Preprocessing:

1.Because there is “tag” function in twitter,so I eliminate “@user” . It means I can focus on their text without “@” noise.

2.There is also hashtag function like #jesus, #romantic , that will affect our result,so I also remove it.

3.Read the json file,extract text and tweet id ,and clean the text with 1. And 2.

4.Use Keras Tokenizer,and use fit\_on\_text + texts\_to\_sequences to turn it into numerical that machine can read it.

5.Turn emotion into onehot matrix.

Method:

1. Build Embedding layer and Dropout
2. Convolution layer and MaxPoling
3. LSTM layer, LSTM usally deal with sentiment analysis problem
4. Dropout,in order to prevent overfiting problem
5. Fully connencted layer ,activation use “softmax”
6. Loss: categorical\_crossentropy ; optimizer: adam

Postprocessing:

1. Label decoding
2. Predict the result,and choose the maximum probability that represent which emotion you want through softmax.
3. Write it to CSV form.
4. Done!

Conclusion:

I tried only 1 LSTM layer and 3 LSTM for different training,it did not gain the performance .

However I try CNN-liked,it improved.

Note: [DM\_Lab2\_Kaggle\_Template\_0.45463](https://github.com/hopewu0320/DM_LAB2/blob/master/DM_Lab2_Kaggle_Template_0.45463.ipynb).ipynb is my Kaggle code in my github