CS563 - NLP

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

Assignment - 3: Neural Parts-of-speech Tagger

Deadline: 20th May 2020 **Date: 07th May 2020**

Design a Part-of-Speech (POS) tagger which assigns syntactic categories to each word in the text (Brown_Train.txt).

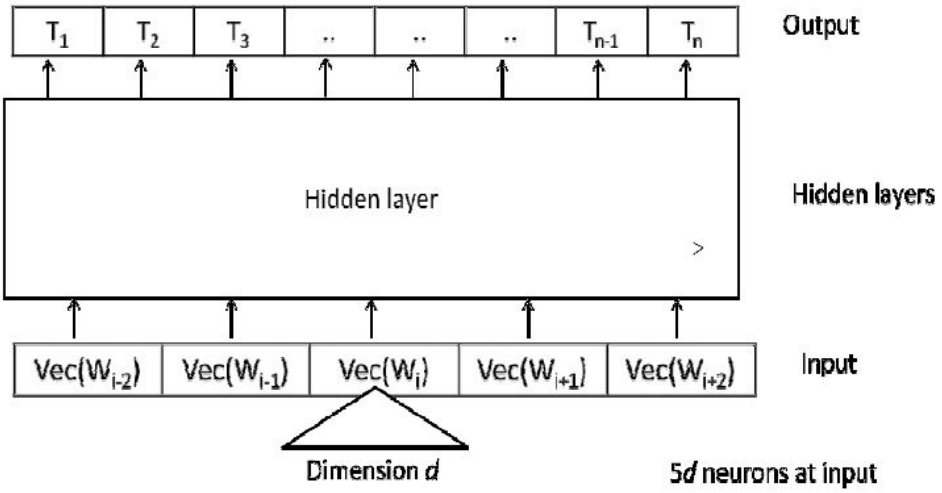
Input: A tokenized sentence.

Output: POS tags for each token of the sentence.

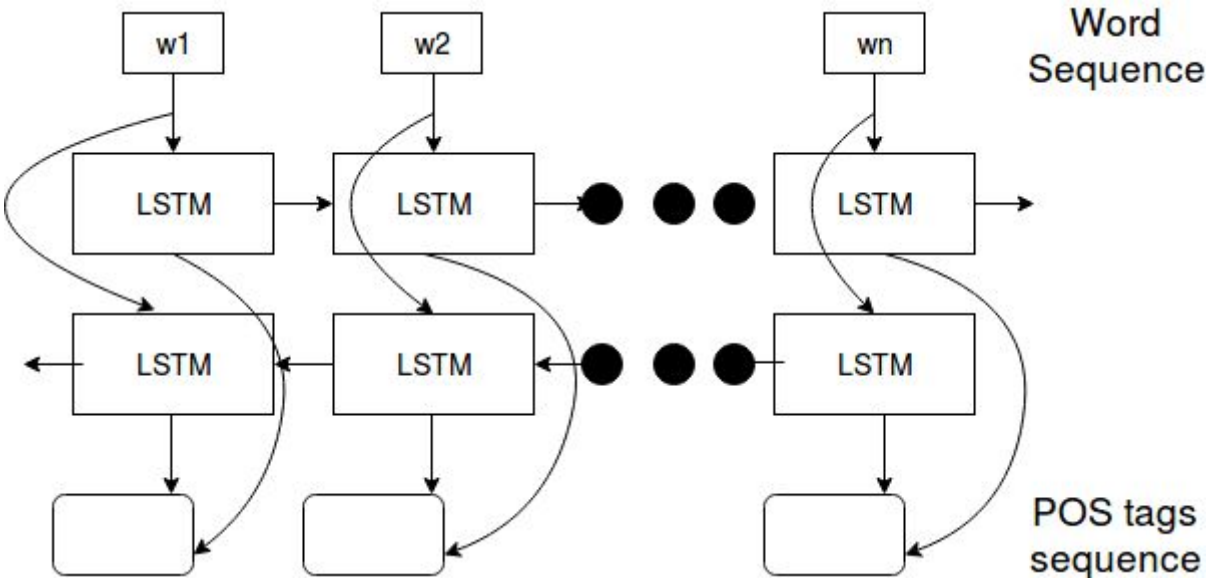
Approach: Solve the problem of POS tagging through the following approaches and compare their performance.

- **Hidden Markov Model (HMM)**
  - You have to implement HMM on your own. Do not use any existing libraries. Calculate emission and transition probabilities and use Viterbi to get the POS tagged sequence.

- **Feed-forward Neural Network:**
  - You may consider the following architecture for the implementation.
    i. Output (Ti): Tags of the POS.
    ii. Input Vec(W_i): Word embedding for the word W_i. Concatenate contextual words (W_{i-2} …. W_{i+2}) to tag W_i



- **Bi-directional LSTM Network**
  - Embed each word with the corresponding pre-trained word embedding.
  - Use a Bi-LSTM to extract the hidden feature (concatenation of forward and backward) for each word.
  - Employ a feed-forward network to classify each word to their corresponding POS tag from a tag set.



You may use any deep learning libraries such as TensorFlow, PyTorch, Keras etc. for the implementation.

Evaluation:

Perform 3-fold cross-validation on the dataset (Brown_Train.txt) and report

- Overall precision, recall and F1-score
- Tag-wise precision, recall and F1-score

- Confusion matrix (Each element Aij of matrix A denotes the number of times tags *i* classified as tag *j*
- Statistics of the tag set.

Dataset Format:

- Each line represents one sentence.
- Sentences are already tokenized.
- Words in a line have the format word_tag.

Submission guidelines:

- Please adhere to the following guidelines while submitting your assignment.
- Please submit your assignment **on or before the deadline**.
- Compress all your files **(Input / Output / Codes / Analysis)** in zip file. It should be named as **Roll-No-Assignment-#.zip**
- Please submit your assignment on "https://bit.ly/3f17O0n**".**