Abstract

Instagram is a popular social media platform where users can share photos and videos, and also post comments on other users postings. Although there are many benefits on sharing information and posting comments, the freedom of using a social media platform also has a negative effect. One of non-constructive actions performed by social media users is cyberbullying, a misuse of technology through social media to embarrass or threaten others. Cyberbullying could affect the social media users, especially the target/victim, hence we have to build a system that can limit the negative posts. In this research, we tackle the cyberbullying detection on Instagram comments as a classification problem and employ the SVM classifier. As a supervised machine learning approach, the SVM method has limitation on processing new words that never found in the training data. Therefore, we employed a semantic information derived from pre-trained word embeddings to gather similar words that appear in the training data to substitute the unknown words in the testing data. The experimental results show that the use of semantic similarity information improve the classification accuracy by 7%, from 67% to 74%.

Keywords: cyberbullying, support vector machine, semantic similarity.