Types of POS Tagging

- There are various techniques that can be used for POS tagging such as

1. **Rule-based POS tagging**: The rule-based POS tagging models apply a set of handwritten rules and use contextual information to assign POS tags to words. These rules are often known as context frame rules. One such rule might be: "If an ambiguous/unknown word ends with the suffix 'ing' and is preceded by a Verb, label it as a Verb".

   either we use the dictionary to assign the tags, else then we assign the tags by using hand-written rules based on the context.

2. **Transformation Based Tagging:** The transformation-based approaches use a pre-defined set of handcrafted rules as well as automatically induced rules that are generated during training.

   is some type of combination of both , we assing rule based tags initially and then later using machine models we transform the POS tag based on context and statistics.

3. **Stochastic (Probabilistic) tagging**: A stochastic approach includes frequency, probability or statistics. The simplest stochastic approach finds out the most frequently used tag for a specific word in the annotated training data and uses this information to tag that word in the unannotated text. But sometimes this approach comes up with sequences of tags for sentences that are not acceptable according to the grammar rules of a language. One such approach is to calculate the probabilities of various tag sequences that are possible for a sentence and assign the POS tags from the sequence with the highest probability. Hidden Markov Models (HMMs) are probabilistic approaches to assign a POS Tag.

   word frequency , n grams to assign the pos , or we can use HMMs

-

Issues in POS Tagging

Why is p-o-s tagging hard?

- **Ambiguity**
  - He will race/VB the car.
  - When will the race/NOUN end?
  - The boat floated/' VBN down the river sank.
- **Average of ~2 parts of speech for each word**
- **The number of tags used by different systems varies a lot. Some systems use < 20 tags, while others use > 400.**

Approaches to NLP Tasks

There are 3 main groups of approaches to solving NLP tasks

1.RULE –BASED: Rule-based approaches are the oldest approaches to NLP. It is still used because they are tried and true, and have been proven to work well. Rules applied to text can offer a lot of insight.

- Regular languages and context free grammers are textbook examples of rule-based approaches to NLP.

Rule-based approaches:-

- tend to focus on pattern-matching or parsing
- are low precision, high recall, meaning they can have high performance in specific use cases, but often suffer performance degradation when generalized.

2. . "Traditional" Machine Learning or Probabilistic Modeling

- "Traditional" machine learning approaches include probabilistic modeling, likelihood maximization, and linear classifiers. Notably, these are not neural network models (see those below).

Traditional machine learning approaches are characterized by:

- training data - in this case, a corpus with markup
- feature engineering - word type, surrounding words, capitalized, plural, etc.
- training a model on parameters, followed by fitting on test data (typical of machine learning systems in general)
- inference (applying model to test data) characterized by finding most probable words, next word, best category, etc.
- "semantic slot filling"

Neural Network Model

- This is similar to "traditional" machine learning, but with a few differences:
- feature engineering is generally skipped, as networks will "learn" important features (this is generally one of the claimed big benefits of using neural networks for NLP)
- instead, streams of raw parameters ("words" -- actually vector representations of words) without engineered features, are fed into neural networks
- very large training corpus
- Specific neural networks of use in NLP include recurrent neural networks (RNNs) and convolutional neural networks (CNNs).

- Linguists is the scientific study of language, including its grammar, semantics, and phonetics.
- Classical linguistics involved devising and evaluating rules of language. Great progress was made on formal methods for syntax and semantics, but for the most part, the interesting problems in natural language understanding resist clean mathematical formalisms.
- Broadly, a linguist is anyone who studies language, but perhaps more colloquially, a self-defining linguist may be more focused on being out in the field.
- Mathematics is the tool of science. Mathematicians working on natural language may refer to their study as mathematical linguistics, focusing exclusively on the use of discrete mathematical formalisms and theory for natural language (e.g. formal languages and automata theory).
-

- In a data-driven approach, **decisions are made based on data instead of intuition**. Following a data-driven approach offers measurable advantages. That's because a data-driven strategy uses facts*Data-Drive methods for natural language processing have now become so popular that they must be considered mainstream approaches to computational linguistics. … A strong contributing factor to this development is undoubtedly the increase amount of available electronically stored data to which these methods can be applied; another factor might be a certain disenchantment with approaches relying exclusively on hand-crafted rules, due to their observed brittleness.*

- *We know machine needs to learn,so our task is to create a learning framework and provide properly formatted clean data for machine to learn from.* Machine Learning is **an application of AI that provides systems the ability to automatically learn and improve from experience without** being explicitly programmed. Machine Learning can be used to help solve AI problems and to improve NLP by automating processes and delivering accurate responses.

Feature-based Models

- Features encode evidence from the context for a particular tag:

(title caps, `NNP`)	Citibank, Mr.
(suffix `-ing`, `VBG`)	running, cooking
(POS tag `DT`, `I-NP`)	the bank, a thief
(current word `from`, `I-PP`)	from the bank
(next word `Inc.`, `I-ORG`)	Lotus Inc.
(previous word `said`, `I-PER`)	said Mr. Vinken

Feature-based Tagging

- How do we incorporate features into a probabilistic tagger?
- Hack the Markov Model tagger to incorporate features
  - estimate probabilities directly from feature counts
- Maximum Entropy (MaxEnt) Tagging
  - principled way of incorporating features
  - requires sophisticated estimation method

Unknown Words in Markov Model Tagging

- Calculate $p(w_i|t_i)$ separately for unknown words:

$$p(w_i|t_i) = p(\texttt{unknown}|t_i)\ p(\texttt{caps}|t_i)\ p(\texttt{suffix}|t_i)$$

- Feature probabilities calculated using relative frequencies

- Assumes independence between features
  $\Longrightarrow$ does not account for feature interaction

- Cannot incorporate more complex features

The Model

$$p(t|C) = \frac{1}{Z(C)} \exp\left(\sum_{i=1}^{n} \lambda_i f_i(C,t)\right)$$

- f_i is a feature
- λ_i is a weight (large value implies informative feature)
- $Z(C)$ is a normalisation constant ensuring a proper probability distribution
- Also known as a *log-linear* model
- Makes no independence assumptions about the features

Maximum Entropy Model

- Simple HMM fails to work as desired when sufficient labeled data is not provided to estimate the model parameters.
- Maximum Entropy is an adaptable and flexible modeling system. This Model determines the probabilities based upon constraints. Upon the application of constraints the most probable sequence of tags is produced. These constraints are determined from the preparation information, keeping up connection between the history and probable Outcomes. Outcomes are the sets of permissible tags. Maximum Entropy model permits the estimation of P(t | h) for given t in the space of aggregate conceivable outcomes T, for each 'h' chosen among the set of histories, H
- Information Entropy or Shannon's entropy **quantifies the amount of uncertainty (or surprise) involved in the value of a random variable or the outcome of a random process**

Language Modelling

- Language modeling (LM) is the use of various statistical and probabilistic techniques to determine the probability of a given sequence of words occurring in a sentence. Language models analyze bodies of text data to provide a basis for their word predictions. They are used in natural language processing (NLP) applications, particularly ones that generate text as an output. Some of these applications include , machine translation and question answering.
- In Machine Translation, you take in a bunch of words from a language and convert these words into another language. Now, there can be many potential translations that a system might give you and you will want to compute the probability of each of these translations to understand which one is the most accurate.

Word Ordering

p(The cat is small) > p(small the is cat)

- In the above example, we know that the probability of the first sentence will be more than the second, right? That's how we arrive at the right translation.
- This ability to model the rules of a language as a probability gives great power for NLP related tasks. Language models are used in speech recognition, machine translation, part-of-speech tagging, parsing, Optical Character Recognition, handwriting recognition, information retrieval, and many other daily tasks.

Types of Language Models

There are primarily two types of Language Models:

1. **Statistical Language Models:** Statistical models include the development of probabilistic models that are able to predict the next word in the sequence, given the words that precede it. A number of statistical language models are in use already These models use traditional statistical techniques like N-grams, Hidden Markov Models (HMM) and certain linguistic rules to learn the probability distribution of words

2. **Neural Language Models:** These language models are based on neural networks and are often considered as an advanced approach to execute NLP tasks. Neural language models overcome the shortcomings of classical models such as n-gram and are used for complex tasks such as speech recognition or machine translation. These are new players in the NLP town and have surpassed the statistical language models in their effectiveness. They use different kinds of Neural Networks to model language