Sentiment Analysis in Sinhala Texts using Convolution Neural Networks

H.M.S.S. Herath* and M. Siyamalan

Department of Computer Science, University of Jaffna
**Corresponding Author E-mail: sewmiherath@gmail.com, TP: +9477465934*

Because of the rapid development of the information communication technology, an enormous amount of data is produced, shared across the internet and other media. Opinion mining, also known as Sentiment Analysis, is a technique, which can be used to detect the opinion of a given sentence or to make a judgement based the given sentence, can play a major role in automatically analysing this data. In addition, the development of Natural Language Processing in Sri Lanka leads the Sri Lankan native user to browse web in their native language and to express their opinions in their mother tongue. But in most of the cases Sinhala language was named as the morphological rich,less resourced language. An automatic solution for the text categorisation and opinion mining could be very useful for analysing sentences from Sinhala language. This work explores a Convolution Neural Network (CNN) based sentiment analysis techniqe, where, each word of a sentence is converted into a numerical representation using a pretrained FastText word embedding model. These numerical representations obtained for each word of the sentences are then used to train the CNN, in order to predict the opinion of the given sentences at test time. This CNN model is trained and tested on a Sinhala news comments dataset which consists of 5010 comments. There were 2520 negative comments and 2490 positive comments. Dataset used in this project is collected by crawling Sinhala online news sites, mainly www.lankadeepa.lk.This data was initially preprocessed by removing non-Sinhala characters, punctuation marks and stop-words. Our model was trained and tested on 70% and 30% of the data respectively. Experiments report a testing accuracy of 85%.

Keywords: Convolution Neural Network; Natural Language Processing; Sentiment Analysis; Text analysis for Sinhala Language