CS224n - Lecture 5

Dependency Parsing means establishing dependency or relatedness links between words of a sentence.

It is essential in resolving the various ambiguities that might arise in natural language sentences.

The approach used by Nivere was similar to the bottom-up parser, with operations: shift, left-arc and right-arc. The strategy worked pretty well, but as it turned out, it requires the construction of a massive amount of sparse features which are expensive to calculate.

This drawback gave rise to the neural approach, which improved the efficiency drastically while achieving similar accuracy.

https://datascience.stackexchange.com/questions/16379/word-embedding-word2vec-for-pos-taggingThe important parts of the neural approach are:

1. The construction of dense input. The input was a triple of words, the POS tags and the label tags. Each of these is a d dimensional dense vector representation. The word was simply the word embeddings. For POS and Label tags, the embeddings were generated similar in spirit to word2vec, the only difference being that the sentences were just the sequences of POS Tags. The intuition behind using the word2vec like dense embeddings of POS and Label tags is that it can be shown that Tags are affected by the context in which they appear. If you read this pm me a smily.

2. Use of cubic activation function: They used a new activation function called the cubic activation function: $(WX + b)^3$. The intuition behind this is that cubic function expansion have a simultaneous correlation between each of the terms.

3. While backpropogating, the input features as well as the weights were updated.