

First step was to search all available framework available that I can use to train NER (new language).

List come down to 3 Stanford NLP, AllenNLP and SpaCy.

I was facing problem that SpaCy works very well with the language pack available along with it, but using it to for new language like Hindi was challenging task (without tagger, parser). As it more based on feature engineering so does it have elements for other language. (References:- <https://spacy.io/usage/adding-languages> , <https://spacy.io/usage/linguistic-features> )

I researched for some article that can help me with this but I found none.

Then I was trying for AllenNLP, some of the articles were available but it was difficult in understanding it's working. I wasn't sure that I can do it 3 days.

I took chance on Spacy for 1 day else I have to go with AllenNLP.

In Spacy I changed its vector for Hindi using FastText and changed the given data into spacy format. After some error It started giving good results around 70% , but after studying data I knew that it was low and can do it better.

On searching I have found combination of randomness with mini batch to improve more as more randomness with batch stops overfitting over earlier batches.

Now I have done with SpaCy I am trying with AllenNLP as it has ELMO and BiLSTM CRF so I can use this for complex NER like using it for Hindi News and integrating it with my earlier project.

This assignment was interesting, it challenged me to do something new within limited time.

Thanks

Other references:-

<https://spacy.io/usage/training#ner>

<https://spacy.io/usage/models>