Sentiment Analysis

What is Sentiment ?

- Sentiment = feelings
  - Attitudes
  - Emotions
  - Opinions
- Generally, a binary opposition in opinions is assumed
- For/against, like/dislike, good/bad, etc.

What is Sentiment Analysis?

- Using NLP, statistics or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit.
- Sometimes referred to as opinion mining, although the emphasis in this case is on extraction.

Challenges in Sentiment Analysis

- People express opinions in complex ways
- In opinion texts, lexical content alone can be misleading
- Intra-textual and sub-sentential reversals, negation, topic change common
- Rhetorical devices/modes such as sarcasm, irony, implication, etc.

Techniques used for sentiment analysis

- Random Forest
- LSTM

Dataset

- Rotten Tomatoes movie review dataset
  - Has five labels :
    - Negative
    - Somewhat negative
    - Neutral
    - Somewhat positive
    - Positive
  - Training Data : 156060 sentences
  - Test Data : 66292

Dataset

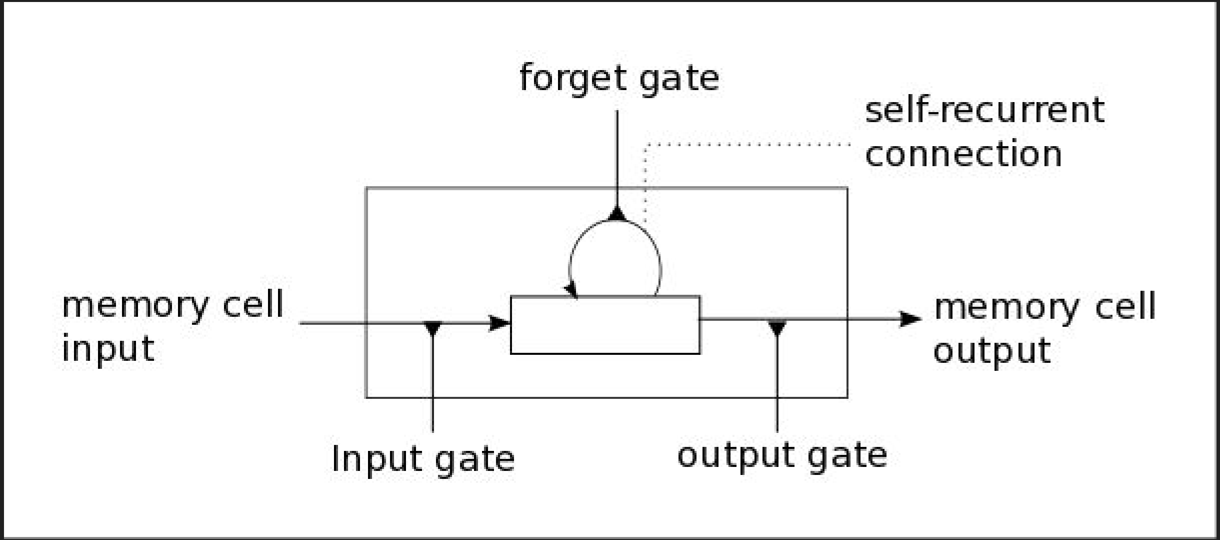
- Large Movie Review Dataset
  - Has two labels :
    - Positive
    - Negative
  - Training Data : 25000
  - Test Data : 25000

Random Forest

- Prediction is done from a combination of trees built on random subset of data and random feature set at each split in the tree.
- Features : n-gram and wordnet based features

LSTM

- Implemented Recurrent Neural Network (RNN) using the Long Short Term Memory (LSTM) architecture using Theano
- Vanishing Gradient Problem



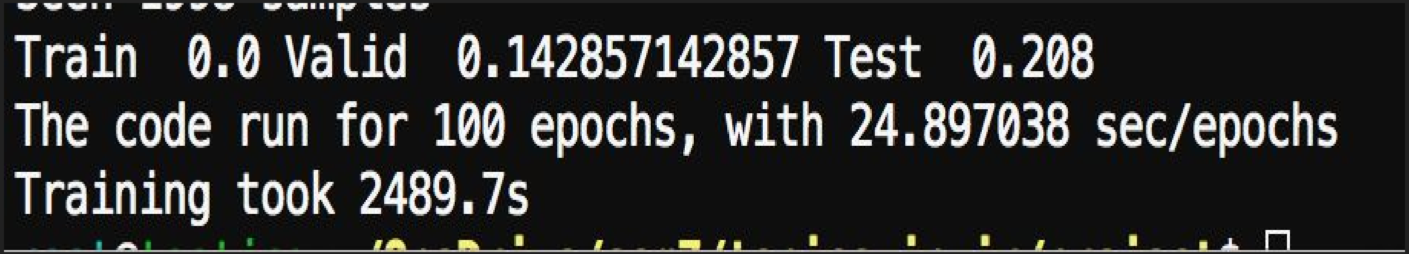
LSTM Architecture

- Our model is composed of a single LSTM layer followed by an average pooling and a logistic regression layer
- From an input sequence x_0, x_1, x_2, ..., x_n, the memory cells in the LSTM layer will produce a representation sequence h_0, h_1, h_2, ..., h_n.
- This representation sequence is then averaged over all timesteps resulting in representation h.
- Finally, this representation is fed to a logistic regression layer whose target is the class label associated with the input sequence

Results

- Accuracy of LSTM
  - Rotten Tomatoes : -
  - IMDB : - 82%
- Accuracy of Random Forest
  - Rotten Tomatoes : - 65 %
  - IMDB : - 70 %

Statistics



```
Train  0.0 Valid  0.142857142857 Test  0.208
The code run for 100 epochs, with 24.897038 sec/epochs
Training took 2489.7s
```

Activation	Validation Error
Sigmoid	0.14
ReLU	0.12
Leaky ReLU (alpha=0.3)	0.12*