LING L-545/CSCI B-659: Computation and Linguistic analysis

Name: Manasi Swaminathan

Username: mswamina

Morphological Disambiguation

Improve Perceptron tagger

First the tagger.py is being trained and tested with UD_Portuguese.
Before making the changes the evaluation on **pt_gsd-ud-dev.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	99.14	99.14	99.14	99.14
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	99.14	99.14	99.14	99.14
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

After adding the features on context
add('i+3 word', context[i+1]) and **add('i-3 word', context[i-1])**
Evaluation on **pt_gsd-ud-dev.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	99.21	99.21	99.21	99.21
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	99.21	99.21	99.21	99.21
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

After adding features on suffix
add(‘ i suffix’, word[-2:]), **add(‘i suffix, context[i])** and **add(‘i suffix’, context[i-1][-3:])**
Evaluation on **pt_gsd-ud-dev.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	99.27	99.27	99.27	99.27
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	99.27	99.27	99.27	99.27
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

After adding features on prefix
add(‘i+1 word pref1’, context[i+1]), **add(‘i pref1’, word[:3])** and **add(‘i pref1’, word[:2])**
After adding all the additional features. The evaluation on **pt_gsd-ud-dev.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	99.37	99.37	99.37	99.37
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	99.37	99.37	99.37	99.37
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

Evaluation on pt_gsd-ud-test.conllu after all the features adding

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	96.25	96.25	96.25	96.25
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	96.25	96.25	96.25	96.25
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

Evaluation on pt_gsd-ud-test.conllu after all the features adding UD_

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	96.60	96.60	96.60	96.60
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	96.60	96.60	96.60	96.60
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

Now testing it on another language UD_English
Evaluating on en_ewt-ud-test.conllu before adding features

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	93.45	93.45	93.45	93.45
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	93.45	93.45	93.45	93.45
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

Evaluating on en_ewt-ud-test.conllu after adding features

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	93.97	93.97	93.97	93.97
XPOS	100.00	100.00	100.00	100.00
Feats	100.00	100.00	100.00	100.00
AllTags	93.97	93.97	93.97	93.97
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

The feature addition in perceptron tagger has improved the performance on two different languages.