Opinion Mining, Subjectivity and Factuality

Rodolfo Delmonte
University Ca'Foscari
Italy
delmont@unive.it

Abstract

We assume that in order to properly capture opinion and sentiment expressed in a text or dialog any system needs a deep text processing approach. In particular, the idea that the task may be solved by the use of Information Retrieval tools like Bag of Words Approaches (BOWs) is totally flawed. BOWs approaches are sometimes also camouflaged by a keyword based Ontology matching and Concept search, based on such lexica as SentiWordNet, by simply stemming a text and using content words to match its entries and produce some result. Any search based on keywords and BOWs is fatally flawed by the impossibility to cope with such fundamental issues as the following ones:

- presence of **negation** at different levels of syntactic constituency;
- presence of **lexicalized negation** in the verb or in adverbs;
- presence of conditional, counterfactual subordinators;
- double negations with copulative verbs;
- presence of modals and other modality operators.

In order to cope with these linguistic elements we propose to build a Flat Logical Form (FLF) directly from a Dependency Structure representation augmented by indices and where anaphora resolution has operated pronoun-antecedent substitutions. We implemented these additions our the system called **venses** that we will show. The output of the system is an xml representation where each sentence of a text or dialog is a list of attribute-value pairs, like **polarity**, **attitute** and **factuality**. In order to produce this output, the system makes use of FLF and a vector of semantic attributes associated to the verb at propositional level and then memorized. Important notions required by the computation of opinion and sentiment are also the distinction of the semantic content of each proposition into two separate categories:

- Objective vs Subjective

This distinction is obtained by searching for **factivity** markers again at propositional level. In particular we take into account:

- tense;
- voice;
- mood;
- modality operators;
- modifiers and attributes adjuncts at sentence level;
- lexical type of the verb (in Levin's classes and also using WordNet classification).