Morphological Disambiguation

Practical: Tagger Comparison

Shreyasi Deshmukh

Report:

The dataset used in this practical is the English ATIS (Airline Travel Information Systems) Corpus.

The taggers used are as follows:

1. Conllu Perceptron Tagger
2. UDPipe Tagger

Based on observation, the part of speech tagged by each of the taggers are mostly similar.

The performance of both the taggers is almost similar, with a UPOS (Universal Part of Speech) score of 98.5%.

The only slight difference that is noticed in the AllTags, and Lemmas Score, where Perceptron performs slightly better, in Precision, Recall and F1 Score.

Results:

1. Perceptron Tagger Result:

```
eval.py --verbose en_atis-ud-test.conllu output.conllu
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       |     98.59 |     98.59 |     98.59 |     98.59
XPOS       |    100.00 |    100.00 |    100.00 |    100.00
Feats      |    100.00 |    100.00 |    100.00 |    100.00
AllTags    |     98.59 |     98.59 |     98.59 |     98.59
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. UDPipe Tagger Result:

```
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       |     98.80 |     98.80 |     98.80 |     98.80
XPOS       |    100.00 |    100.00 |    100.00 |    100.00
Feats      |     98.47 |     98.47 |     98.47 |     98.47
AllTags    |     97.96 |     97.96 |     97.96 |     97.96
Lemmas     |     99.41 |     99.41 |     99.41 |     99.41
UAS        |    100.00 |    100.00 |    100.00 |    100.00
LAS        |    100.00 |    100.00 |    100.00 |    100.00
```