

Abstract

Product reviews in e-commerce can help manufacturers improve products quality and help consumers to make decisions. Unfortunately, the accumulation amount of reviews complicate to conclude the reviews. Summarizing product reviews with sentiment analysis is a solution of the mass numbers by product reviews. The first stage is to determine the extracted product features using the pattern knowledge which is then compared to the features in document to find the identified product features. Sentiment analysis is done by using a SentiWordNet dictionary to determine the orientation of opinion in a sentence. Summarizing product reviews conducted by using tf-idf scoring for each sentences, then do centroid based clustering for each feature of orientation. Results of preprocessing's evaluation process that obtained from this research is that preprocessing using lemmatization and labeling with POS Tagging is the most suitable and affect the performance compared than add stopword removal. The performance of the extraction stage around the numbers 60% to 90% in the dataset used. The results of classification using SentiWordNet are between 65% and 72% depending on the structure and complexity of its dataset. Documents summarization can be done with a dataset that has a number sentence reviews and product features are adequate so that it can be evaluated as well as clustering.

Keywords: product reviews, pattern knowledge, sentiment analysis, summarizing, tf-idf, centroid-based clustering.