Examine different methods for POS tagging and their pros and cons

Introduction:

Labeling is a sort of arrangement that might be characterized as the programmed task of depiction to the tokens. Here the descriptor is called tag, which might address one of the grammatical form, semantic data, etc.

Presently, assuming we talk about Part-of-Speech (POS) labeling, it could be characterized as the most common way of allocating one of the grammatical features to the given word. It is for the most part called POS labeling. In basic words, we can say that POS labeling is an undertaking of marking each word in a sentence with its proper grammatical feature. We definitely realize that grammatical forms incorporate nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.

A large portion of the POS labeling falls under Rule Base POS labeling, Stochastic POS labeling and Transformation based labeling.

Body:

Probably the most seasoned procedure of labeling is rule-based POS labeling. Rule-based taggers use word reference or vocabulary for getting potential labels for labeling each word. Assuming the word has more than one potential tag, rule-based taggers use manually written standards to distinguish the right tag. Disambiguation can likewise be acted in rule-based labeling by dissecting the phonetic elements of a word alongside its previous just as following words. For instance, assume in the event that the first expression of a word is article, word should be a noun.

One more method of labeling is Stochastic POS Tagging. Presently, the inquiry that emerges here is which model can be stochastic. The model that incorporates recurrence or likelihood (insights) can be called stochastic. Quite a few distinct ways to deal with the issue of grammatical feature labeling can be alluded to as stochastic tagger.

Transformation based labeling is additionally called Brill labeling. It is a case of the transformation based learning (TBL), which is a standard based calculation for programmed labeling of POS to the given text. TBL, permits us to have etymological information in a clear structure, changes one state to one more state by utilizing transformation rules.

It draws the motivation from both the past clarified taggers – rule-based and stochastic.

Assuming we see similitude between rule-based and transformation tagger, similar to rule based, it is likewise founded on the principles that determine what labels should be doled out to what in particular words. Then again, assuming we see comparability among stochastic and transformation tagger, as stochastic, it is AI procedure in which rules are consequently actuated from information.

Conclusion:

Rule-based POS taggers have the accompanying properties –

These taggers are information driven taggers.
The principles in Rule-based POS labeling are constructed physically.
The data is coded as rules.
We have some predetermined number of rules roughly around 1000.
Smoothing and language displaying is characterized expressly in rule-based taggers.

Stochastic POS taggers have the accompanying properties –

This POS labeling depends on the likelihood of tag happening.
It requires preparing corpus
There would be no likelihood for the words that don't exist in the corpus.
It utilizes distinctive testing corpus (other than preparing corpus).
It is the easiest POS labeling on the grounds that it picks most successive labels related with a word in preparing corpus.

The benefits of TBL are as per the following –

We learn little arrangement of straightforward standards and these principles are sufficient for labeling.
Advancement just as troubleshooting is exceptionally simple in TBL on the grounds that the learned guidelines are straightforward.
Intricacy in labeling is decreased on the grounds that in TBL there is joining of machine learned and human-created rules.
Transformation based tagger is a lot quicker than Markov-model tagger.

The drawbacks of TBL are as per the following –

Transformation based learning (TBL) doesn't give label probabilities.
In TBL, the preparation time is extremely long particularly on enormous corpora.

Reference:

Hasan, F. (2006). COMPARISON OF DIFFERENT POS TAGGING TECHNIQUES FOR SOME SOUTH ASIAN LANGUAGES, https://core.ac.uk/download/pdf/61799956.pdf
Kumawat, D& Jain,V (2015). POS Tagging Approaches: A Comparison, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.1924&rep=rep1&type=pdf
Arvindpdmn. (2019, September 8). Part-of-speech tagging. Devopedia. Retrieved November 8, 2021, from https://devopedia.org/part-of-speech-tagging.