
Sentiment Analysis From Text

by Team 8

DESCRIPTION OF DATASET

The ACL Internet Movie Database (IMDb) dataset used was created in for learning word vectors. The dataset consists of (50,000) reviews are paired with a label of 0 or 1 to represent negative and positive sentiment, respectively. These labels were linearly mapped from the IMDb's star rating system where reviewers can rate a movie a certain number of stars from 1 to 10. The reviews with labels are split in half; each set has 12,500 positive reviews and 12,500 negative reviews to keep the data balanced

The Proposed Method

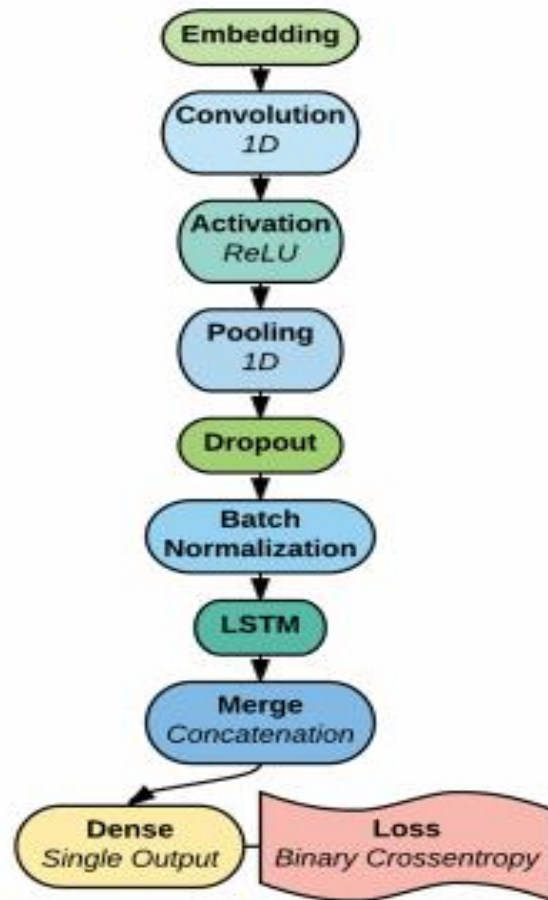


Figure 3. Diagram of basic structure of CNN based LSTM network. We use this in our experiments as a baseline model to compare with our novel model.

EXPERIMENTAL SETUP

The IMDb review sentiment dataset was used for all experiments. The dataset was preprocessed to a dictionary size of 5,000 words with a zero-padded maximum sequence of 500 words per a review; anymore data became insignificant to the networks objective.

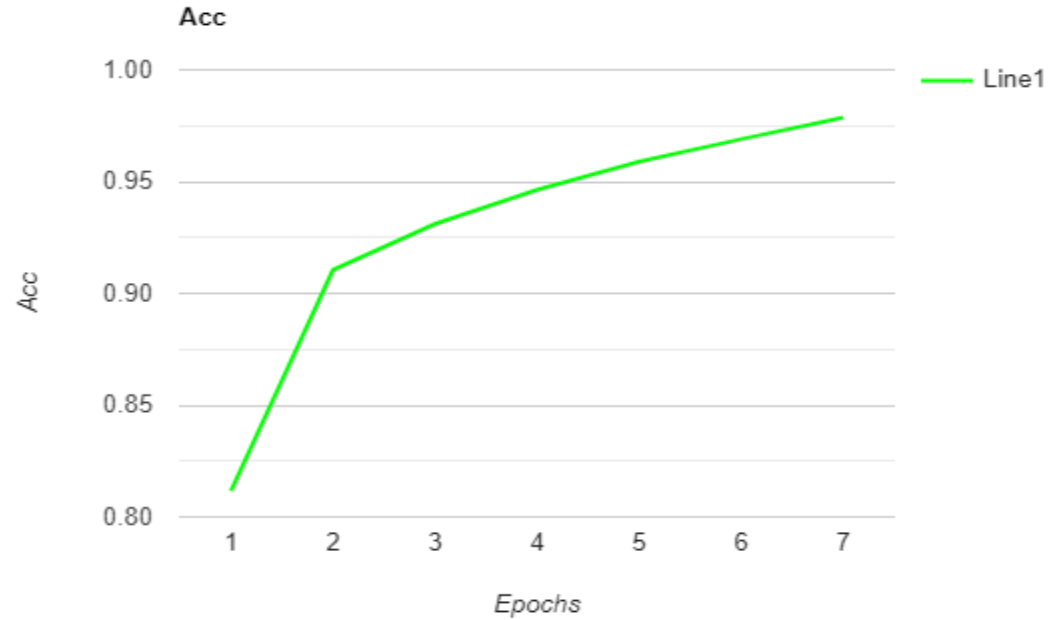
Max Pooling & Dropout

Although max pooling assisted with the chief issue of overfitting, large pooling sizes proved to decrease accuracy. The optimum kernel size during the experiments was 2; reducing the height of the input by half.

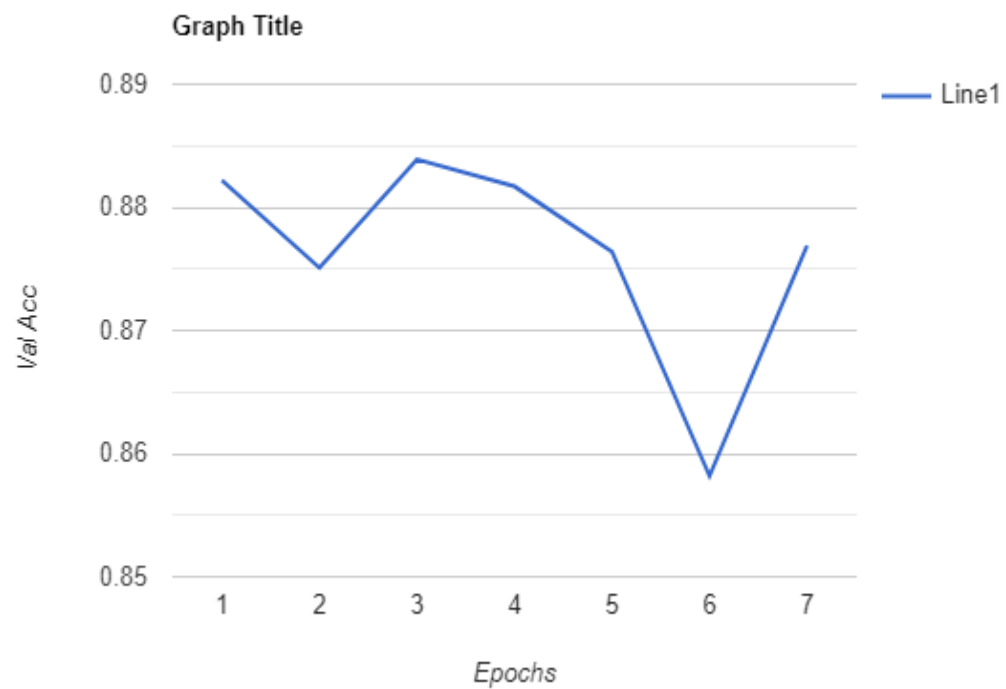
The dropout layer was recognized to be the best option to reduce overfitting.

RESULTS AND ANALYSIS

Accuracy :



Overfitting :



Result

... Please Enter your valuable review:



1 s



... Please Enter your valuable review:

The movie is the best movie ever

[] 1 s



Please Enter your valuable review:

The movie is the best movie ever

Positive review

[] 1 s