Abstract

Sentiment Analysis of large-scale data has become increasingly important and has attracted many researchers, urging them to use new platforms and tools that can handle large volumes of data. In this paper, we present new evaluation experiments of sentiment analysis for a large-scale dataset of online customer's reviews under Apache Spark data Processing System. Apache Spark's scalable machine learning library (MLlib) is used and three classification techniques from the library are applied; Naïve Bayes, Support vector machine, and logistic regression. The results are evaluated using the accuracy metric. Experimental results show that Support vector machine classifier outperforms Naïve Bayes and logistic regression classifiers.