**Prof. Dr. Andreas Hotho,**
**M.Sc. Janna Omeliyanenko**
Lecture Chair X for Data Science, Universität Würzburg

# 2. Exercise for "Sprachverarbeitung und Text Mining"

12.11.2021

## 1 Knowledge Questions

1. What is Part of Speech Tagging and what may Part of Speech Tagging be used for?

> Part of Speech Tagging refers to the classification of words of a sentence into syntactic units (word types).
>
> As input features for:
>
> - Parsers
> - Information Extraction
> - Speech synthesis
> - Machine Translation

2. What is the difference between open and closed classes of word types in POS-tagging?

> Closed classes, in contrast to open classes, contain only a finite set of words, and new ones cannot be added.
>
> Example: Pronouns, Auxiliaries, Prepositions

3. Which two typical approaches are used for POS-tagging?

   - Rule-based approaches (z.B. ENGTWOL)
   - Stochastic approaches (z.B HMM, MEMM, CRF)

4. Which two types of probabilities are needed in the modeling approach of the Hidden Markov Model? What elements does a Hidden Markov Model generally consist of?

   - Transition probabilities $P(T_i|T_{i-1})$
   - Likelihood probabilities $P(W_i|T_i)$

   Generally, an HMM consists of:

   - A list **Q** with the possible states
   - a matrix **A** with the transition probabilities
   - a list **O** with the possible observations
   - a matrix **B** with observation likelihoods
   - start and end states

5. What does the Markov-assumption state when applied to POS-tagging?

   The probability of a tag depends only on the previous tag.

6. Which POS-tagsets do you know for the German and English language?

   - German: Besides some minor modifications, the STTS (Stuttgart-Tübingen Tagset) is actually always used in German.
   - English: Penn TreeBank tagset

7. All POS-tagging methods presented in the lecture require a list (dictionary) that lists all possible POS-tags for a word. However, such lists are usually not complete. Give a solution for POS-tagging with unknown words.

   - Assume that each tag can be assigned to the unknown word with equal probability.

- Use additional information about word properties and compute probability for each POS-tag given observed properties (multipliable under assumption of independence).
  - e.g. English word ending with 's' $\rightarrow$ noun in the plural with $x\%$ probability.
  - e.g. English word beginning with capital letter $\rightarrow$ proper name with $y\%$ probability.

8. Name three types of evaluation of POS-tagging methods.

- Overall error rate
- Error rates on particular tags
- Error rates on particular words
- Confusion-Matrix

with respect to sample solution (gold standard)

# 2 POS Tagging

Part-of-speech taggers model the words as observations (observed variables), and the associated word types (part-of-speech tags) as hidden variables (hidden states). We assume in the following a highly simplified set of generalized word types (tag-set):

| | | |
|---|---|---|
| DET | Determiner | the,a,... |
| N | Noun | year,home,costs,time,... |
| PRO | Pronoun | he,their,you |
| V | Verb | said,took,saw |

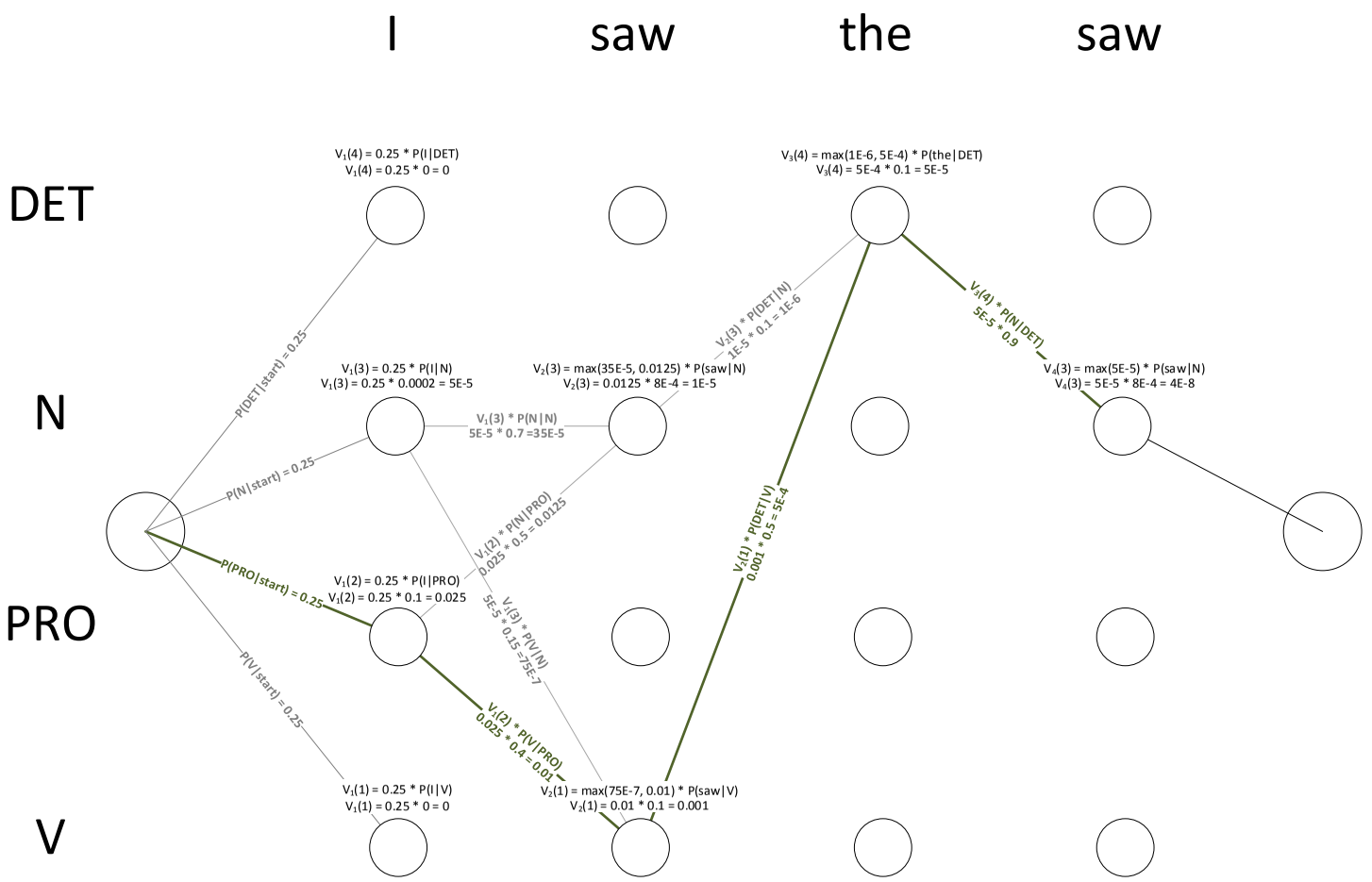
Using the table below, calculate the most likely sequence of tags for the following sentence:
„`I saw the saw`“

1. Apply the Viterbi algorithm.
2. Which are the main factors influencing the complexity of the calculation?

Assume uniformly distributed start transition probabilities.

| from\to | DET | N | PRO | V |
|---|---|---|---|---|
| DET | 0.05 | 0.9 | 0.05 | 0 |
| N | 0.1 | 0.7 | 0.05 | 0.15 |
| PRO | 0.05 | 0.5 | 0.05 | 0.4 |
| V | 0.5 | 0.1 | 0.4 | 0 |

| w\t | DET | N | PRO | V |
|---|---|---|---|---|
| I | 0 | 0.0002 | 0.1 | 0 |
| saw | 0 | 0.0008 | 0 | 0.1 |
| the | 0.1 | 0 | 0 | 0 |


I
saw
the
saw
DET
N
PRO
V
V1(4) = 0.25 * P(I|DET)
V1(4) = 0.25 * 0 = 0
V3(4) = max(1E-6, 5E-4) * P(the|DET)
V3(4) = 5E-4 * 0.1 = 5E-5
P(DET|start) = 0.25
V3(3) * P(DET|N)
1E-5 * 0.1 = 1E-6
V3(4) * P(N|DET)
5E-5 * 0.9
V1(3) = 0.25 * P(I|N)
V1(3) = 0.25 * 0.0002 = 5E-5
V2(3) = max(35E-5, 0.0125) * P(saw|N)
V2(3) = 0.0125 * 8E-4 = 1E-5
V4(3) = max(5E-5) * P(saw|N)
V4(3) = 5E-5 * 8E-4 = 4E-8
V1(3) * P(N|N)
5E-5 * 0.7 =35E-5
P(N|start) = 0.25
V1(2) * P(N|PRO)
0.025 * 0.5 = 0.0125
V3(1) * P(DET|V)
0.001 * 0.5 = 5E-4
P(PRO|start) = 0.25
V1(2) = 0.25 * P(I|PRO)
V1(2) = 0.25 * 0.1 = 0.025
V4(3) * P(V|N)
5E-5 * 0.15 =75E-7
P(V|start) = 0.25
V1(2) * P(V|PRO)
0.025 * 0.4 = 0.01
V1(1) = 0.25 * P(I|V)
V1(1) = 0.25 * 0 = 0
V2(1) = max(75E-7, 0.01) * P(saw|V)
V2(1) = 0.01 * 0.1 = 0.001

*Note: The unreachable states or states with probability 0 are not shown in the graph for readability reasons.*
Main influence factors are the length of the sequence (linear) and the amount of available POS tags (quadratical).