

Performance Evaluation of Hybrid Feature Selection Technique for Sentiment Classification based on Food Reviews

Suryanti Awang^a, Nur Syafiqah Mohd Nafis^b

^a Faculty of Computing Centre For Data Science & Artificial Intelligence
(Data Science Centre) Soft Computing & Intelligent Systems (SPINT),
Universiti Malaysia Pahang, Kuantan, Malaysia

^b Faculty of Computing, Universiti Malaysia Pahang, Kuantan, Malaysia

ABSTRACT

This paper presents an evaluation of the performance efficiency of sentiment classification using a hybrid feature selection technique. This technique is able to overcome the issue of lack in evaluating features importance by using a combination of TF-IDF+SVM-RFE (Term Frequency-Inverse Document Frequency (TF-IDF) and Supports Vector Machine (SVM-RFE)). Feature importance is measured and significant features are selected recursively based on the number of significant features known as k-top features. We tested this technique with a food reviews dataset from Kaggle to classify a positive and negative review. Finally, SVM has been deployed as a classifier to evaluate the classification performance. The performance is observed based on the accuracy, precision, recall and F-measure. The highest accuracy is 80%, precision is 82%, recall is 76% and F-measure is 79%. Consequently, 24.5% of the features to be classified in this technique have been reduced in obtaining these highest results. Thus, the computational resources are able to be utilized optimally from this reduction and the classification performance efficiency is able to be maintained.

KEYWORDS

Sentiment Classification; Sentiment Analysis; NLP; Computational Intelligence

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Ministry of Higher Education of Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2019/ICT02/UMP/02/1) (Universiti Malaysia Pahang internal grant: RDU1901103) for supporting this research.