A quantitative assessment of the impact of Valence Shifters and Emoji in lexicon for Italian Sentiment Analysis

Elena Catanese, Luca Valentino, Giorgia Sacco

Istituto Nazionale di Statistica – catanese@istat.it; luvalent@istat.it; giorgia.sacco@istat.it

Abstract

In recent years, due to the enormous spread of Social Media, Sentiment Analysis has become a popular task and still is a growing field. Sentiment Analysis attempts to automatically determine the sentiment, or positive/negative opinion, contained in text. Although Sentiment Analysis can be performed in a supervised setting, the use of lexicons (list of words with a pre-computed score) keeps remaining the most popular. One of the biggest advantages, at least, for Official Institutions, is that lexicons ensure the transparency of the adopted methodologies and users' understandability. Some special words, such as negations and amplifiers, can alter the meaning (and the sentiment) of a sentence. These words are known as Valence Shifters. Another element that can have an impact is the increasingly frequent use of emoji. This study proposes an enrichment of the current lexicon based approach used to produce the Italian Social Mood on Economy Index (SMEI), integrating it with the treatment of Emoji and Valence Shifters. On the one hand we use annotated Italian tweets datsets to assess the accuracy of the methodologies. On the other hand, we analyse the inpact of these in the daily time series of the SMEI; the accuracy is assessed by analyzing the relevance of the valleys and peaks of the index and by analyzing the trend of the index itself.