ASPECT BASED SENTIMENT ANALYSIS

TEAM NO : 32

Revanth Talluri : 2020102068
Rohit Reddy : 2020102035
Josh Ujwal : 2020102018
SriRam Prasad : 2020102033

OBJECTIVE :

To perform Aspect based sentiment analysis by modelling interdependencies of sentences on a Restaurant review data set with hierarchical bidirectional LSTM.

To implement Simple RNN model and hierarchical Bidirectional LSTM

To compare the accuracies obtained using hierarchical BI-LSTM and a simple RNN model

Approach :

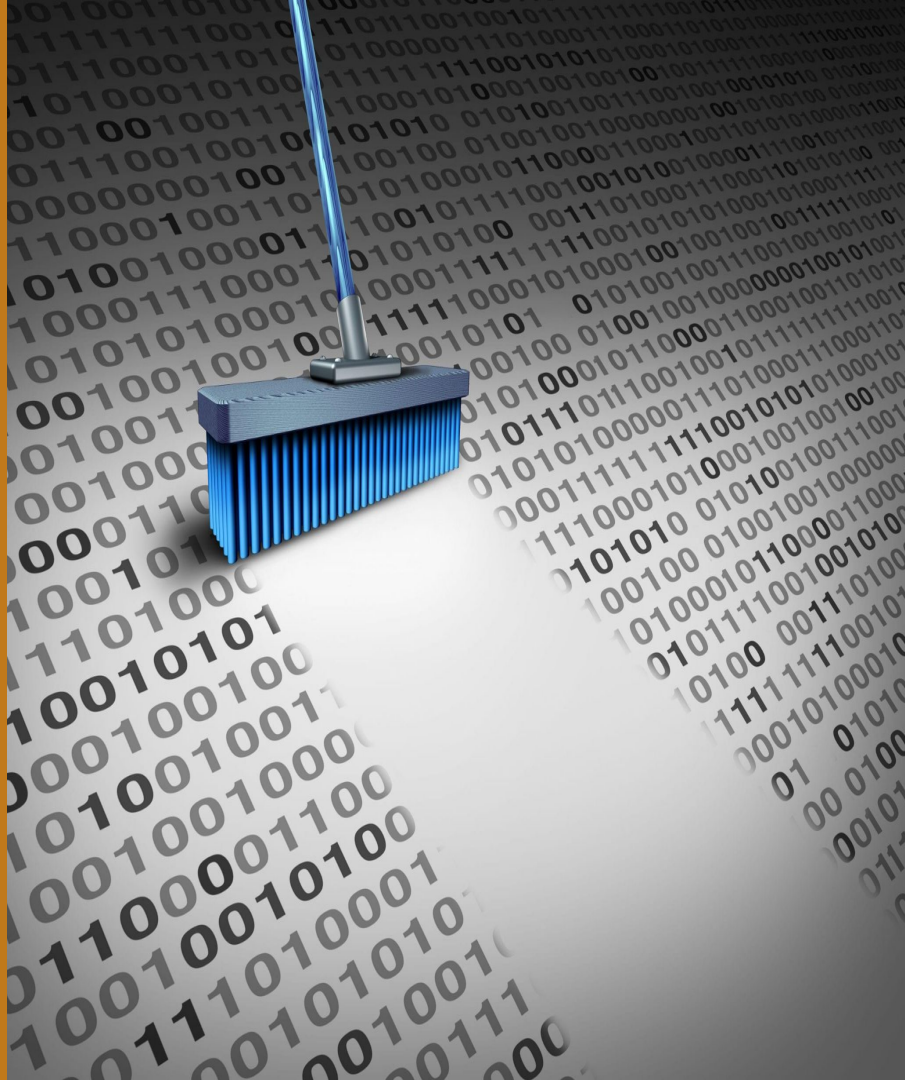
- Learn about Different types of RNN models LSTM,BI-LSTM and HI-BI-LSTM
- Get the Review data set online ,load it ,clean it, do the required padding and stuff and give it as input to a standard RNN model
- Learn about Embeddings and make a dictionary of words to vector mappings which is useful for finding aspect
- Implement the BI-LSTM and HI-BI-LSTM with the same data set and compare accuracies obtained with the simple RNN model

DATA EXTRACTION

- The first step is data extraction.
- Each sentence has two attributes:
    1. Aspect
    2. Entity
- In this step we will extract the required parameters from the dataset.
- So here we arrange the data in a systematic manner in a pandas dataframe with columns as:
    - Review Id (Which review the text belongs to)
    - Aspect
    - text
    - Polarity (positive, negative)

DATA CLEANING

- This is the second step in the process.
- In this step we remove HTML, emotes, links and punctuations from the dataset.
- We will also remove stopwords like “a”, “the”, “is”, “are” etc, from the dataset to make it more efficient.



DATA PREPROCESSING

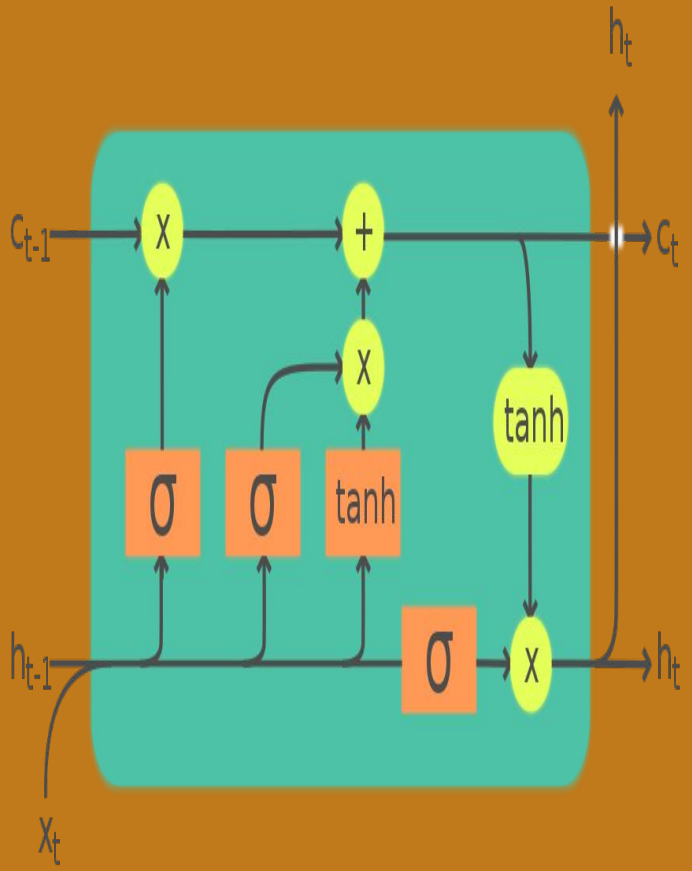
- This is the third step in the process.
- Here we will convert the group of words present in each sentence into vectors of sentence size and represent them using word index.
- So, basically here we will do the tokenizing of words and also do padding to make all vectors of same length, and we will prepare the required deliverables for the embedding matrix.

WORD EMBEDDINGS

- Word Embeddings are a method of extracting features out of text so that we can input those features into a machine learning model to work with text data.
- It is used in our project to convert the entity and attribute of a sentence to it’s corresponding vector.
- Embedding Matrix was created using GLOVE word embeddings for every word in the aspect array.
- Glove.twitter.27B.100d was used, hence each word will be represented as a 100 Dimensional Vector

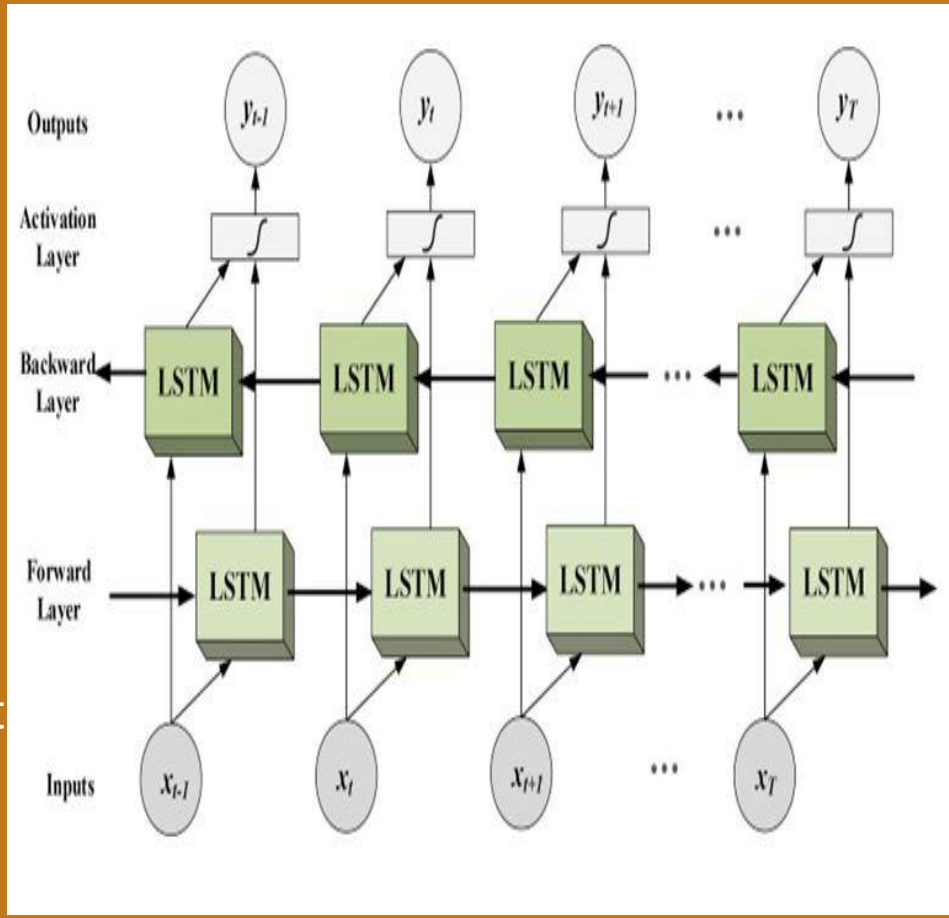
LSTM

- LSTM(Long short-Term Memory) are a special type of RNN(recurrent neural network)
- It is capable of learning long term dependencies
- RNN is unable to learn to connect the information in large gap while LSTM don’t have this problem
- The tanh function gives us the estimate of long term memory/ short term memory to remember



BI-LSTM

- Generally Sentiment of a sentence depends not only on the preceding words but also on the successive words
- Unlike in LSTM where we predict a given word from it's previous words ,In Bi-LSTM we will perform forward LSTM and backward LSTM
- The result is taken by concatenating the states of forward and backward LTSM, some of the sentences many not be differentiated by LSTM but will be differentiated by Bi-LSTM

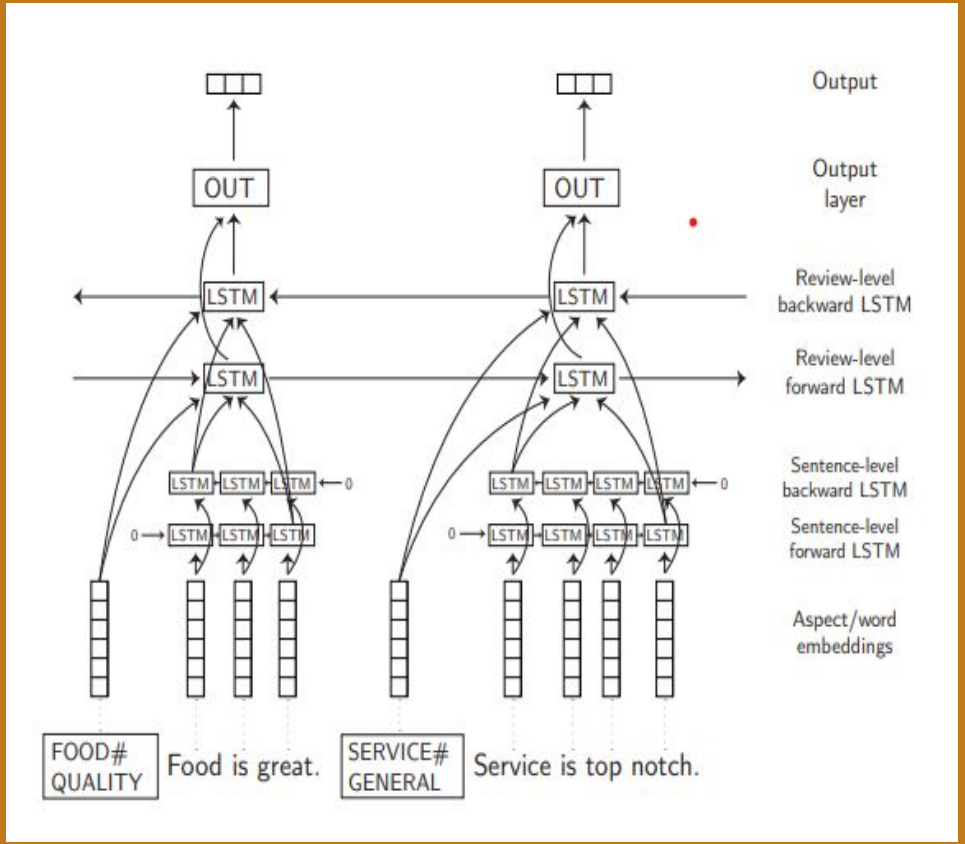


SENTENCE LEVEL CLASSIFICATION WITH BI-LSTM

- The Embedding layer generates a sequence of word-embeddings from given sequence of sentences
- The Bidirectional LSTM generates embeddings from the sequence of word embeddings
- The Dense layer is for the classification

HI-BI-LSTM

- Performing Bi-LSTM on Review level along with sentence level will give us hierarchical bidirectional LSTM
- Concatenation of final output sentence level BI-LSTM with aspect vector is given as input to the Review level BI-LSTM



NOTES ON ASPECT VECTOR

- Every sentence is associated with an aspect. Aspects consist of an entity and an attribute, e.g. FOOD#QUALITY
- An aspect is the combination of these entity and attribute
- Entity and attribute are mapped to a 100 dimensional vector using the embeddings and aspect is calculated by doing average of both of them.

REVIEW LEVEL BI-LSTM AND ACCURACIES

- Final states of BI-LSTM is concatenated with the aspect vector and fed into the review level BI-LSTM
- Calculated the accuracies obtained for simple RNN model and Review Level BI-LSTM

  SimpleRNN : 67.44

  Review LevelBI-LSTM :80.95

Conclusion

We can see that Hierarchical LSTM gives us better accuracies compared to normal Bi-LSTM

THANKYOU