SNS COLLEGE OF ENGINEERING

Kurumbapalayam(Po), Coimbatore – 641 97

Accredited by NAAC-UGC with 'A' Grade

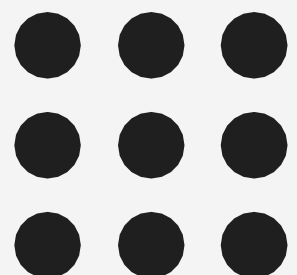**Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai**

Department of Artificial Intelligence and Data Science

Course Name – 19AD601 – Natural Language Processing

III Year / VI Semester

Unit 2 – WORD LEVEL ANALYSIS

Topic 6- Rule based Tagging

Rule Based 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".

There are two stages in rule-based taggers:
First Stage: Uses a dictionary to assign each word a list of potential parts-of- speech.

Second Stage: Uses a large list of handcrafted rules to window down this list to a single part-of-speech for each word.

Rule Based Tagging

The ENGTWOL is a rule-based tagger
– In the first stage, uses a two-level lexicon transducer
– In the second stage, uses hand-crafted rules (about 1100 rules).
Rule-1: if (the previous tag is an article)
 then eliminate all verb tags
• Rule-2: if (the next tag is verb)
 then eliminate all verb tags

Properties of Rule-Based POS Tagging
⬚   Rule-based POS taggers possess the following properties –
⬚   These taggers are knowledge-driven taggers.
⬚   The rules in Rule-based POS tagging are built manually.
⬚   The information is coded in the form of rules.
⬚   We have some limited number of rules approximately around 1000.
⬚   Smoothing and language modeling is defined explicitly in rule-based taggers.

Rule Based Tagging

Example: He had a fly.
The first stage:
– he   he/pronoun
– had   have/verbpast   have/auxiliary past
– a   a/article
– fly   fly/verb fly/noun

The second stage:
apply rule: if (the previous tag is an article)
        then eliminate all verb tags
– he   he/pronoun
– had   have/verbpast   have/auxliarypast
– a   a/article
– fly   fly/verb fly/noun

.

Rule Based Tagging

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.

Transformation-based tagging is also known as Brill Tagging.

Brill Tagging uses transformation rules and rules are learned from a tagged corpus.

Then these learned rules are used in tagging. Before the rules are applied, the tagger labels every word with its most likely tag. We get these most likely tags from a tagged corpus.

.

Brill Tagger algorithm has three major steps.

• Tag the corpus with the most likely tag for each (unigram model)

• Choose a transformation that deterministically replaces an existing tag with a new tag such that the resulting tagged training corpus has the lowest error rate out of all transformations.

• Apply the transformation to the training corpus.

These steps are repeated until a stopping criterion is reached.

• The result (which will be our tagger) will be:

– First tags using most-likely tags

– Then apply the learned transformations in the learning order.

.

Rule Based Tagging

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.

THANK YOU