Assignment 3A

Natural Language Processing (CS563)

Department of CSE, IIT Patna

(Read all the instructions carefully and adhere to them.)

Date: 3rd-April-2021 **Deadline**:- 17th-April-2021

Instructions:

1. Markings will be based on the correctness and soundness of the outputs.
2. Marks will be deducted in case of plagiarism.
3. Proper indentation and appropriate comments (if necessary) are mandatory.
4. You should zip all the required files and name the zip file as:
<roll_no>_assignment_<#>.zip, eg. 1501cs11_assignment_01.zip.
5. Upload your assignment (the zip file) in the following link:

https://www.dropbox.com/request/WMtFjtWFir417V1U2iFc

For any queries regarding this assignment contact:

Dushyant Singh Chauhan (dushyantchauhan27@gmail.com) and

Kshitij Mishra (kmishra.kings@gmail.com)

Sentiment Analysis:

Sentiment analysis is an important task in natural language processing and has a wide range of real-world applications. The typical sentiment analysis focuses on predicting the positive or negative polarity of the given sentence(s). This task works in the setting that the given text has only one aspect and polarity. A more general and complicated task would be to predict the aspects mentioned in a sentence and the sentiments associated with each one of them.

As per the above statement, sentiment analysis can be performed in two ways:

A. Sentence-level Sentiment Analysis
B. Aspect-based Sentiment Analysis

(A) Sentence-level Sentiment Analysis

- **Problem Statement:** For a given message (tweet) classify the sentiment of the sentence as positive, negative, or neutral.
    - **Input:** Given Sentence
    - **Output:** Sentiment of the sentence (positive, negative, or neutral).

- **Embedding vectors**
    - Use Fasttext embedding for experiment and download the feature vector from crawl-300d-2M.vec.zip.
    - For more information and usage please go through this link (https://fasttext.cc/docs/en/english-vectors.html)

- **Methods:** LSTM
- **Dataset:** Download the dataset for sentiment classification from here https://drive.google.com/drive/folders/1_IH0MMD7chR7Sxog0ly7kiUgZoe5otd_?usp=sharing
    - There will be three files *twitter-train.tsv*, *twitter-dev.tsv*, and *twitter-test.tsv* in the dataset.
    - Train your model on *twitter-train.tsv* and validate your model on *twitter-dev.tsv* and test your model on *twitter-test.tsv*

- **Evaluation Metrics:** Evaluate your model based on the following metrics:
    - Accuracy
    - F-score
    - Precision and recall
    - Confusion matrix