Sentiment Analysis-CuDNNLSTM

January 22, 2020

1 Sentiment Analysis with an RNN

Run in Google Colab

View source on GitHub

http://www.polyvista.com/blog/wp-content/uploads/2015/06/sentiment-customer-exp-large.png

1.1 What is Sentiment Analysis?

Sentiment Analysis also know as opinion mining refers to the identification, extraction and study of sentiment states by using natural language processing, text analysis, computational linguistics and biometrics.

1.2 Sentiment Analysis with an Recurrent Neural Network

We will use a RNN for sentiment analysis because we care for the sequence in the data.

1.2.1 Imports

[1]:
```
import re
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt

from tensorflow.keras.models import Sequential, load_model
from tensorflow.compat.v1.keras.layers import CuDNNLSTM, Embedding,␣
↪Dropout,Dense
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.optimizers import RMSprop, Adamax , Adam


# import keras
# from keras.models import Sequential, load_model
```

```
# from keras.layers import Dense, Embedding, Dropout
# from keras.preprocessing.text import Tokenizer
# from keras.preprocessing.sequence import pad_sequences
import tensorflow as tf
from tensorflow.python.client import device_lib
```

[2]:
```
from tensorflow.compat.v1 import ConfigProto, InteractiveSession



config = ConfigProto()
config.gpu_options.per_process_gpu_memory_fraction = 0.6
config.gpu_options.allow_growth = True
session = InteractiveSession(config=config)
```

[3]:
```
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all" #This is for multiple print␣
 ↪statements per cell
```

[4]:
```
value = tf.test.is_gpu_available(
    cuda_only=False,
    min_cuda_compute_capability=None
)
print ('***If TF can access GPU: ***\n\n',value) # MUST RETURN True IF IT CAN!!
```

```
WARNING:tensorflow:From <ipython-input-4-cb50da41978a>:3: is_gpu_available (from
tensorflow.python.framework.test_util) is deprecated and will be removed in a
future version.
Instructions for updating:
Use `tf.config.list_physical_devices('GPU')` instead.
***If TF can access GPU: ***

 True
```

[5]:
```
value = tf.config.list_physical_devices('GPU')
print(value)
```

```
[PhysicalDevice(name='/physical_device:GPU:0', device_type='GPU')]
```

[6]:
```
print(device_lib.list_local_devices())
```

```
[name: "/device:CPU:0"
device_type: "CPU"
memory_limit: 268435456
locality {
}
incarnation: 8209267694858045925
, name: "/device:XLA_CPU:0"
```

```
device_type: "XLA_CPU"
memory_limit: 17179869184
locality {
}
incarnation: 8311693489633492000
physical_device_desc: "device: XLA_CPU device"
, name: "/device:XLA_GPU:0"
device_type: "XLA_GPU"
memory_limit: 17179869184
locality {
}
incarnation: 5503782070576639138
physical_device_desc: "device: XLA_GPU device"
, name: "/device:GPU:0"
device_type: "GPU"
memory_limit: 1259942707
locality {
  bus_id: 1
  links {
  }
}
incarnation: 12306573854068510633
physical_device_desc: "device: 0, name: GeForce MX150, pci bus id: 0000:02:00.0,
compute capability: 6.1"
]
```

[7]:
```
tf.debugging.set_log_device_placement(True)
```

[8]:
```
 tf
print("Num GPUs Available: ", len(tf.config.experimental.
 ↪list_physical_devices('GPU')))
```

[8]:
```
<module 'tensorflow' from '/home/erolerten/anaconda3/envs/venv-
tensorflow/lib/python3.7/site-packages/tensorflow/__init__.py'>

Num GPUs Available:  1
```

2 Place tensors on the CPU

3 with tf.device('/GPU:0'):

a = tf.constant([[1.0, 2.0, 3.0], [4.0, 5.0, 6.0]]) b = tf.constant([[1.0, 2.0], [3.0, 4.0], [5.0, 6.0]])

c = tf.matmul(a, b) print(c)

3.0.1 Loading in Dataset

[9]:
```
data1 = pd.read_csv('Tweets.csv')
data2 = pd.read_csv('stanford-tweets.csv',sep=',')
# data1 = data1.sample(frac=1).reset_index(drop=True)
# data2 = data2.sample(frac=1).reset_index(drop=True)
print(data1.shape)
print(data2.shape)

data1.head()
data2.head()
```

```
(14640, 15)
(1600000, 2)
```

[9]:

	tweet_id	airline_sentiment	airline_sentiment_confidence	negativereason	negativereason_confidence	airline	airline_sentiment_gold	name	negativereason_gold	retweet_count	text	tweet_coord	tweet_created	tweet_location	user_timezone
0	570306133677760513	neutral	1.0000	NaN	NaN	Virgin America	NaN	cairdin	NaN	0	@VirginAmerica What @dhepburn said.	NaN	2015-02-24 11:35:52 -0800	NaN	Eastern Time (US & Canada)
1	570301130888122368	positive	0.3486	NaN	0.0000	Virgin America	NaN	jnardino	NaN	0	@VirginAmerica plus you've added commercials t…	NaN	2015-02-24 11:15:59 -0800	NaN	Pacific Time (US & Canada)
2	570301083672813571	neutral	0.6837	NaN	NaN	Virgin America	NaN	yvonnalynn	NaN	0	@VirginAmerica I didn't today… Must mean I n…	NaN	2015-02-24 11:15:48 -0800	Lets Play	Central Time (US & Canada)
3	570301031407624196	negative	1.0000	Bad Flight	0.7033	Virgin America	NaN	jnardino	NaN	0	@VirginAmerica it's really aggressive to blast…	NaN			
4	570300817074462722	negative	1.0000	Can't Tell	1.0000	Virgin America	NaN	jnardino	NaN	0	@VirginAmerica and it's a really big bad thing…	NaN			

```
3  2015-02-24 11:15:36 -0800            NaN  Pacific Time (US & Canada)
4  2015-02-24 11:14:45 -0800            NaN  Pacific Time (US & Canada)
```

[9]:

	sentiment	text
0	negative	@switchfoot http://twitpic.com/2y1zl - Awww, t…
1	negative	is upset that he can't update his Facebook by …
2	negative	@Kenichan I dived many times for the ball. Man…
3	negative	my whole body feels itchy and like its on fire
4	negative	@nationwideclass no, it's not behaving at all…

Removing all columns except the airline_sentiment and text column.

[10]:
```
data1 = data1[['airline_sentiment', 'text']]
new_columns = ['sentiment','text']
data1.columns = new_columns
data1.head()
```

[10]:

	sentiment	text
0	neutral	@VirginAmerica What @dhepburn said.
1	positive	@VirginAmerica plus you've added commercials t…
2	neutral	@VirginAmerica I didn't today… Must mean I n…
3	negative	@VirginAmerica it's really aggressive to blast…
4	negative	@VirginAmerica and it's a really big bad thing…

[11]:
```
df = data1.append(data2, ignore_index = True)
print(df.shape)
df
```

```
(1614640, 2)
```

[11]:

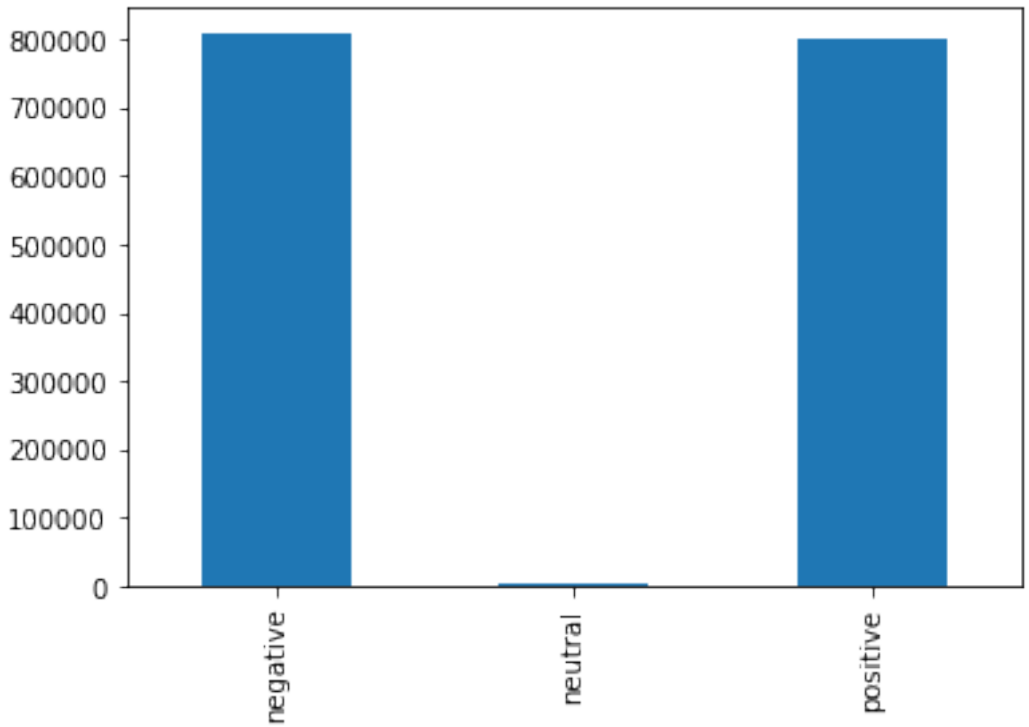
	sentiment	text
0	neutral	@VirginAmerica What @dhepburn said.
1	positive	@VirginAmerica plus you've added commercials t…
2	neutral	@VirginAmerica I didn't today… Must mean I n…
3	negative	@VirginAmerica it's really aggressive to blast…
4	negative	@VirginAmerica and it's a really big bad thing…
…	…	…
1614635	positive	Just woke up. Having no school is the best fee…
1614636	positive	TheWDB.com - Very cool to hear old Walt interv…
1614637	positive	Are you ready for your MoJo Makeover? Ask me f…
1614638	positive	Happy 38th Birthday to my boo of alll time!!! …
1614639	positive	happy #charitytuesday @theNSPCC @SparksCharity…

[1614640 rows x 2 columns]

3.0.2 Data exploration

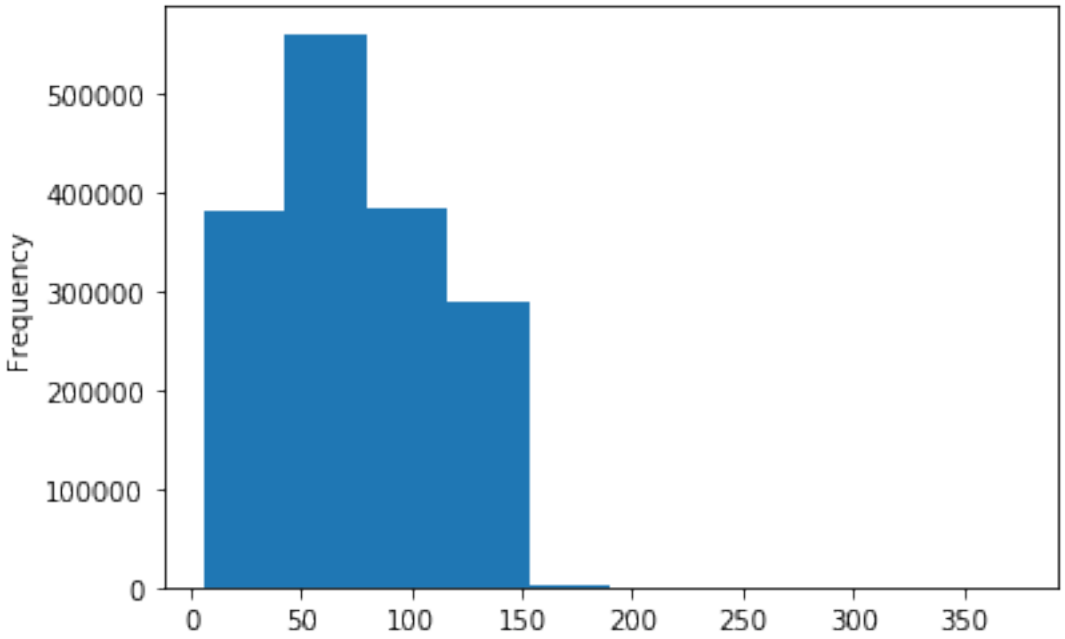
[12]:
```
df['sentiment'].value_counts().sort_index().plot.bar()
```

[12]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2d9a547a90>
```



[13]:
```
df['text'].str.len().plot.hist()
```

[13]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f2d887a54d0>
```



3.0.3 Preprocessing

[14]:
```
# data['text'] = data['text'].str.replace('@VirginAmerica', '')
# data.head()
df = df.sample(frac=1).reset_index(drop=True)
df
```

[14]:

	sentiment	text
0	negative	Enjoyed a beautiful day of softball and laying…
1	positive	Oh stfu. Imma start callin you yohn cause that…
2	positive	@tommcfly have a nice trip back
3	positive	I miss my little cupcake at home waitin on me …
4	positive	It was yummy…dinner just wasn't the same wit…
…	…	…
1614635	negative	managed 2 work 4 three hrs. today. back to bed
1614636	negative	I wanna tell my mom.. But I know I can't.
1614637	negative	is sick of friends fighting wish some people …
1614638	negative	@duplantier yes ma'am!. I remember u were 16 &…
1614639	negative	@THE_REAL_SHAQ im very upset, coz my brother j…

[1614640 rows x 2 columns]

[15]:
```
df['text'].apply(lambda x: x.lower()) #transform text to lowercase
df['text'] = df['text'].apply(lambda x: re.sub('[^a-zA-z0-9\s]', '', x))
```

```
df['text'].head()
```

```
[15]: 0          enjoyed a beautiful day of softball and laying…
      1          oh stfu. imma start callin you yohn cause that…
      2                            @tommcfly have a nice trip back
      3          i miss my little cupcake at home waitin on me …
      4          it was yummy…dinner just wasn't the same wit…
                                      …
      1614635      managed 2 work 4 three hrs. today. back to bed
      1614636           i wanna tell my mom.. but i know i can't.
      1614637    is sick of friends fighting  wish some people …
      1614638    @duplantier yes ma'am!. i remember u were 16 &…
      1614639    @the_real_shaq im very upset, coz my brother j…
      Name: text, Length: 1614640, dtype: object
```

```
[15]: 0    Enjoyed a beautiful day of softball and laying…
      1    Oh stfu Imma start callin you yohn cause thats…
      2                      tommcfly have a nice trip back
      3    I miss my little cupcake at home waitin on me …
      4    It was yummydinner just wasnt the same without…
      Name: text, dtype: object
```

```
[16]: df['sentiment']
```

```
[16]: 0          negative
      1          positive
      2          positive
      3          positive
      4          positive
                   …
      1614635    negative
      1614636    negative
      1614637    negative
      1614638    negative
      1614639    negative
      Name: sentiment, Length: 1614640, dtype: object
```

```
[17]: df = df[df['sentiment'] != 'neutral']
```

```
[18]: df
```

```
[18]:         sentiment                                               text
      0        negative  Enjoyed a beautiful day of softball and laying…
      1        positive  Oh stfu Imma start callin you yohn cause thats…
      2        positive                      tommcfly have a nice trip back
      3        positive  I miss my little cupcake at home waitin on me …
      4        positive  It was yummydinner just wasnt the same without…
```

```
…          …                                                    …
1614635  negative      managed 2 work 4 three hrs today back to bed
1614636  negative             I wanna tell my mom But I know I cant
1614637  negative  is sick of friends fighting  wish some people …
1614638  negative  duplantier yes maam I remember u were 16 amp u…
1614639  negative  THE_REAL_SHAQ im very upset coz my brother jus…

[1611541 rows x 2 columns]
```

[19]:
```
# from numba import jit, cuda
```

[20]:
```
words = 5000
tokenizer = Tokenizer(num_words=words, split=" ")
tokenizer.fit_on_texts(df['text'].values)

X = tokenizer.texts_to_sequences(df['text'].values)
X = pad_sequences(X) # padding our text vector so they all have the same length
X[:5]
```

[20]:
```
array([[   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
        1016,    4,  321,   31,   12, 2866,    6, 1192,  119,    3,   24,
          50,    2,  256,    3,   47,  142,  230],
       [   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
          81, 1516,  256,    7,  385,  102,   71,    1,  169,    9,    6,
           1,   58,   71,   84,    7,   45,    9],
       [   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,  589,   17,    4,  133,  488,   51],
       [   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    1,   83,    5,
         161, 3006,   23,   76, 3074,   14,   15],
       [   0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    0,
           0,    0,    0,    0,    0,    0,    0,    0,    0,    0,    9,
          26,   20,  388,    3,  244,  376,  670]], dtype=int32)
```

3.0.4 Creating model

[26]:
```
model = Sequential()
model.add(Embedding(words, 256, input_length=X.shape[1]))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
```

```
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256, return_sequences=True))
model.add(Dropout(0.3))
model.add(CuDNNLSTM(256))
model.add(Dropout(0.3))
model.add(Dense(2, activation='sigmoid'))
```

```
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
```

[27]:
```
optimizer_RMS = RMSprop(learning_rate=0.001, rho=0.9)
optimizer_adam = Adam(learning_rate=0.001, beta_1=0.9, beta_2=0.999,␣
 ↪amsgrad=False)
model.compile(loss='binary_crossentropy', optimizer=optimizer_adam,␣
 ↪metrics=['accuracy'])
model.summary()
```

```
Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_1 (Embedding)      (None, 40, 256)           1280000
_________________________________________________________________
dropout_8 (Dropout)          (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_7 (CuDNNLSTM)     (None, 40, 256)           526336
_________________________________________________________________
dropout_9 (Dropout)          (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_8 (CuDNNLSTM)     (None, 40, 256)           526336
_________________________________________________________________
dropout_10 (Dropout)         (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_9 (CuDNNLSTM)     (None, 40, 256)           526336
_________________________________________________________________
dropout_11 (Dropout)         (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_10 (CuDNNLSTM)    (None, 40, 256)           526336
_________________________________________________________________
```

```
dropout_12 (Dropout)         (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_11 (CuDNNLSTM)    (None, 40, 256)           526336
_________________________________________________________________
dropout_13 (Dropout)         (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_12 (CuDNNLSTM)    (None, 40, 256)           526336
_________________________________________________________________
dropout_14 (Dropout)         (None, 40, 256)           0
_________________________________________________________________
cu_dnnlstm_13 (CuDNNLSTM)    (None, 256)               526336
_________________________________________________________________
dropout_15 (Dropout)         (None, 256)               0
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 514
=================================================================
Total params: 4,964,866
Trainable params: 4,964,866
Non-trainable params: 0
_________________________________________________________________
```

[28]:
```
y = pd.get_dummies(df['sentiment']).values
[print(df['sentiment'][i], y[i]) for i in range(0,5)]
```

```
negative [1 0]
positive [0 1]
positive [0 1]
positive [0 1]
positive [0 1]
```

[28]:
```
[None, None, None, None, None]
```

[29]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=0)
```

3.0.5 Training model

[30]:
```
batch_size = 32
epochs = 12

import time

start = time.time()
model.fit(X_train, y_train, epochs=epochs, batch_size=batch_size, verbose=2)
end = time.time()
elapsed = end - start
```

```python
print(elapsed/60," minutes")
```

```
Executing op RangeDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op RepeatDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op MapDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op PrefetchDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op FlatMapDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op TensorDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op RepeatDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op ZipDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op ParallelMapDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op DatasetCardinality in device
/job:localhost/replica:0/task:0/device:CPU:0
Train on 1289232 samples
Epoch 1/12
Executing op ModelDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op AnonymousIteratorV2 in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op MakeIterator in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op AssignVariableOp in device
/job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op LogicalNot in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op Assert in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
```

```
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op VarHandleOp in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op __inference_distributed_function_8464 in device
/job:localhost/replica:0/task:0/device:GPU:0
1289232/1289232 - 3881s - loss: 0.4347 - accuracy: 0.7981
Epoch 2/12
1289232/1289232 - 3898s - loss: 0.4013 - accuracy: 0.8185
Epoch 3/12
1289232/1289232 - 3877s - loss: 0.3934 - accuracy: 0.8227
Epoch 4/12
1289232/1289232 - 3850s - loss: 0.3901 - accuracy: 0.8249
Epoch 5/12
1289232/1289232 - 3869s - loss: 0.3897 - accuracy: 0.8248
Epoch 6/12
1289232/1289232 - 3877s - loss: 0.3903 - accuracy: 0.8245
```

```
Epoch 7/12
1289232/1289232 - 3868s - loss: 0.3920 - accuracy: 0.8239
Epoch 8/12
1289232/1289232 - 3876s - loss: 0.3961 - accuracy: 0.8223
Epoch 9/12
1289232/1289232 - 3874s - loss: 0.3963 - accuracy: 0.8217
Epoch 10/12
1289232/1289232 - 3879s - loss: 0.3989 - accuracy: 0.8203
Epoch 11/12
1289232/1289232 - 3870s - loss: 0.4019 - accuracy: 0.8194
Epoch 12/12
1289232/1289232 - 3870s - loss: 0.4023 - accuracy: 0.8183
Executing op DeleteIterator in device
/job:localhost/replica:0/task:0/device:CPU:0
```

[30]:
```
<tensorflow.python.keras.callbacks.History at 0x7f2c7d8a9090>

774.8522711634636  minutes
```

[31]:
```
model.save('sentiment_analysis-20012020.h5')
```

```
Executing op ReadVariableOp in device
/job:localhost/replica:0/task:0/device:GPU:0
Executing op Identity in device /job:localhost/replica:0/task:0/device:GPU:0
Executing op ReadVariableOp in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op ReadVariableOp in device
/job:localhost/replica:0/task:0/device:GPU:0
Executing op Identity in device /job:localhost/replica:0/task:0/device:GPU:0
```

3.0.6 Testing model

[32]:
```
predictions = model.predict(X_test)

[print(df['text'][i], predictions[i], y_test[i]) for i in range(0, 5)]
```

```
Executing op RangeDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op RepeatDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op MapDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op PrefetchDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op FlatMapDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op TensorDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op RepeatDataset in device
```

```
/job:localhost/replica:0/task:0/device:CPU:0
Executing op ZipDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op ParallelMapDataset in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op ModelDataset in device /job:localhost/replica:0/task:0/device:CPU:0
Executing op AnonymousIteratorV2 in device
/job:localhost/replica:0/task:0/device:CPU:0
Executing op __inference_distributed_function_1460080 in device
/job:localhost/replica:0/task:0/device:GPU:0
Enjoyed a beautiful day of softball and laying by the poolNow its time to start
the work week early  [0.07400684 0.92599314] [0 1]
Oh stfu Imma start callin you yohn cause thats how i say it and i know how much
you love it  [0.9449446  0.05505533] [1 0]
tommcfly have a nice trip back  [0.8786865  0.12131354] [1 0]
I miss my little cupcake at home waitin on me  httpmypictme40EB [0.99174964
0.00825042] [1 0]
It was yummydinner just wasnt the same without yall  [0.97249275 0.02750724] [1
0]
```

[32]:
```
[None, None, None, None, None]
```

[33]:
```
accurate_prediction_count, inaccurate_prediction_count = 0, 0
for i, prediction in enumerate(predictions):
    if np.argmax(prediction)==np.argmax(y_test[i]):
        accurate_prediction_count += 1
    else:
        inaccurate_prediction_count += 1

total_predictions = accurate_prediction_count + inaccurate_prediction_count
print('Number of predictions: ', total_predictions)
print('Number of accurate predictions: ', accurate_prediction_count)
print('Number of false predictions: ', inaccurate_prediction_count)
print('Accuracy: ', accurate_prediction_count/total_predictions)
```

```
Number of predictions:  322309
Number of accurate predictions:  263015
Number of false predictions:  59294
Accuracy:  0.8160336819635815
```

[34]:
```
# pos_count, neu_count, neg_count = 0, 0, 0
# real_pos, real_neu, real_neg = 0, 0, 0
# for i, prediction in enumerate(predictions):
#     if np.argmax(prediction)==2:
#         pos_count += 1
#     elif np.argmax(prediction)==1:
#         neu_count += 1
#     else:
#         neg_count += 1
```

```
#     if np.argmax(y_test[i])==2:
#         real_pos += 1
#     elif np.argmax(y_test[i])==1:
#         real_neu += 1
#     else:
#         real_neg +=1

# print('Positive predictions:', pos_count)
# print('Neutral predictions:', neu_count)
# print('Negative predictions:', neg_count)
# print('Real positive:', real_pos)
# print('Real neutral:', real_neu)
# print('Real negative:', real_neg)
```

3.1 Improvements we could implement

Weight classes (because data is skew)

Train more epochs

Use bigger network

Try other word number

3.2 Resources

Recurrent Neural Networks Explained (my own post and video)

Sentiment Analysis (Wikipedia)

What is the best way to do sentiment analysis with Python? (Quora)

How to Do Sentiment Analysis (Siraj Raval)