

Tutorial Lecture

Sentiment Analysis

Dr. Pushpak Bhattacharyya

*Professor, Department of Computer Science and Engineering
Indian Institute of Technology, Mumbai*

Abstract

Sentiment analysis is an exciting new field of research in Artificial Intelligence combining Natural Language Processing, Machine Learning and Psychology. Since 2000, due to the proliferation of huge amounts of opinions in electronic form on the web, on social networks and on blogs, automatic means of polarity (positive, negative and neutral) detection in texts flourished in leaps and bounds. Individual and organizations with public interface can no longer afford to be oblivious of sentiments expressed about them in electronic form.

In the present tutorial, we will first discuss the foundations of sentiment analysis, covering knowledge based and machine learning based techniques. Feature engineering forms an important part of the task. Starting from word level features we move to more sophisticated text units and explore their efficacy for SA.

After this, we describe our research on sentiment analysis that attempts to advance the state of the art by tackling new text form (Tweets), a new task (Thwarting), new language (Indian languages) and new features (word senses instead of words). Tweets are noisy texts; they need text cleaning and normalization. Exploitation of discourse relations like 'although', 'still', 'but' and so on improve the accuracy considerably. Languages differ in terms of annotated resources. However, projecting the parameters learnt from one language to another can ameliorate the problem of resource scarcity. We show the use of these ideas for sentiment analysis of Indian languages. We demonstrate that working with senses instead of just words is not only advantageous from the point of view of multilinguality, but also for accuracy. Finally, problems like thwarting (reversal of polarity just by a single statement at a critical place in the text) and sarcasm are hard problems. We present some of our findings in these tasks too.