Abstract:
In this work, we created subjectivity lexicons of positive and negative expressions for Indonesian language by automatically translating English lexicons, and by intersecting and unioning the translation results. We compared the performances of the resulting lexicons using a simple prediction method that compares the number of occurrences of positive and negative expressions in a sentence. We also experimented with weighting the expressions by their frequency and relative frequency in unannotated data. A modification in prediction method using machine learning was later used to better incorporate the information that cannot be captured by the simple prediction. We showed that the lexicons were able to reach high recall but low precision when predicting whether a sentence is evaluative (positive or negative) or not (neutral). Scoring the expressions improve the recall or precision but with comparable decrease in the other measure. The machine learning prediction was able to minimize the sensitivity of the performances to the size of the lexicon, but further experiments are required to explore the best choice for the prediction method.