Creating Subjective and Objective Sentence Classifiers from Unannotated Texts

Janyce Wiebe[1] and Ellen Riloff[2]

[1] Department of Computer Science
University of Pittsburgh
Pittsburgh, PA 15260
`wiebe@cs.pitt.edu`

[2] School of Computing
University of Utah
Salt Lake City, UT 84112
`riloff@cs.utah.edu`

Abstract. This paper presents the results of developing subjectivity classifiers using only unannotated texts for training. The performance rivals that of previous supervised learning approaches. In addition, we advance the state of the art in objective sentence classification by learning extraction patterns associated with objectivity and creating objective classifiers that achieve substantially higher recall than previous work with comparable precision.

1 Introduction

There has been a recent swell of interest in the automatic identification and extraction of attitudes, opinions, and sentiments in text. Motivation for this task comes from the desire to provide tools for information analysts in government, commercial, and political domains, who want to automatically track attitudes and feelings in the news and on-line forums. How do people feel about recent events in the Middle East? Is the rhetoric from a particular opposition group intensifying? What is the range of opinions being expressed in the world press about the best course of action in Iraq? A system that could automatically identify opinions and emotions from text would be an enormous help to someone trying to answer these kinds of questions. Applications that could benefit from this technology include multi-perspective question answering, which aims to present multiple answers to the user based on opinions derived from different sources, and multi-document summarization, which aims to summarize differing opinions and perspectives.

There is also a need to explicitly recognize objective, factual information for applications such as information extraction and question answering. Linguistic processing alone cannot determine the truth or falsity of assertions, but we could direct the system's attention to statements that are objectively presented, to lessen distractions from opinionated, speculative, and evaluative language.