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


Index Terms
Computer Science
Circuits and Systems

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