

Mechanism for Sarcasm Detection and Classification in Malay Social Media

Abstract

The classification of users' sentiment from social media data can be used to learn public opinion on certain issues. The presence of sarcasm in sentences can hamper the performance of the classification as it tends to "fool" the system. In this paper, we investigate mechanisms for detecting sarcasm in Malay social media data that contain sarcastic contents; more specifically the public comments on economic related posts on Facebook. Two features were investigated; the n-gram and punctuation marks. Features selection in the form of Pearson's correlation was then applied to reduce the features size. To measure the performances of the selected features, two supervised classification techniques were employed which are k-Nearest Neighbors and non-linear Support Vector Machine. Experiments on sarcasm detection and classification were conducted. Results show that combination of n-gram and punctuation marks produced the best F -measure and Area Under Curve of 0.818 for sarcasm detection. Extended experiment on sarcasm classification recorded F -measure of 0.991 with Area Under Curve of 0.994 for sarcasm positivity while F -measure of 0.902 with Area Under Curve of 0.846 for sarcasm negativity.