Name: Kenny Patel

Roll No. : 19074010

Lab Task-5: Isolated Word Morphological Analysis

Morphological Analysis

Morphology is the study of formation of words, i.e. how words are build up from smaller pieces or how they are structured and formed.

Morphemes are the the most important part of Morphology. Word forms are analysed as arrangements of Morphemes. Hence, Morphemes are the smallest linguistic unit with grammatical function.

Eg: unhappiness : un - prefix ; happy - stem ; ness - suffix

un, happy, ness --> Morphemes

Morphological Analysis is the grammatical analysis of how words are formed by using Morphemes.

इसके अतिरिक्त गुग्गुल कुंड, भीम गुफा तथा भीमशिला भी दर्शनीय स्थल हैं ।

Root	POS	Gender	Number	Person	Case	TAM

इसके == यह pn any sg 3 o ke

In this lab task, we used Multi-Task Deep Morph Analyzer which is a *CNN-RNN model* combined with the potential of task-optimised *phonetic features*. This Analyzer predicts the Lemma, POS category, Gender, Number, Person, Case, and Tense-aspect-mood (TAM) of Hindi and Urdu words.

It was cloned from the repo:

https://github.com/NLPRL/MTDMA-Morph-Analyzer.git

- The root directory 19074010-Kenny-Patel consists of two folders : Morph_Analyzer_Hindi and Morph_Analyzer_Urdu
- Each folder consists of following subfolders: Graphs, Model, Output ; and one Jupyter Notebook
- Model was made to train on each feature of Hindi and Urdu as well as Lemma, hence Output folder consists of individual output files of each feature and one for the Root word.
- Graph folder consists of 6 graphs of one feature each. (By feature, I mean: POS, Gender, Number, etc.)
- Model folder consists of the Model we obtained after training, validating and testing on the dataset.

- Jupyter notebook consists of all the commands I ran for the making up of our model.

- **DataSet** : The Hindi-Urdu Dependency Treebanks were downloaded from this webpage hosted by IIIT-Hyderabad and they were extracted in the datasets folder.
- **Train** : `python main.py --lang (language) --mode train --phonetic true --freezing true`
- **Test** : `python main.py --lang (language) --mode test --phonetic true --freezing true`
- Predict : `python main.py --lang (language) --mode predict --phonetic true --freezing true` language = hindi or urdu

Results:
- Hindi :
    - Accuracy :
        - output0_acc: 0.9556
        - output1_acc: 0.9206
        - output2_acc: 0.9287
        - output3_acc: 0.9401
        - output4_acc: 0.9370
        - output5_acc: 0.9604

- ○ Validation_Accuracy :
  - ■ val_output0_acc: 0.9690
  - ■ val_output1_acc: 0.9407
  - ■ val_output2_acc: 0.9421
  - ■ val_output3_acc: 0.9522
  - ■ val_output4_acc: 0.9481
  - ■ val_output5_acc: 0.9715
- ○ Average precision score, micro-averaged over all classes : 0.6732071283805671
- ● Urdu :
  - ○ Accuracy :
    - ■ output0_acc: 0.9313
    - ■ output1_acc: 0.8833
    - ■ output2_acc: 0.9032
    - ■ output3_acc: 0.9231
    - ■ output4_acc: 0.8740
    - ■ output5_acc: 0.9396
  - ○ Validation_Accuracy :
    - ■ val_output0_acc: 0.9495
    - ■ val_output1_acc: 0.9082
    - ■ val_output2_acc: 0.9192
    - ■ val_output3_acc: 0.9411
    - ■ val_output4_acc: 0.8958
    - ■ val_output5_acc: 0.9575

- Average precision score, micro-averaged over all classes : 0.7154971754964659