

Subjectivity Word Sense Disambiguation (Abstract of Invited Talk)

Janyce Wiebe

Department of Computer Science
University of Pittsburgh
Sennott Square Building, Room 5409
210 S. Bouquet St., Pittsburgh, PA 15260, U.S.A.
wiebe@cs.pitt.edu

Abstract

Many approaches to opinion and sentiment analysis rely on lexicons of words that may be used to express subjectivity. These are compiled as lists of keywords, rather than word meanings (senses). However, many keywords have both subjective and objective senses. False hits – subjectivity clues used with objective senses – are a significant source of error in subjectivity and sentiment analysis. This talk will focus on sense-level opinion and sentiment analysis. First, I will give the results of a study showing that even words judged in previous work to be reliable opinion clues have significant degrees of subjectivity sense ambiguity. Then, we will consider the task of distinguishing between the subjective and objective senses of words in a dictionary, and the related task of creating “usage inventories” of opinion clues. Given such distinctions, the next step is to automatically determine which word instances in a corpus are being used with subjective senses, and which are being used with objective senses (we call this task “SWSD”). We will see evidence that SWSD is more feasible than full word sense disambiguation, because it is more coarse grained – often, the exact sense need not be pinpointed, and that SWSD can be exploited to improve the performance of opinion and sentiment analysis systems via sense-aware classification. Finally, I will discuss experiments in acquiring SWSD data, via token-based context discrimination where the context vector representation is adapted to distinguish between subjective and objective contexts, and the clustering process is enriched by pair-wise constraints, making it semi-supervised.