NLP Assignment 3 : Dependency Parsing

Instructor : Prof. Sudeshna Sarkar

TAs : Debanjana Kar, Sukannya Purkayastha, Ishani Mondal

Task : The objective of this assignment is to train and evaluate transition-based dependency parser on cross-lingual corpora.

Corpora: Use the corpora folder uploaded in moodle.

Note : Please refer to the link to write your code.

Experiments : The steps to be followed are as follows -

- Read the CONLLU data into dependency graphs.
  Example : train_graph = DependencyGraph.load(train_path)
- Write the configuration states and corresponding transitions into the training file as required by the learning algorithm in sklearn.
- NLTK trains a SVM classifier to predict the transition corresponding to a configuration. In this assignment, we will train the parser using :
  - Logistic Regression
  - Multi Layer Perceptron Classifier.
  - Use any other classifier of your choice (using sklearn)

- Use the following combination for training and testing the parser:
  - Train on English, test on English
  - Train on Spanish, test on Spanish
  - Train on English, test on Spanish
  - Train on Spanish, test on English

Features:

- The NLTK system by default uses the word, PoS, lemma and morphological features.
- Compare the performance of the parser using different combinations of features (particularly with and without morphological features).

Repeat the experiments for the following cases. A note attached with this assignment on these two types of transition parsing will give you a deeper perspective on this topic. :

- Arc-eager transition.
- Arc-standard transitions.

Deliverables:

• Python code in .ipynb notebook.
• A report containing the results of the experiments, your analysis and observations (for each of the combination stated above) in pdf format.
(Please don’t insert your code in the report)