7th International Conference on Computer Science and Information Technology (CoSIT 2020)
January 25 ~ 26, 2020, Zurich, Switzerland
https://cosit2020.org/

Sentiment Classification for Under-Resourced Language Using Word2Vec Neural Network: Amharic Language Social Media Text

[1]Zewdie Mossie and [2]Jenq-Haur Wang, [1]Department of Computer Science and Information Engineering, National Taipei University of Technology, Taiwan and [2]Department of International Graduate Program in Electrical Engineering and Computer Science, National Taipei University of Technology, Taipei, Taiwan

ABSTRACT

Sentiment classification becomes popular task in social network texts which express opinions on different issue to analyze and produce useful knowledge. However, many linguistic computational resources are available only for English language. In the recent years, due to the emergence of social media platforms, opinion-rich resources are booming abundant for under-resourced languages with the need to perform Sentiment Analysis. On the other hand,most of the existing researches focus on how to extract the effective features, such as lexical and syntactic features,while limited work has been done on semantic features, which can make more contributions to both under-resourced and resourceful languages. In this paper, we proposed sentiment classification based on Word2Vec for Amharic Language text on political domain. The Word2Vec establishes the neural network models to learn the vector representations of words to extract the deep semantic relationships. Firstly, we cluster the similar features together and apply language modeling Ngram to check sentiment-bearing Co-occurring Terms (COT). Word2Vec and TF-IDF were used to learn the word representations as a candidate feature vector. Secondly, The Gradient-Boosting Tree(GBT) and Random forest machine learning classifiers were used to train and test in the Apache Spark platform. In our experiments, we use the Amharic language in Ethiopia and adopt a standard natural language pre-processing techniques on the crawled Facebook datasets to categorize into positive and negative opinions. Experimental results of feature extraction using Word2Vec technique performs better in the GBT classifier achieving an average accuracy of 82.29%. Therefore, our proposed approach can successfully discriminate among posts and comments expressing positive and negative opinions.

KEYWORDS

Amharic text Sentiment, Word2Vec Semantic, Social Media, Under-resourced Language

Read More Accepted Papers List :https://cosit2020.org/papers.html

Participants are Invited for the conference!!!.

Contact :cosit@cosit2020.org or Cosit23_conf@yahoo.com