Sentiment Analysis Using LSTM, CNN & GloVe for Exploration of social media

A.C. Nanayakkara*, B.T.G.S. Kumara, and R.M.K.T. Rathnayaka
Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka
*chethana@ccs.sab.ac.lk

Sentiment analysis is an automated method that makes use of computing energy to identify the polarity of a given context. In particular, because of the improvement of social media, there shall be a substantial demand for digging up significant statistics from the big data on the Internet via sentiment analysis. Inspired by the contribution of using the achievements of deep learning, the study was enthused by coping with the sentiment analysis task by using the deep learning models. In this paper, we propose a framework called Glove + deep learning. Initially, we utilized the Glove embedding to compute vector representations of words, which will be the input for the Neural net. Precisely, the GloVe word embedding model is followed right here with the intention of creating the best use of the worldwide and nearby records for the training corpus. Then we designed a simple neural network, Convolutional Neural network (CNN) and Recurrent based Long Short-Term Memory (LSTM) neural network architectures for the sentiment classification of well-known YouTube video review. In order to overcome the shortcomings of the traditional recurrent neural community which is the simplest capability to managing short-time period dependencies and the inability to examine the lengthy-time period records of textual content, a neural network model combining LSTM is proposed in the study. We used the Rectified Linear Unit (ReLU) Activation function, Adam Optimizer, and binary cross entropy as a loss function to improve the accuracy and generalizability of the model. The model appreciably generated higher effects and specifically, the LSTM model was appreciably sensitive to overfitting even without parameter tuning. Through the comparison test of the simple deep neural network model, CNN model, and LSTM model, evaluating the training accuracy (81%, 86%, 87%), testing accuracy (64%, 74%, 84%), and overfitting indicators (15, 12, 3), it was identified that the LSTM model with Glove word embedding demonstrated optimally. In this effort, the number of layers, neurons, hyperparameters, and activation functions were randomly selected. Future studies can change these parameters and test which works optimally for different embedding methods and within diversified datasets.

Keywords: CNN, Deep Learning, Glove Word Embedding, LSTM, Sentiment Analysis