

NLP Assignment 2:

Task 2: Dependency Parsing

--Utsav Basu (20CS30057)

Problem Statement

Given the gold-standard dependency graphs and the Oracle (set of heuristics/rules) below, we need to train a simple linear classifier to generate dependency graphs for unknown sentences. The UAS (Unlabelled Attachment Score) metric over the test set is to be reported.

Dataset

We are using the UD-English-GUM dataset. This dataset consists of many sentences already tokenized into words. Each word is annotated with its POS tag and its head word in the dependency graph, along with the type of dependency relation. A fictional root word is added as the root of the dependency graph. Each sentence starts with a unique `# sent_id` and a `# text` describing the text of the sentence.

Methods and Results

Arc-eager parsing is used with the features as described in the problem statement. The model is trained for 10 epochs and the weights are saved in `dependency_model_on.npy`.

As an added metric, the fraction of time the model predicts correct transition on train set is also calculated. It is as follows:

Correct preds / Total preds = 0.8863175824498546

The UAS score obtained on test set is as follows:

UAS Score: 0.5261707988980716