

NLP ASSIGNMENT 3

15CS30033

Spandan Kumar Sahu

Here are the following reported Accuracy:

ARC STANDARD:

	SVM	Logistic Regression	Multilayer Perceptron
Standard Morphological Features	LAS: 0.8609221466364324 UAS: 0.7694633408919124	LAS: 0.8034769463340892 UAS: 0.6893424036281179	LAS: 0.8193499622071051 UAS: 0.7105064247921391
All Morphological Features	LAS: 0.9123204837490552 UAS: 0.8321995464852607	LAS: 0.872260015117158 UAS: 0.7739984882842026	LAS: 0.873771730914588 UAS: 0.780045351473923
No Morphological Features	LAS: 0.8473167044595616 UAS: 0.763416477702192	LAS: 0.800453514739229 UAS: 0.6923658352229781	LAS: 0.798941798941799 UAS: 0.6825396825396826

ARC EAGER:

	SVM	Logistic Regression	Multilayer Perceptron
Standard Morphological Features	LAS: 0.8828420256991686 UAS: 0.792139077853363	LAS: 0.8616780045351474 UAS: 0.751322751322751	LAS: 0.8571428571428571 UAS: 0.748299319727891

	6	3	2
All Morphological Features	LAS: 0.911564625850340 1 UAS: 0.824640967498110 3	LAS: 0.897203325774754 4 UAS: 0.798941798941799	LAS: 0.874527588813303 1 UAS: 0.770975056689342 4
No Morphological Features	LAS: 0.869236583522297 8 UAS: 0.771730914588057 4	LAS: 0.851095993953136 8 UAS: 0.737717309145880 6	LAS: 0.831443688586545 7 UAS: 0.719576719576719 5

LAS = Labeled Attachment Score

UAS = Unlabeled Attachment Score

Conclusions:

1. From the comparisons of the above two tables, inclusion of more morphological features increased the accuracy (both LAS and UAS accuracy). Hence morphological features are important for dependency parsing.
2. All the models were constructed using the default hyper-parameters. The hyper-parameters could be further tuned to increase efficiency. But using the default values for the hyper-parameters, the SVM classification had a consistently higher accuracy compared to the MLP (Multilayer Perceptron) which was further better than the Logistic Regression Classifier.
3. On the type of Dependency Parsing, Arc-Standard was outperformed by Arc-Eager, in almost all the cases.

The default features used by NLTK are:

1. The word itself
2. Lemma
3. PoS tag
4. Feats

The steps to introduce additional features:

1. Append the feature, separated by '[' symbol to the existing column for features (column #6).
2. In this case, I appended the entire 10th column to the 6th column, in a row by row fashion

The steps to remove a morphological features:

1. To remove the entire 6th column, replace it by ' _ ' symbol.
2. To remove individual feature, create a new conllu file by parsing the existing file, and omitting the particular feature whenever it occurs.