Graduate Assignment

Implementation and Evaluation of a Transition Based Dependency Parsing Using Universal Dependencies

Introduction:
Transition-based dependency parsing using the English-EWT corpus from the Universal Dependencies. The study incorporated a neural network that utilizes word and POS tag embeddings to enhance the accuracy of parsing actions, a method essential for identifying sentence structures within computational linguistics. This could potentially enhance handling diverse linguistic data, which is crucial for training phases to adapt parsers more effectively across different languages.

Algorithm:
Transition-based dependency parsing is a method in Natural Language Processing (NLP) for constructing syntactic dependency trees by determining the relationships between words in a sentence. The primary objective is to establish dependencies between "head" words and their dependents through a series of parser actions. This approach uses a transition system that iteratively modifies the state of a parse until the entire sentence is parsed into a dependency tree.
Parsing Process
Using these actions, the parser iteratively builds a dependency tree:

- Start with the initial state.
- At each step, the parser chooses an action (SHIFT, LEFT-ARC, RIGHT-ARC) based on the current state analyzed using features like the top words of the stack and buffer and the dependency relations already established.
- Continue applying actions until the buffer is empty and the stack only contains the ROOT and possibly one other node depending on whether the ROOT participates in the dependency relations.

Implementation:
The comprehensive setup for processing dependency parsing data using the conllu format. It includes steps for reading the data, parsing it, and setting up a PyTorch-based neural network model for dependency parsing.
DependencyParsingDataset: A custom class inheriting from `torch.utils.data.Dataset`. This class is tailored to handle dependency parsing datasets formatted in conllu.
Transition Based Dependency Parsing
Main Execution:
A neural network model defined using PyTorch's `nn.Module`. This model is designed to predict parsing actions based on input features word and POS tag indices.
The script parses the sample conllu data to demonstrate how the data can be handled and printed out. This part of the code is mostly illustrative and helps in understanding how to manipulate and access parsed data.

https://universaldependencies.org/udw18/PDFs/9_Paper.pdf