

A LSTM model for improving movie review

Monir Hossain¹, Rafi Md. hasib², T. M. Sazzad Hossain³, and Md Habibur Rahman⁴

¹ID : 2017-1-60-18, East West University

²ID : 2017-1-60-138, East West University

³ID : 2015-1-60-152, East West University

⁴ID : 2015-1-60-142, East West University

May 26, 2021

1 Introduction

Natural language processing(NLP) is one of the interesting sectors in machine learning research. NLP is highly applicable in two types of problems: the first type is various computation tasks like spell checking, grammar checking, language translation etc . It is also highly used in reorganization of psychological and theoretical knowledge. It determines people's liking of a topic, event, issues etc . Everyday social networks and other sites generate more and more information about their views on political or any social topic. NLP can play a crucial role in Understanding people's views on a topic.

We proposed a LSTM model to find a better solution to this problem. We utilize LSTM(Long term memory concept) of Recurrent Neural Network to understand long term dependencies of word sequences. For word embedding we use a pre trained Word2vector model GloVe. For experiment we use [movie review data-set](#) from stanford University Our main contribution in the study as follows

1. Create word embedding from pre trained word2vector model glove 6b 200d
2. To understand word sequence and their relation we use a sequential model LSTM to understand deeper semantic of words.

Some of the work related traditionally scientists proposed a support vector machine, random forest, Naive Bayes to solve sentiment analysis problems which achieve better solutions. Later researchers applied a variant of Naive Bayes which resulted in a better solution than previous. Later researchers proposed a paragraph2vec approach for feature extraction which outperformed most of the TF, TF-IDF, SVM, NB models based on accuracy. Letter a new proposal was to solve Image based sentiment analysis problem using CNN framework approach. Which was experimented on a dataset with 1269 images which are collected from twitter. This system places high performance in terms of accuracy and precision.

2 Methodology

In this section we describe detail of out proposed model. Initially we feed our LSTM model a word embedding to understand long term dependency between the sequence of word and and in the end of LSTM we apply a classifier layer to output our desire result.

Our LSTM layer is combined of Embedding layer ,LSTM layer , Dropout technique, Regularization, Danselayer and for activation RELU and sigmoid activation function.

1. Embedding Layer : We use word2vec model for word embedding. we use pre trained glove word2vec model for word embedding
2. LSTM layer : We use a LSTM layer after embedding layer
3. Dropout layer : We use a dropout layer with keep_prob=0.2

4. Dense layer : After dropout layer we use a dense layer with 256 neuron , l2 regularization and RELU activation.
5. Dense layer : Again a Dropout layer with 128 neuron , l2 regularization and RELU activation
6. Dropout layer : After dense layer a dropout layer with keep _prob = 0.2
7. Finally a sigmoid activation layer for classifing the output.

3 Experiment

for experiment we use IMDB movie review dataset form [stanford](#). this dataset contain 50 thousand movie review with positive and negative review. we divide the dataset into 80:20 for train and test dataset. For model architecture we use keras library from tensorflow and glove model for word embedding. we also use natural language processing library nltk for processing data. after 30 epoch we get a 90 percent accuracy on test set.

4 Conclusion

Our model is successfully able to classify user review with 90 present accuracy. However it still have a high variance problem. so with hydrometer tuning it is possible to get an accuracy more than that. In future we will try to reduce high variance problem and we also try to combine CNN with LSTM to create a hybrid model which might result a good solution.