

Natural Language Process
IIT Kanpur
Assignment 1 Report
Date: Sun, Oct 16

Name: Neeraj Kumar

Roll: 13427

Email: neerajkr@iitk.ac.in

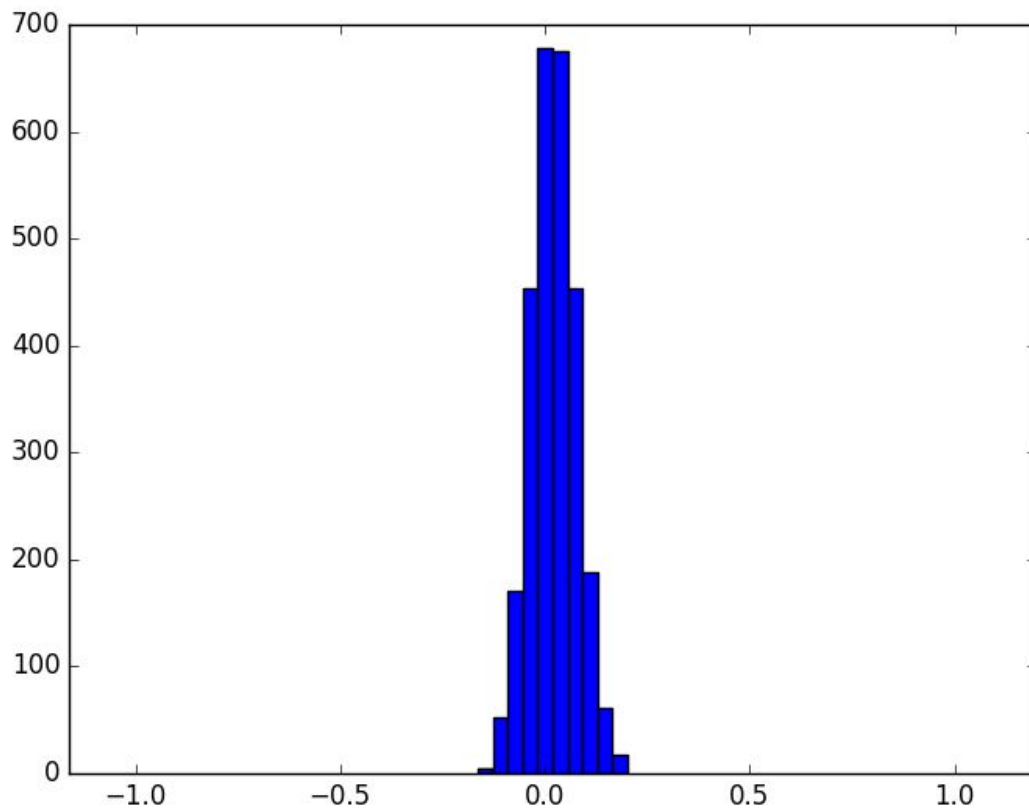
Dept: EE

Question 1:

- a) most frequent tag in the learning corpus
Accuracy= 88.9846 %
Approx. Time= 1 min
- b) Viterbi Algorithm
Accuracy= $64498/94749=68.072486253$ %
Approx. Time= 20 min
- c) LSTM
Accuracy= 99.647709491321923 %
Approx. Time= 5 min
- d) NLTK
Accuracy= 92.7815262913 %
Approx. Time= 15 min
- e) opennlp
Accuracy= 95.952515664 %
Approx. Time= 5 min

Question 2:

Similarity Histogram Between Google Word2Vec and Brown Word2Vec:



Question 3:

- a) Binary Bag of Words (bBoWs) - simple presence/absence.
Accuracy= 51.98 %
Approx. Time= 5 min
- a) BoWs using tf-idf
Accuracy= 50.73 %
Approx. Time= 5 min
- a) Word2Vec vectors averaged for all words in the document
Accuracy= 60.988 %
Approx. Time= 10 min
- a) Bag of Vectors (BoVs) - represent each word by its word2vec
Accuracy= 53.288 %

Approx. Time= 30 min

- a) Weighted Bag of Vectors - weight by tf-idf
Accuracy= 55.367 %
Approx. Time= 30 min