CS 5590 – Python and Deep learning

Deep Learning

Lab Assignment-2

Submitted By

Vamsi Krishna Challa

16242688

**Introduction:**

This lab assignment is very much about implementing text classification and depicting it using tensor flow for the classification performed over convolution neural networks. We are asked to perform text classification with CNN over a new data set that is not previously used in the class. Therefore, I have considered using few personalised data sets for this purpose. We were also asked to perform analysis by changing multiple hyper parameters, the results for these new datasets are obtained. The accuracy of system in all the cases is noted down and analysis is done using that data.

**Objective:**

* The primary objective of our assignment is to implement text classification on CNN i.e., convolutional neural networks over new datasets.
* Now we would need to evaluate the changes put on hyper parameters and calculate the trends of the accuracy and losses over the values considered.

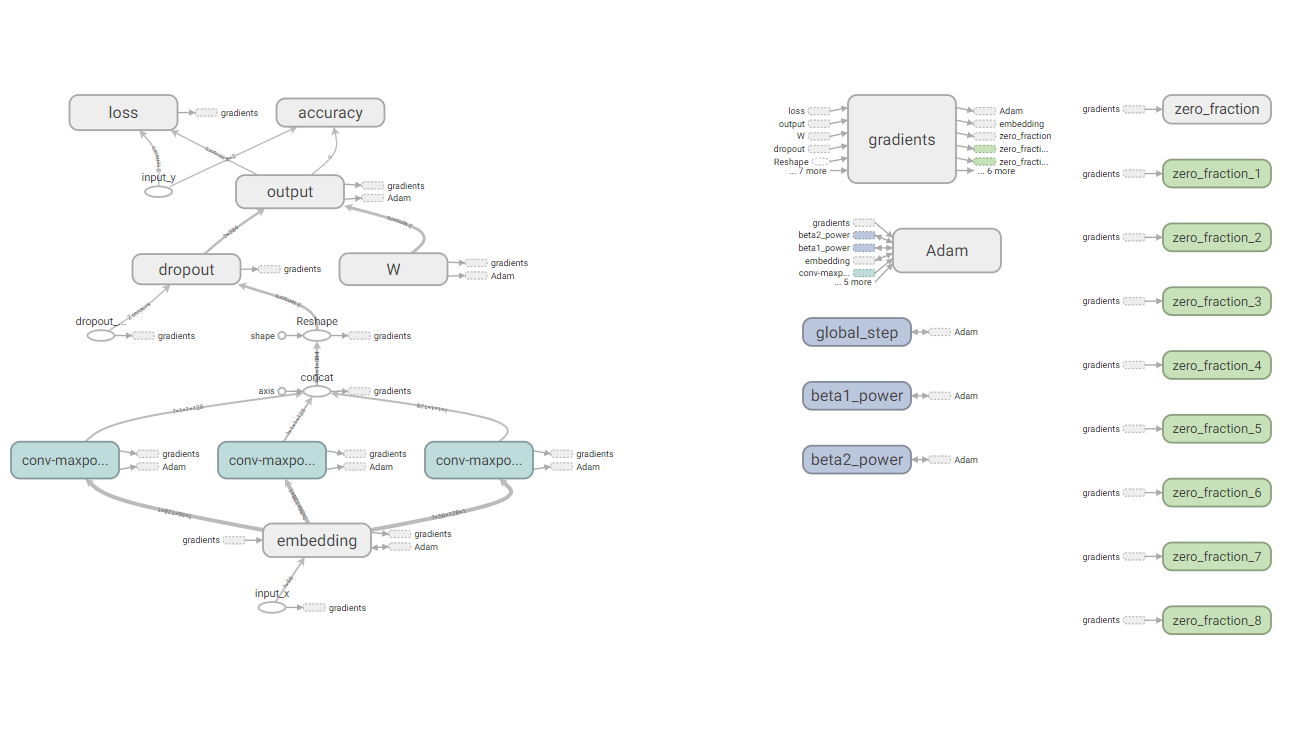
**Approaches and methods:**

* Here we would consider the following datasets in the data folder
* We are going to use Adam optimizer to construct the model.
* We would need to initially read the data line by line, build the model and train the model over the considered datasets.
* We need to process the raw data in datasets by reading the data, optimizing it over the spaces, special symbols, and generate labels for every word in the dataset.
* We need to create placeholders for X, Y variables over features and labels respectively.
* I would now split the data into train data and test data in ratio of 20/80 respectively.
* We would store the summary of the session generated in the designated folder.
* The procedure for training the dataset here considered will be defined using the global step, descent-optimizer and training-set-optimizer.

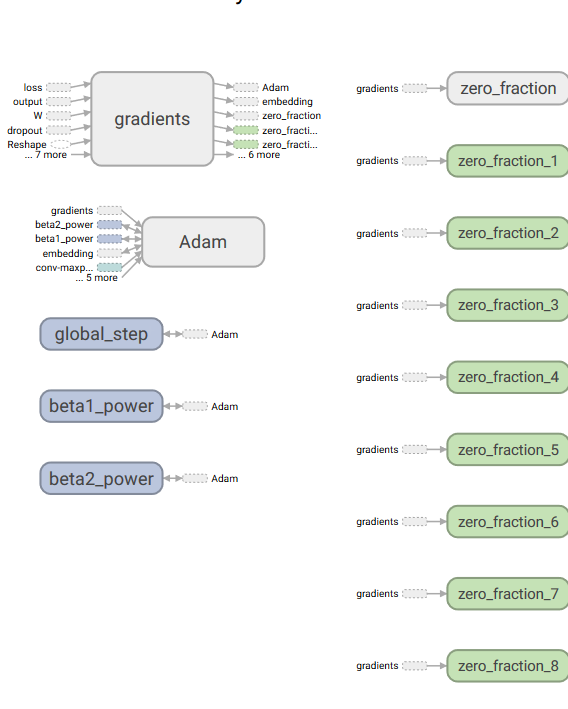
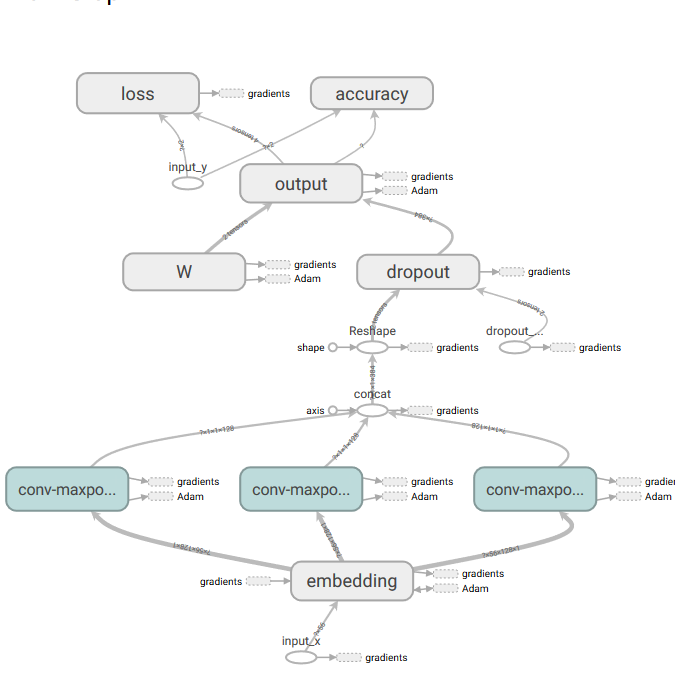
**Workflow:**

* Obtain the new datasets, on which we need to perform text classification using convolutional neural networks.
* We update the hyper parameters and note the trends of accuracy and loss for change in the parameters.

You Can see the workflow in the graph below more clearly:



For more clear understanding look at the individual parts of the graph:



**Configuration:**

* PyCharm 3.4
* Tensor Board

**Datasets:**

Here I am considering IPL Matches dataset, which have many classes in them, but we consider only few for our working.

Source to dataset: https://www.kaggle.com/manasgarg/ipl/data

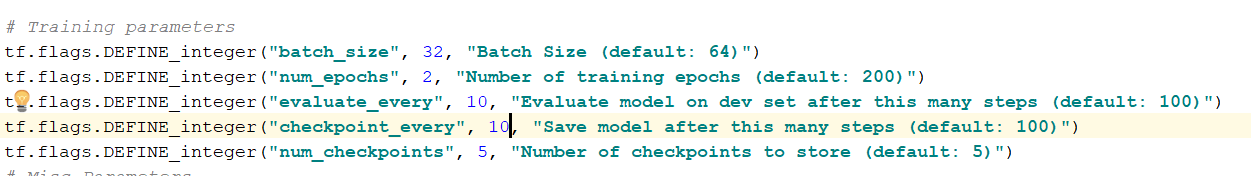
**Parameters:**

Batch size, number of epochs, number of checkpoints, learning rate etc are the parameters that I have considered here.

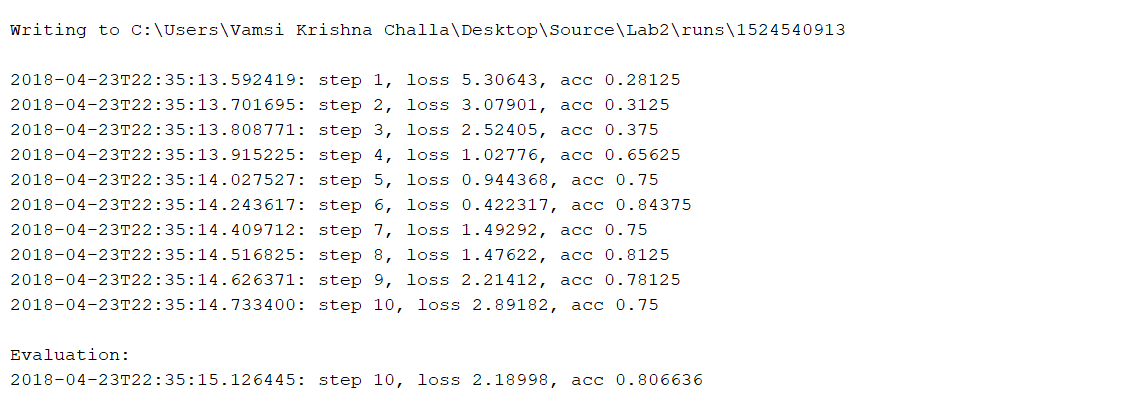
**Evaluation and Discussion:**

Using data\_eval.py file, we would initially tokenize the input data that is considered for evaluation. This would remove the unnecessary content from the dataset and make it ready for evaluation. Now we need to evaluate the effects of changing the parameters on accuracy and loss. Now for this purpose we change the learning rate parameter for this evaluation purpose.

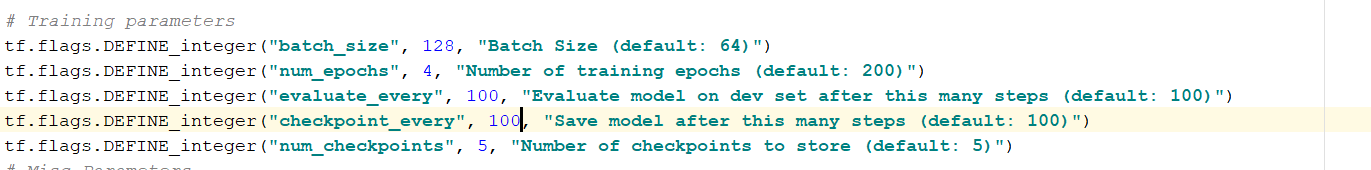
Initially parameters were set at



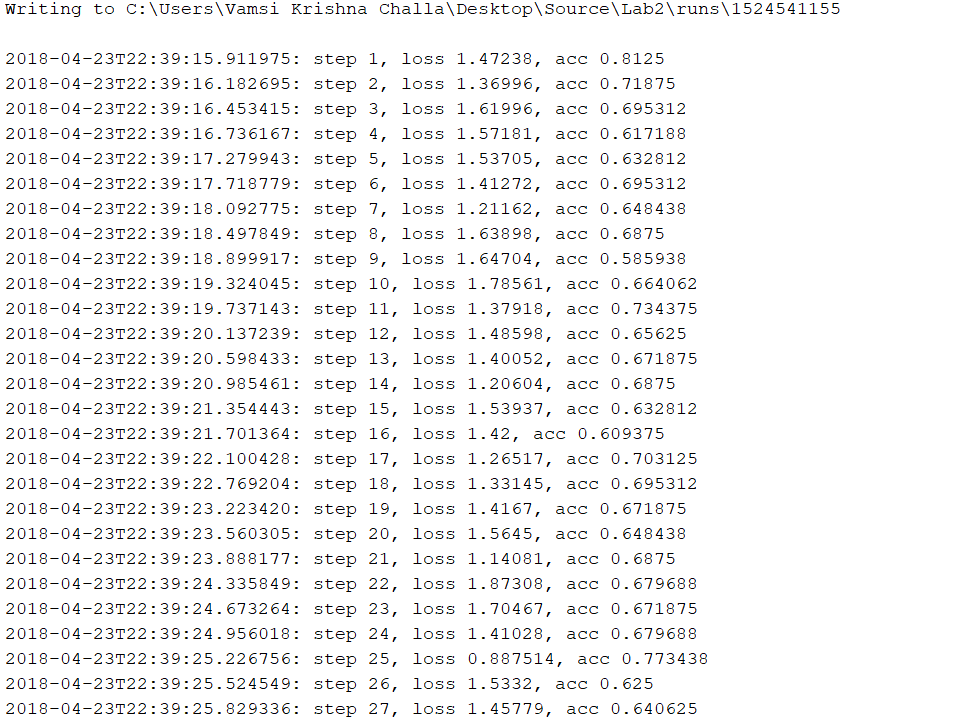
Results obtained are:



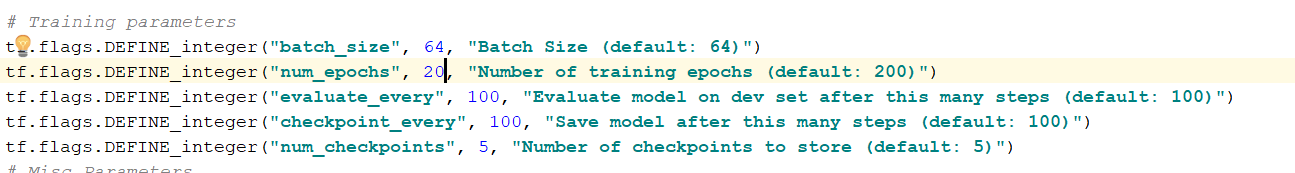
Later the parameters were changed to:



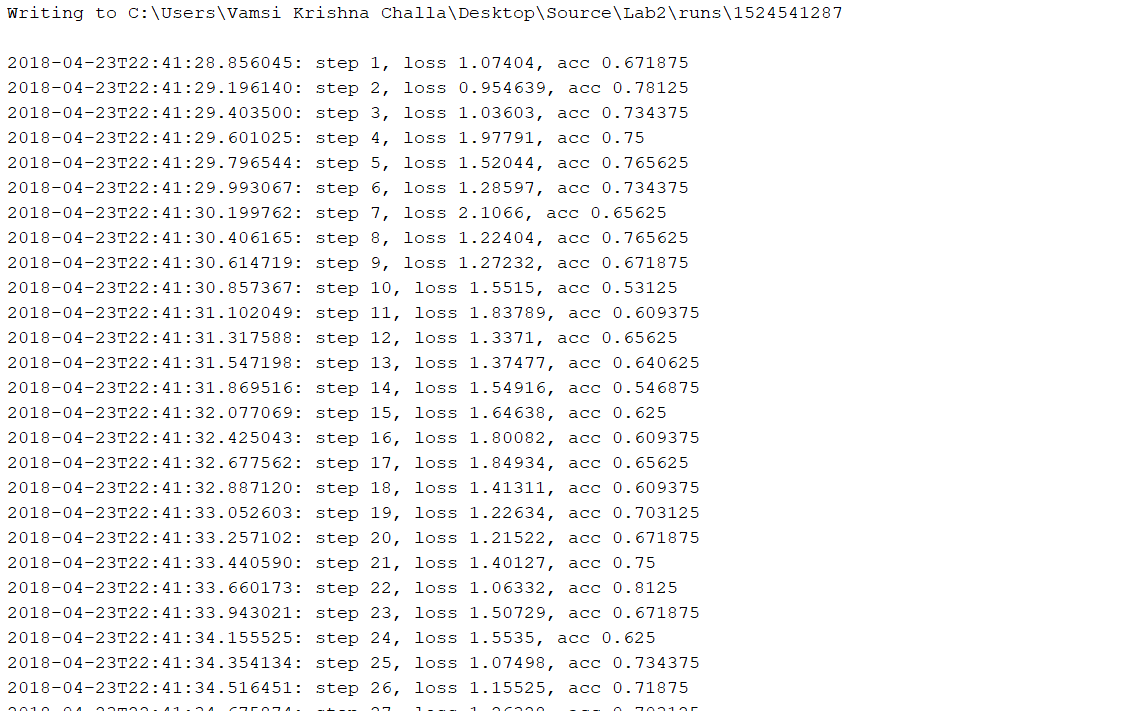
Results obtained are:



Later I have checked by changing the parameters:



Results obtained:



**Conclusion:**

I have obtained optimum results on this training set. We might have got better accuracy rates when performing classification using RNN’s Etc. However, when working with CNN, we have achieved optimum performance of the system. Due to low capacity of the system, I have tested the parameters with minimal hyper parameters and obtained the results.