

DEEP LEARNING LAB ASSIGNMENT – 2

Introduction:

The given task is to implement Text Classification using CNN and display the graph on Tensorboard. CNNs are layers of convolutions with nonlinear functions applied to the results. Convolutions are used above the input layer to process the output in CNN. We apply our CNN model on consumer complaints dataset.

Objectives:

The main objective is to apply the CNN on the dataset for text classification. CNN architecture works well in obtaining good results in performance for the dataset. Our main objective for this task is to classify the given dataset into multiple classes.

Approaches:

Firstly, we will develop multiple layers for the model. Initial layer performs the word embedding. The embedded words are convoluted using multiple filter sizes. Then this layer will be implemented by max pool which creates a feature vector. Using matrix multiplication and max pooling, the predictions are generated.

Workflow:

1. Dataset is read using read csv.
2. We do the embedding to the file and then we do convolution on the data in the TextCNN.
3. We train our data then and evaluate the trained model using the test data
4. The predictCNN here is the resulted output in which the predicted output will be present.
5. The datahelpers will help making the data nonlinear initially
6. Using the Gradient Descent Optimizer to optimize the cost obtained.
7. We create a batch and train our model based on that batch for iterations of the size of epochs.
8. Using the tf.summary.filewriter, we create the graphs on the tensorboard and view our results on the tensorboard.

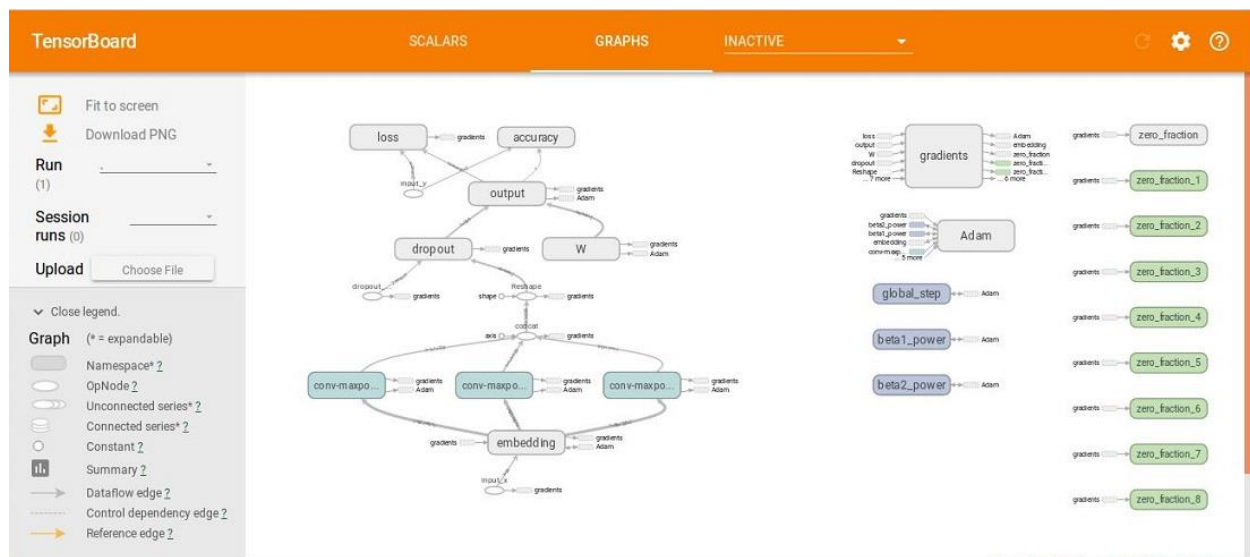
Dataset:

Dataset we used in this assignment is a financial related dataset. It has the complaints from the consumers regarding the bank accounts and various other factors. It is a very large dataset and has 11 classes. Using the convolutional neural networks and tensorflow, we built the model.

Parameters:

The main parameters considered while building model are vocabulary size, embedding size, filter sizes, num_filters, r2lambda value, num_classes. The 11 classes in the dataset are: Bank Account, Consumer Loan, Credit Card, Credit Reporting, Debt Collection, Money Transfers, Mortgage, Payday Loan, Prepaid Loan, Student Loan, Other Financial Services.

Evaluation & Discussion:



After increasing the embedding size and num_layers, there is an increase in the accuracy. The model graph is shown above. Also after increasing the batch size and running the model, it took more time to run the model and process the output.

Conclusion:

Hence, as the batch size used was small while training the model the resulted output is seems to be smooth. However, the increase in it makes the metrics smooth.

.

References:

<http://www.wildml.com/2015/12/implementing-a-cnn-for-text-classification-in-tensorflow/>

<https://github.com/dennybritz/cnn-text-classification-tf>

<https://github.com/dennybritz/cnn-text-classification-tf/issues/127>

<https://github.com/jiegzhan/multi-class-text-classification-cnn>