COMP 4107

PROJECT - SENTIMENT ANALYSIS

A Neural Network Performing Sentiment Analysis

JONATHAN AUSTIN | HRISHI MUKHERJEE

Introduction

Sentiment Analysis has been around for just over a decade [2] and has proven to be useful in many real world applications like recommender systems and stock market prediction. Many different instruments have been used to accomplish this task, and one very recent one has been the use of neural networks. Our project aims to do exactly that by using a convolutional neural network and a corpus of movie reviews.

PROBLEM

This project aims to solve the problem of analyzing the sentiment of a body of text. A dataset of movie reviews extracted from Rotten Tomatoes is the corpus of text which needs to be analyzed [3]. This dataset contains movie reviews with corresponding sentimental values for each review. The task is to classify the movie reviews into one of five categories - very negative, negative, neutral, positive, and very positive. [4]

NEURAL NETWORK ARCHITECTURE

A Convolutional Neural Network has been chosen to solve this problem. The base neural network has six layers as mentioned below. A third convolutional layer is added during the testing process.

1. Input Layer
2. Embedded Layer
3. Convolutional Layer 1
4. Convolutional Layer 2
5. Dropout Layer
6. Output Layer

SOLUTION

The movie reviews are classified into one of five categories - very negative, negative, neutral, positive, and very positive. Once the network is trained, it is tested against another set of movie reviews in order to calculate how accurately it can predict the sentiment of a given movie review. The neural network is trained in a supervised manner, with corresponding labels provided as expected outputs to the network.

The problem is solved using the following process:

1. The dataset of movie reviews are converted into vectors using a Tokenizer
2. The labels are then converted to a binary representation
3. The data is split into two different sets - training and testing
4. The Embedding Layer is initialized
5. The binary matrix representation of the data is then transformed into a Tensor
6. One by one, the tensors are loaded into the embedded layer and run through the convolutional layers
7. After each run, the network outputs a classification of the tensor
8. The actual output is compared with the expected output to compute the accuracy of the tensor

EXPERIMENTAL METHODOLOGY

In order to test the neural network, a number of variables were modified to generate the accuracies reported. The modified variables directly affect the final classification accuracies. These variables are as follows:

1. Number of Layers
2. Training Batch Size
3. Filter Size

RESULTS

The following results have been documented after training a convolutional neural network on a set of 50,000 comments. The training and testing process has been repeated over 15 epochs. The accuracies for each epoch are nicely formatted in the figures corresponding to each test case. These figures can be found in the figures section from pages 5-10.

Number of Layers	Batch Size	Filter Size	Average Accuracy	Maximum Accuracy	Figure
2	4	5	0.52	0.57	1
2	8	5	0.52	0.62	2
2	16	5	0.54	0.61	3
2	32	5	0.53	0.60	4
3	4	5	0.54	**0.63**	5
3	8	5	0.53	0.57	6
3	16	5	0.53	0.59	7
3	32	5	0.54	0.57	8
3	4	4	0.53	0.59	9
3	8	4	**0.55**	0.61	10
3	16	4	0.54	0.61	11
3	32	4	0.53	0.57	12

CITATIONS

1. (2017, December 7). Retrieved from http://curatti.com/wp-content/uploads/2015/01/Slider-Sentiment-468x340.png
2. Natarajan, V. (2013, November 8). What is the history of Sentiment Analysis? Retrieved December 7, 2017, from https://www.quora.com/What-is-the-history-of-Sentiment-Analysis?share=1
3. (n.d.). Retrieved December 08, 2017, from https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews
4. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. (n.d.). Retrieved December 08, 2017, from https://nlp.stanford.edu/sentiment/
5. K. (2017, December 12). Keras-team/keras. Retrieved December 13, 2017, from https://github.com/keras-team/keras

FIGURES

FIGURE 1

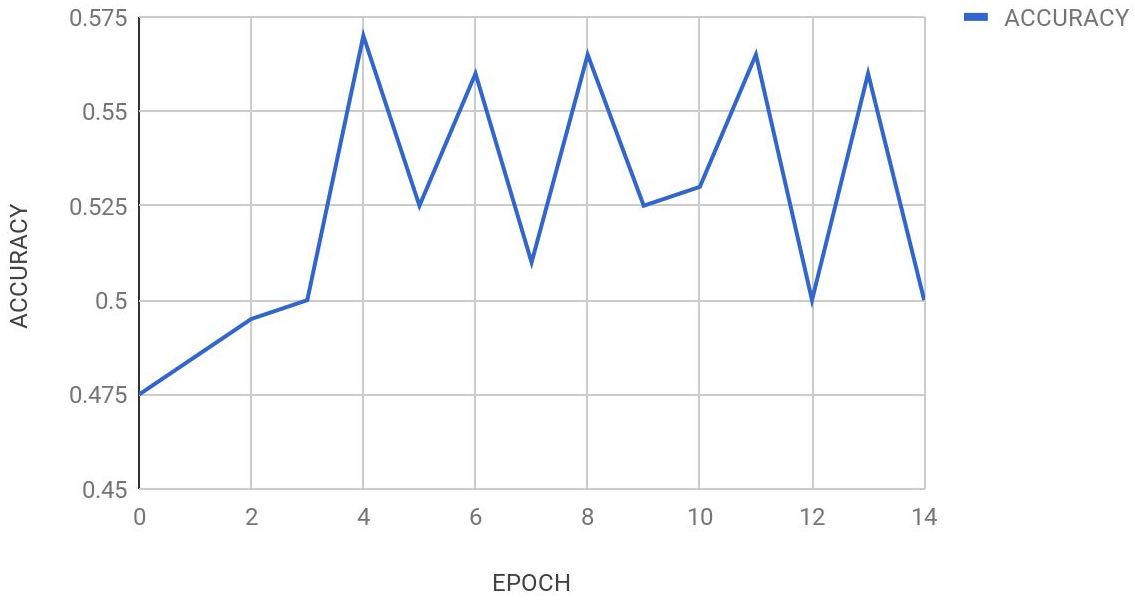

FIGURE 2

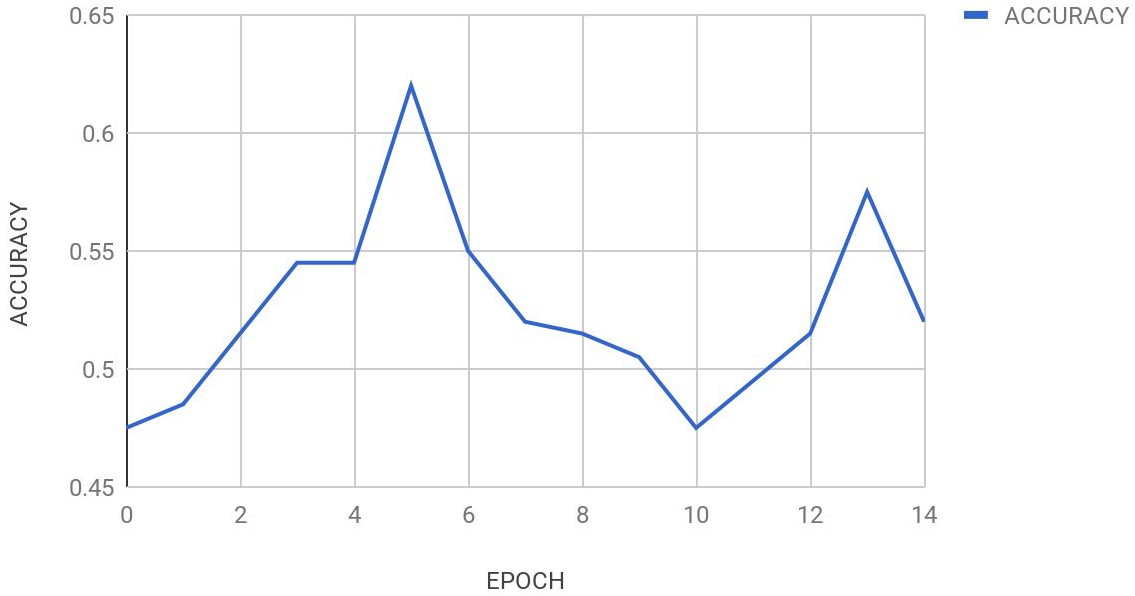

FIGURE 3

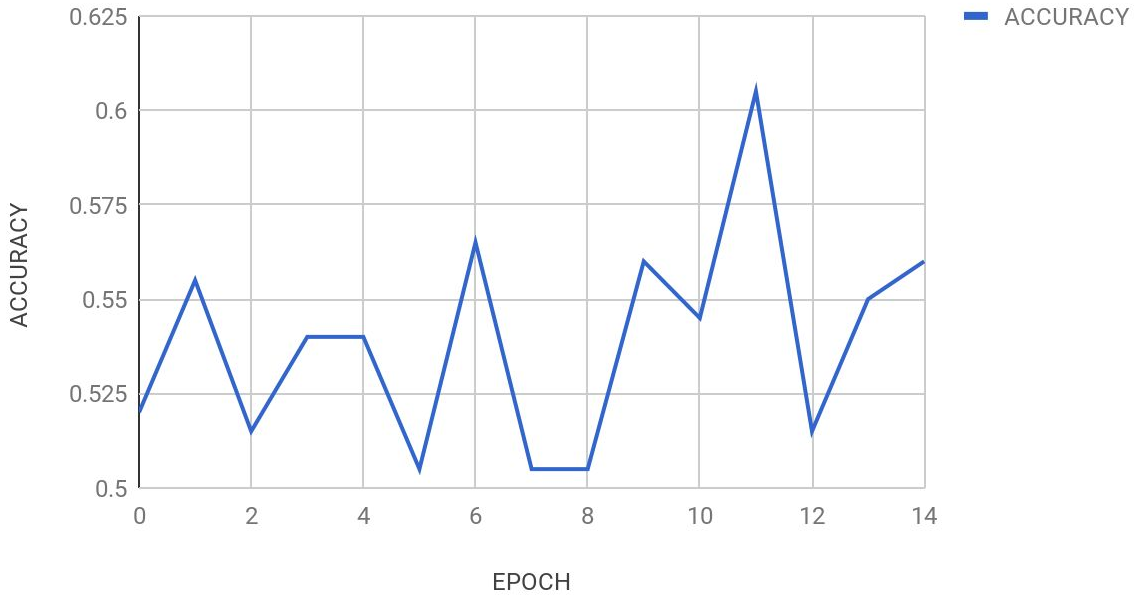

FIGURE 4

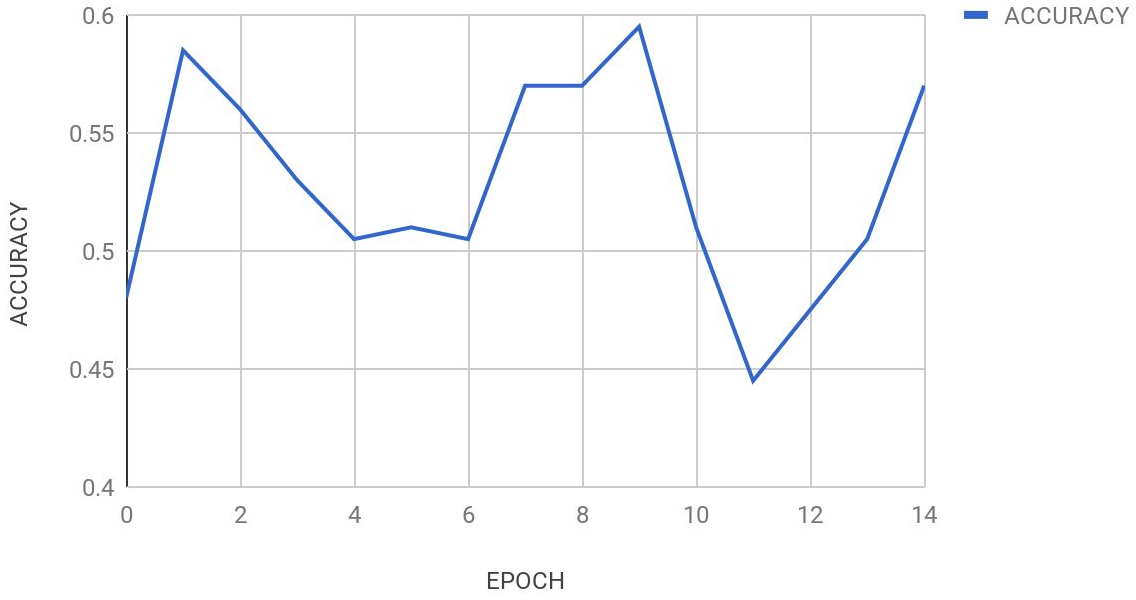

FIGURE 5

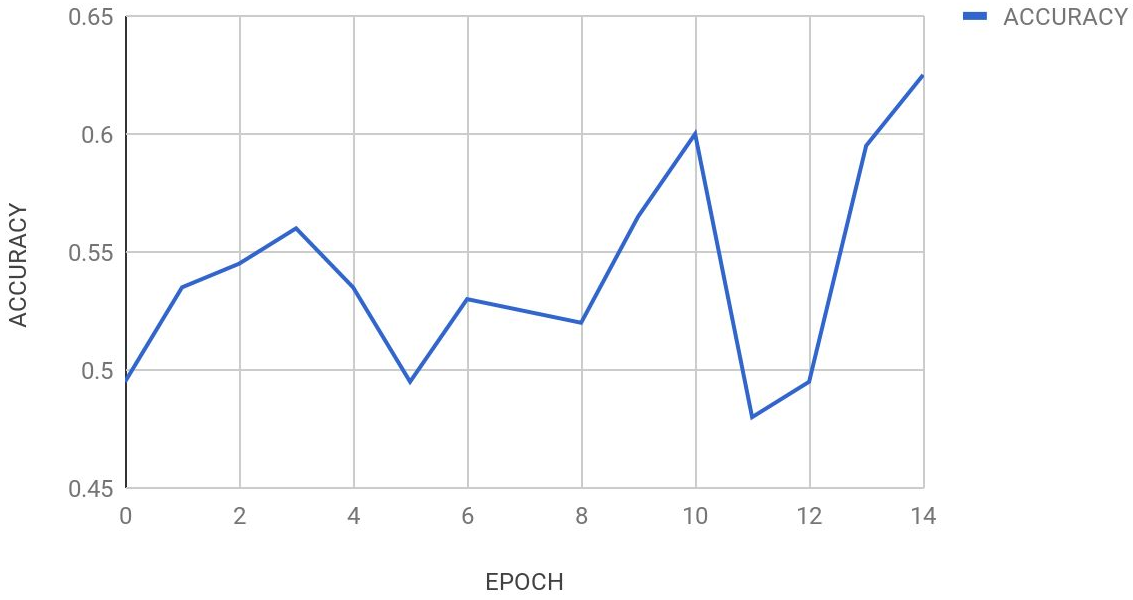

FIGURE 6

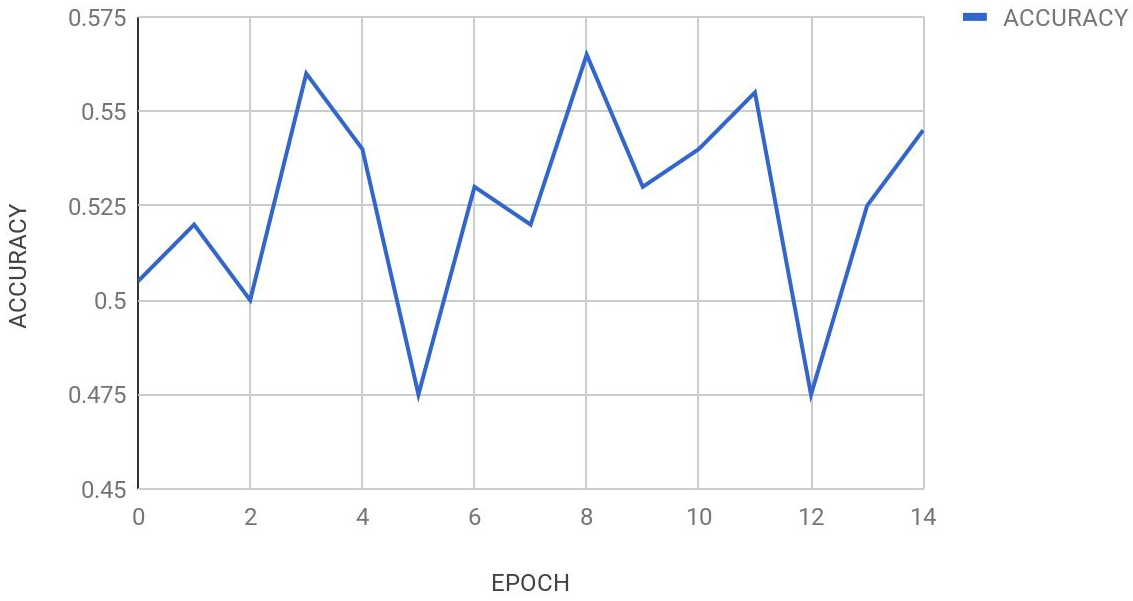

FIGURE 7

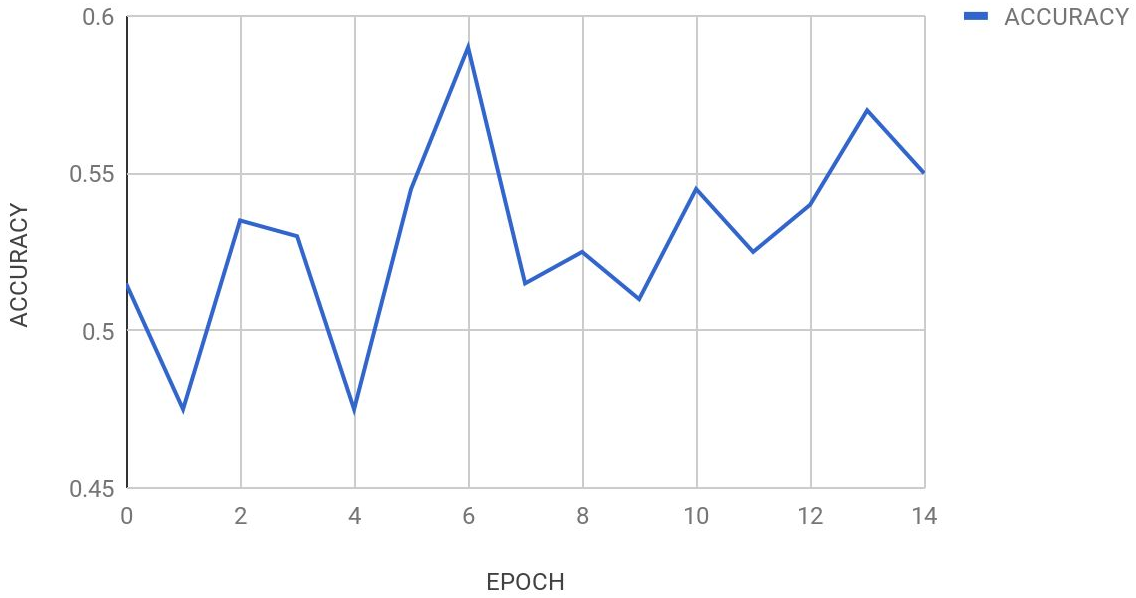

FIGURE 8

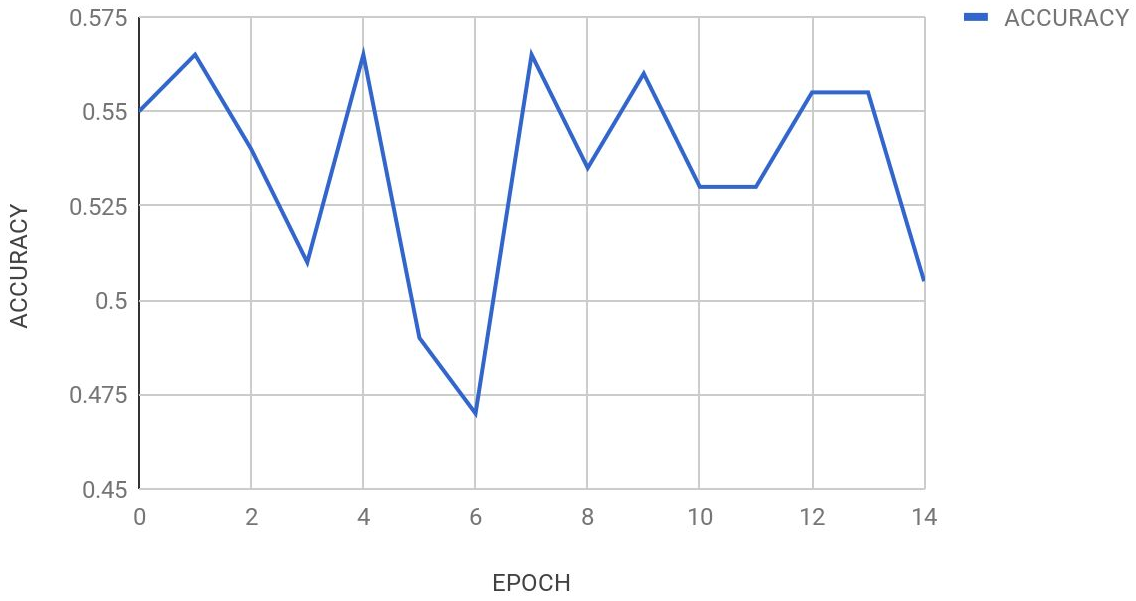

FIGURE 9

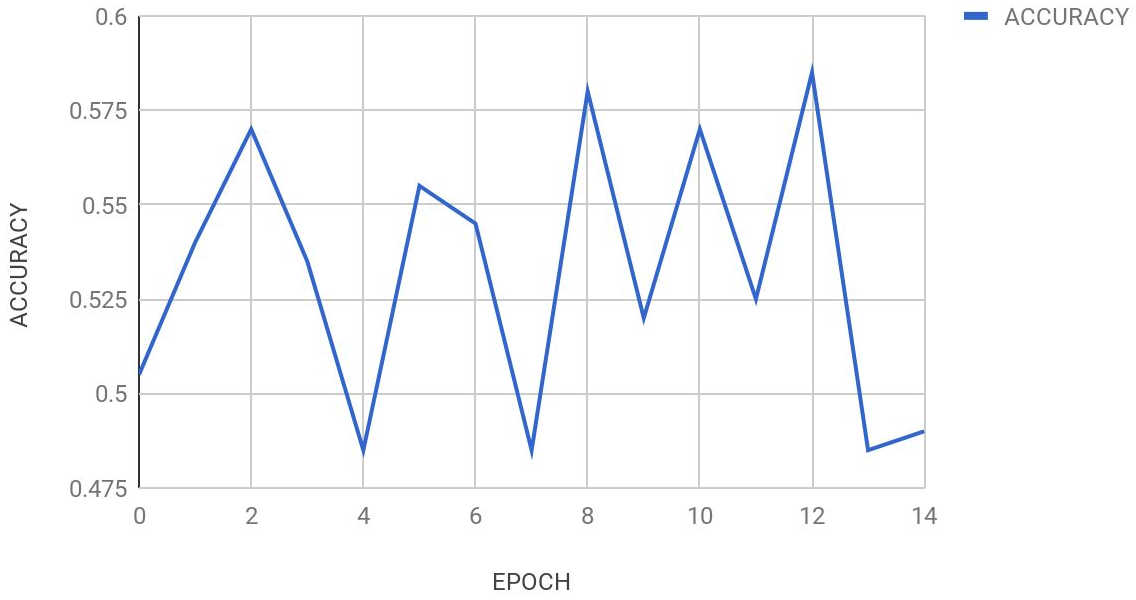

FIGURE 10

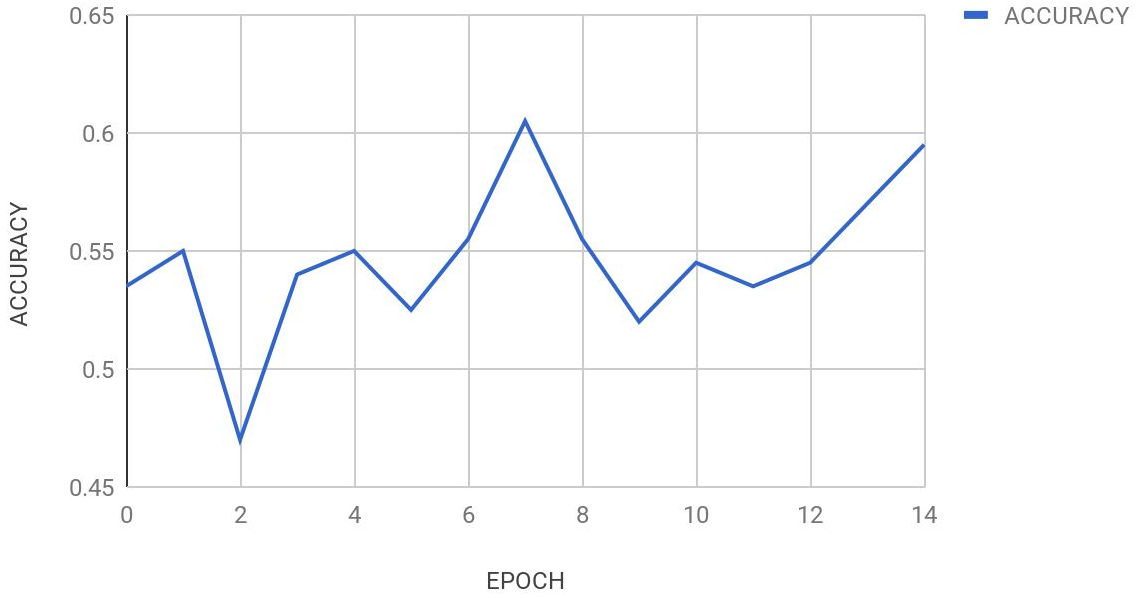

FIGURE 11

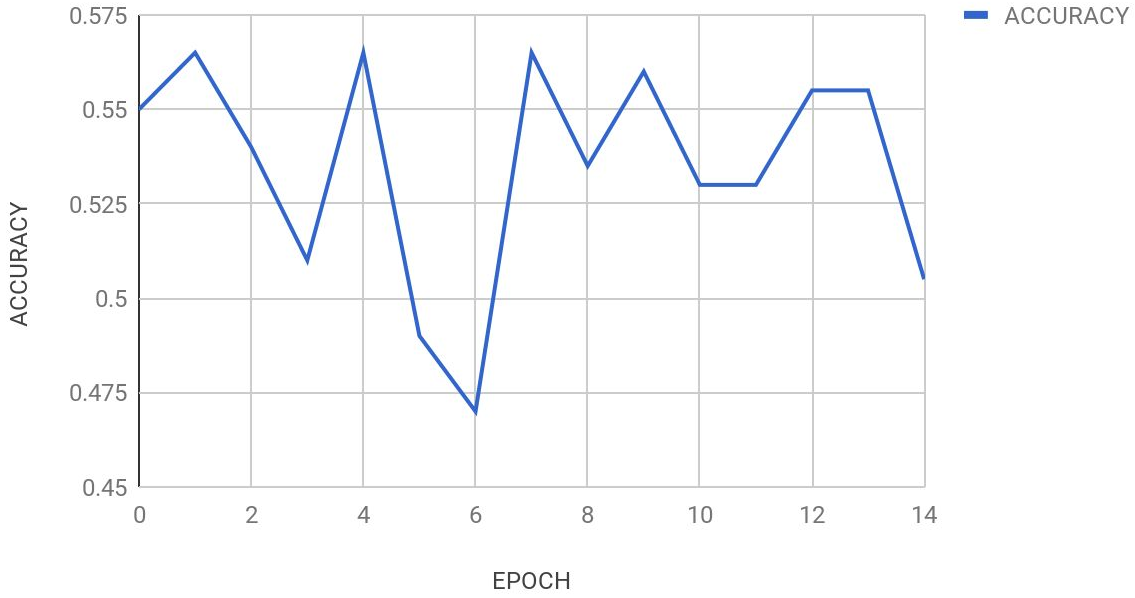

FIGURE 12