

Morphological Disambiguation

Improve perceptron tagger

The model was first executed by following the instruction given in the tutorial. After this, I experimented with three parameters in the code:

1. `word_frequency(freq_thresh)`: This parameter was chosen to experiment because If let's say a particular word is not occurring N times, then according to the code, it is not being added to the dictionary. Initially this parameter was set to 20 in the base file, it was changed to 18. Here are the results:

Metrics	Precision	Recall	F1 Score	AligndAcc
Tokens	100.00	100.00	100.00	
Sentences	100.00	100.00	100.00	
Words	100.00	100.00	100.00	
UPOS	96.28	96.28	96.28	96.28
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	96.28	96.28	96.28	96.28
Lemmas	100.00	100.00	100.00	100.00
UAS	100.00	100.00	100.00	100.00
LAS	100.00	100.00	100.00	100.00

2. Ambiguity threshold: This parameter was experimented with, initially the value of this was around 0.97, for experimenting it was changed to 0.90. Here were the results:

[(base) kartikaykumar@Kartikays-MacBook-Air conllu-perceptro				
Metrics	Precision	Recall	F1 Score	AligndAcc
Tokens	100.00	100.00	100.00	
Sentences	100.00	100.00	100.00	
Words	100.00	100.00	100.00	
UPOS	96.28	96.28	96.28	96.28
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	96.28	96.28	96.28	96.28
Lemmas	100.00	100.00	100.00	100.00
UAS	100.00	100.00	100.00	100.00
LAS	100.00	100.00	100.00	100.00

As is visible from the figure, ambiguity threshold didn't induce significant changes to the result.

3. Number of epochs (n_iters) and Word_freq (freq_thresh): Lastly, the number of epochs was changed to 10 and freq_thresh to 15 and the following results were obtained:

Metrics	Precision	Recall	F1 Score	AligndAcc
Tokens	100.00	100.00	100.00	
Sentences	100.00	100.00	100.00	
Words	100.00	100.00	100.00	
UPOS	96.30	96.30	96.30	96.30
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	96.30	96.30	96.30	96.30
Lemmas	100.00	100.00	100.00	100.00
UAS	100.00	100.00	100.00	100.00
LAS	100.00	100.00	100.00	100.00

A small change is observed in the accuracy of the model.

This concludes the experiments with the model.