CEP Course on Deep Learning for NLP

Assignment: NER in Tweets

Named-entity recognition (NER) seeks to locate and classify named entities in text into predefined categories such as the names of persons, organizations, locations etc.

Design a named entity recognition system for Twitter that identifies the presence of named entities in a tweet.

Input: A tokenized sentence.
Output: NER tags for each token of the sentence.

Setups:
1. Identify all the named entity, i.e., whether a token is a named entity or not.
2. Identify the named entity types in a sentence.

Approach: Solve the problem of NER through following approaches and compare their performances.
- **Bi-LSTM**
- **CNN**
- **Bi-LSTM followed by CNN**

Dataset: Perform 3 fold cross-validation on the below datasets and report both average & individual fold results.

- **NER-Dataset.txt** (Identify the presence of named entity in a tweet.)
- **CS563-NER-Dataset-10Types.txt** (Identify the presence of named entity and classify them into predefined 10 subtypes. 10 Types are *person, product, company, geolocation, movie, music artist, tvshow, facility, sports team and other*.)
- **Format:**
  - Each line contains <Word \t Tag>
  - Sentences are separated with a blank line.

Evaluation:

Report the Precision, Recall and F-Score for each tag.