## What’s done

Read in Large Movie Review Dataset

Read in training and test data

Built Doc2Vec model and trained it.

Saved Doc2Vec model to reload in later.

Trained imported NN on dataset, got around 67% accuracy. Used this to gage how my model did when I started creating it

## What’s left to do

Read in Reuters Dataset

Prepare and Process Dataset

Train Doc2Vec model and CNN

Test Accuracy

Train a Word2Vec model to see if it works.

Fine tune models to improve accuracy to the fullest.

Setup GUI

## Changes made for Bug fixing & Improvements

Cut model size in half gave better results. 63 -> 73% accuracy.

Added in Flatten layer to the model to fix an error. Caused by the output having a different shape.

Change output layer to have 2 neurons/.

Reshaped X\_train and X\_test to be 3 dimensional for the Convolutional layer.

Converting lists to numpy arrays for my model to use.

Used a separate Input layer to ensure the shape was the only issue when my model was breaking. Helped to debug it too, making sure I knew it was only the shape and nothing to do with the Conv1D layer.

Different changes to model accuracy results:

* Imported model accuracy : 67%
* Model with 19 layers : 63%
* Model with 10 layers : 73%
* Activation set to sigmoid with 10 layers : 73%
* Activation set to :
* Adding a hidden layer with 250 neurons before output layer : 60% accuracy.

Working on Reuters Dataset.

* Read in the cat.txt file which contains the id’s and labels for each file in training and testing.
* Split the training and testing labels into 2 different lists.
* Had to remove all the \n that was added when the string was split.
* Had to iterate through the arrays and grab the file names to match up with the files when I do the doc\_id.

Tried 3 different ways to import the dataset to suit me

Using shuffle on tagged Documents seems to break by Doc2Vec training

Took out the shuffle and it trained successfully.

Got it working and got my model trained

Had to reshape the training and testing data correctly

98% accuracy on my model.

Might have been misleading ^

Updated tensorflow version to solve some issues and had to fix more issues

Needed to squeeze the train and test labels.

Acc around 64% atm.

Need to improve.

gensim.utils.simple\_preprocess vs. word\_tokenize

gensim is better. Gives better results for the model.

From changing to gensim. Acc 64 -> 67%

Activation = relu. Acc is around 67

Activation = tanh. Acc is around 66.

Changed layers to only 1 Conv1D layer, 2 dropout, 1 max pooling1D.

Acc went up to 71%.

More tinkering with layers might improve my accuracy.

GUI Issues – Multi-Threading – stop it from freezing

Set up queues to pass variables around to update the GUI as it runs