinion
mining and sentiment analysis, 2008.

Proceedings of the 12th Conference of the European Chapter of the Association for

31. Ellen Riloff and Janyce Wiebe. Learning extraction patterns for subjective expressions.
In Proceedings of the 2003 conference on Empirical methods in natural language processing,

Index Terms

Computer Science
Circuits and Systems

Keywords

Data Mining, Text Mining, Subjectivity Analysis, Hindi Language, Natural Language Processing.