Harshil patel student id:32

ASSIGNMENT 2

Task:

Implement the Text classification with CNN model with new data set (minimum 5 classes) which is not used in class.

In this assignment I have used 5 different class to classify text they are as follow…..

- Art
- Politics
- Sports
- Tech
- Travel

I am using in class program and changing the code where every need to make it work for 5 class with my data set.

I am only using 5 sentences in each class. So, my model accuracy will not be good.

Data_helper python code

This program is used to create data with ans.it will take data text file and specify the content for that text in that.
So it will create x and y value for our model.

```
def load_data_and_labels(a_data_file,
p_data_file,s_data_file,t_data_file,tr_data_file):
    #geting 5 data file as an argument
    a_examples = list(open(a_data_file, "r",encoding='UTF8').readlines())
    a_examples = [s.strip() for s in a_examples]
    p_examples = list(open(p_data_file, "r",encoding='UTF8').readlines())
    p_examples = [s.strip() for s in p_examples]
    s_examples = list(open(s_data_file, "r", encoding='UTF8').readlines())
    s_examples = [s.strip() for s in s_examples]
    t_examples = list(open(t_data_file, "r", encoding='UTF8').readlines())
    t_examples = [s.strip() for s in t_examples]
    tr_examples = list(open(tr_data_file, "r", encoding='UTF8').readlines())
    tr_examples = [s.strip() for s in tr_examples]
    # Split by words
    x_text = a_examples + p_examples + s_examples + t_examples + tr_examples
    x_text = [clean_str(sent) for sent in x_text]
    # Generate labels
    a_labels = [[1,0,0,0,0] for _ in a_examples]
    p_labels = [[0,1,0,0,0] for _ in p_examples]
    s_labels = [[0, 1, 0, 0, 0] for _ in s_examples]
    t_labels = [[0, 1, 0, 0, 0] for _ in t_examples]
    tr_labels = [[0, 1, 0, 0, 0] for _ in tr_examples]
    y = np.concatenate([a_labels, p_labels,s_labels,t_labels,tr_labels], 0)
    return [x_text, y]
```

train.py

it will be the main program to train the model. By using the data set generated by above program.

Loading data:

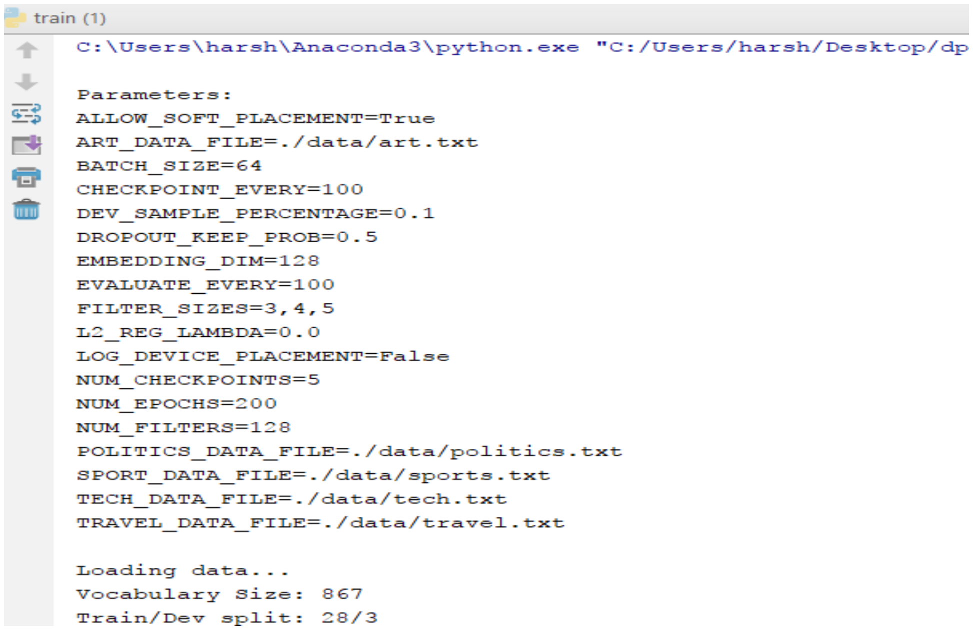
```
tf.flags.DEFINE_string("art_data_file", "./data/art.txt", "Data source for the
positive data.")
tf.flags.DEFINE_string("politics_data_file", "./data/politics.txt", "Data source for
the politics data.")
tf.flags.DEFINE_string("sport_data_file", "./data/sports.txt", "Data source for the
sports data.")
tf.flags.DEFINE_string("tech_data_file", "./data/tech.txt", "Data source for the tech
data.")
tf.flags.DEFINE_string("travel_data_file", "./data/travel.txt", "Data source for the
travel data.")
```

calling data helper:

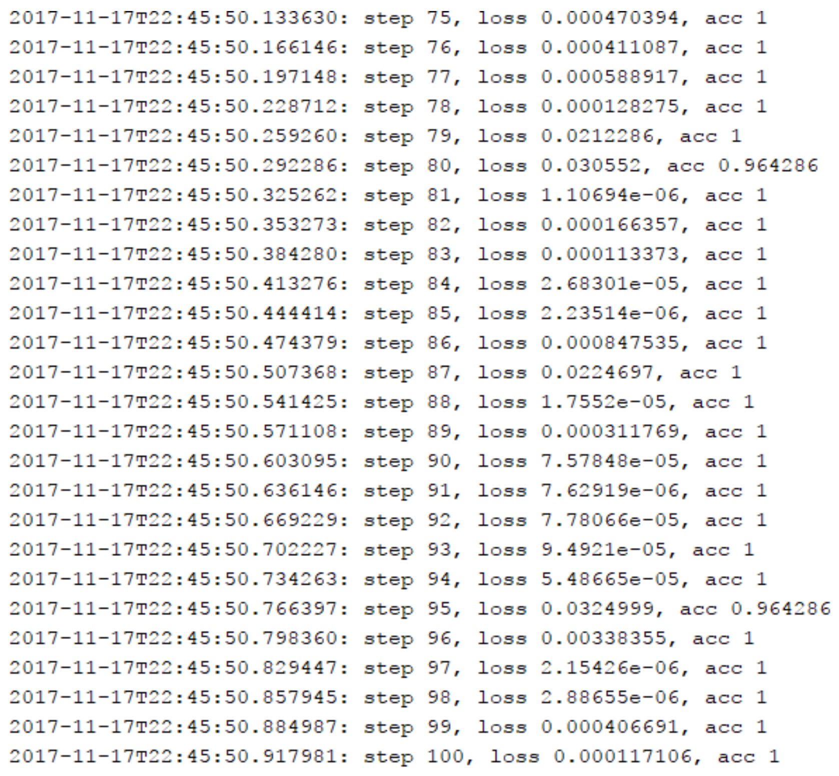
```
x_text, y = data_helpers.load_data_and_labels(FLAGS.art_data_file,
FLAGS.politics_data_file,FLAGS.sport_data_file,FLAGS.tech_data_file,FLAGS.travel_data_
file)
```

output:

Harshil patel student id:32



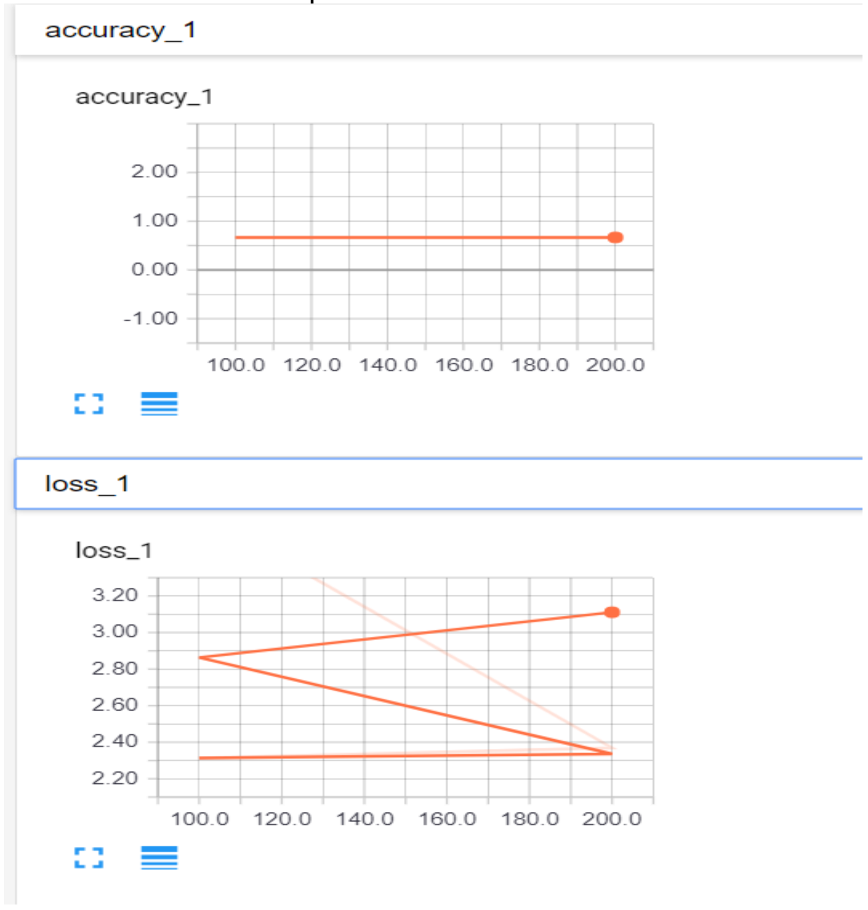
```
C:\Users\harsh\Anaconda3\python.exe "C:/Users/harsh/Desktop/dp

Parameters:
ALLOW_SOFT_PLACEMENT=True
ART_DATA_FILE=./data/art.txt
BATCH_SIZE=64
CHECKPOINT_EVERY=100
DEV_SAMPLE_PERCENTAGE=0.1
DROPOUT_KEEP_PROB=0.5
EMBEDDING_DIM=128
EVALUATE_EVERY=100
FILTER_SIZES=3,4,5
L2_REG_LAMBDA=0.0
LOG_DEVICE_PLACEMENT=False
NUM_CHECKPOINTS=5
NUM_EPOCHS=200
NUM_FILTERS=128
POLITICS_DATA_FILE=./data/politics.txt
SPORT_DATA_FILE=./data/sports.txt
TECH_DATA_FILE=./data/tech.txt
TRAVEL_DATA_FILE=./data/travel.txt

Loading data...
Vocabulary Size: 867
Train/Dev split: 28/3
```



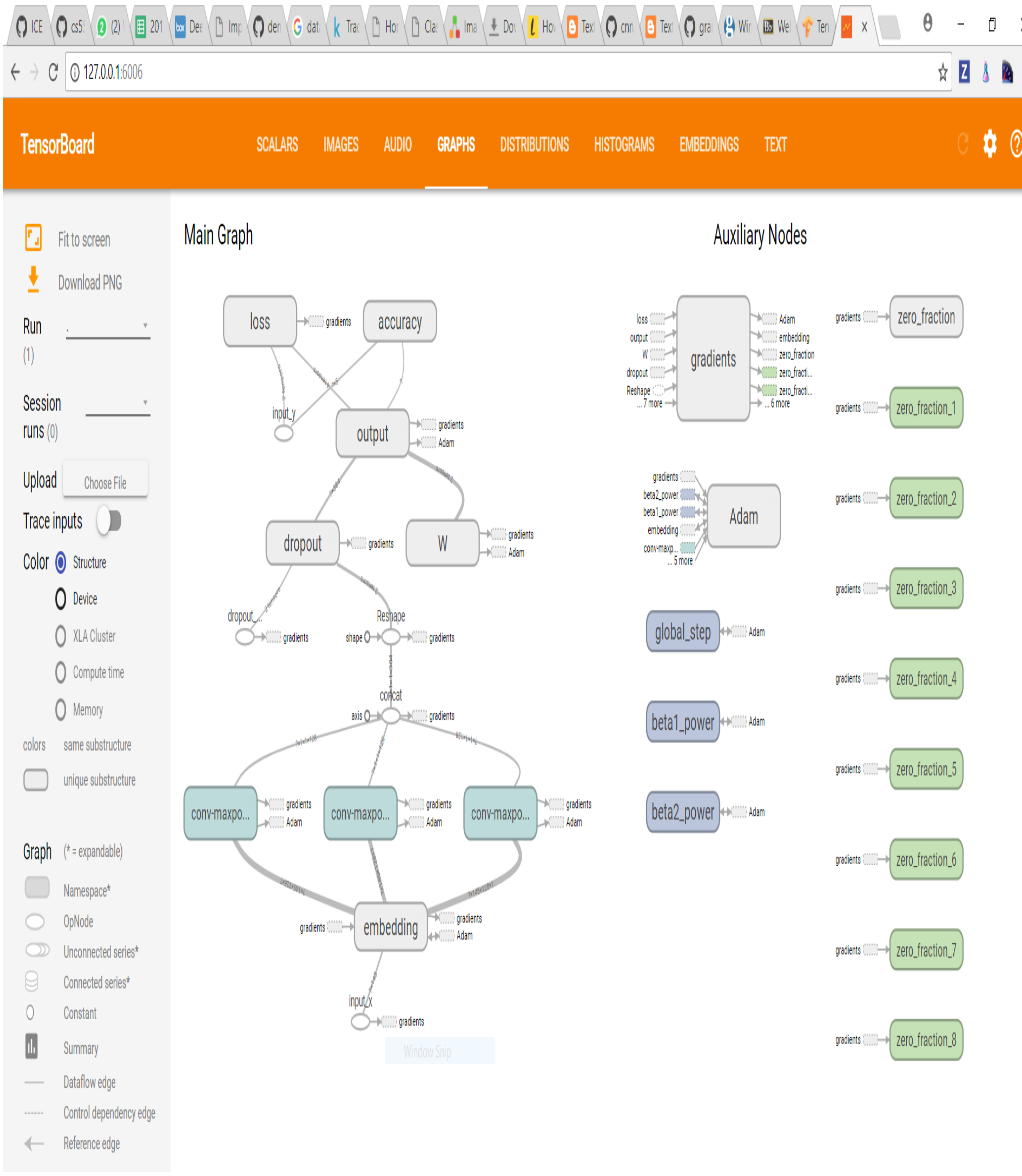
```
2017-11-17T22:45:50.133630: step 75, loss 0.000470394, acc 1
2017-11-17T22:45:50.166146: step 76, loss 0.000411087, acc 1
2017-11-17T22:45:50.197148: step 77, loss 0.000588917, acc 1
2017-11-17T22:45:50.228712: step 78, loss 0.000128275, acc 1
2017-11-17T22:45:50.259260: step 79, loss 0.0212286, acc 1
2017-11-17T22:45:50.292286: step 80, loss 0.030552, acc 0.964286
2017-11-17T22:45:50.325262: step 81, loss 1.10694e-06, acc 1
2017-11-17T22:45:50.353273: step 82, loss 0.000166357, acc 1
2017-11-17T22:45:50.384280: step 83, loss 0.000113373, acc 1
2017-11-17T22:45:50.413276: step 84, loss 2.68301e-05, acc 1
2017-11-17T22:45:50.444414: step 85, loss 2.23514e-06, acc 1
2017-11-17T22:45:50.474379: step 86, loss 0.000847535, acc 1
2017-11-17T22:45:50.507368: step 87, loss 0.0224697, acc 1
2017-11-17T22:45:50.541425: step 88, loss 1.7552e-05, acc 1
2017-11-17T22:45:50.571108: step 89, loss 0.000311769, acc 1
2017-11-17T22:45:50.603095: step 90, loss 7.57848e-05, acc 1
2017-11-17T22:45:50.636146: step 91, loss 7.62919e-06, acc 1
2017-11-17T22:45:50.669229: step 92, loss 7.78066e-05, acc 1
2017-11-17T22:45:50.702227: step 93, loss 9.4921e-05, acc 1
2017-11-17T22:45:50.734263: step 94, loss 5.48665e-05, acc 1
2017-11-17T22:45:50.766397: step 95, loss 0.0324999, acc 0.964286
2017-11-17T22:45:50.798360: step 96, loss 0.00338355, acc 1
2017-11-17T22:45:50.829447: step 97, loss 2.15426e-06, acc 1
2017-11-17T22:45:50.857945: step 98, loss 2.88655e-06, acc 1
2017-11-17T22:45:50.884987: step 99, loss 0.000406691, acc 1
2017-11-17T22:45:50.917981: step 100, loss 0.000117106, acc 1
```

Harshil patel student id:32

I am only using 5 sentences in each class. So, my model accuracy will not be good. And the chart are simple.



Harshil patel student id:32


127.0.0.1:6006
TensorBoard
SCALARS
IMAGES
AUDIO
GRAPHS
DISTRIBUTIONS
HISTOGRAMS
EMBEDDINGS
TEXT
Fit to screen
Download PNG
Run (1)
Session runs (0)
Upload
Choose File
Trace inputs
Color
Structure
Device
XLA Cluster
Compute time
Memory
colors same substructure
unique substructure
Graph (* = expandable)
Namespace*
OpNode
Unconnected series*
Connected series*
Constant
Summary
Dataflow edge
Control dependency edge
Reference edge
Main Graph
Auxiliary Nodes
loss
accuracy
input_y
output
dropout
W
dropout_...
Reshape
shape
concat
axis
conv-maxpo...
embedding
input_x
gradients
Adam
global_step
beta1_power
beta2_power
zero_fraction
zero_fraction_1
zero_fraction_2
zero_fraction_3
zero_fraction_4
zero_fraction_5
zero_fraction_6
zero_fraction_7
zero_fraction_8
... 7 more
... 6 more
... 5 more