# Der Forward-Algorithmus

Kursfolien

Karin Haenelt



# Zweck

Berechnung der Wahrscheinlichkeit
einer Beobachtung
nach einem Hidden Markov Model

Speichern und Wiederverwenden
partieller Zwischenergebnisse

# Hidden Markov Model

Formal spezifiziert durch Fünf-Tupel $(S, K, \Pi, A, B)$

| | | |
|---|---|---|
| $S = \{s_1, ..., s_N\}$ | | Menge der Zustände |
| $K = \{k_1, ..., k_M\} = \{1, ..., M\}$ | | Ausgabe-Alphabet |
| $\Pi = \{p_i\}, i \in S$ | | Wahrscheinlichkeiten der Startzustände |
| $A = \{a_{ij}\}, \quad i, j \in S$ | $\sum_{j=1}^{N} a_{ij} = 1$ | Wahrscheinlichkeiten der Zustandsübergänge |
| $B = \{b_{ijk}\}, \quad i, j \in S, \quad k \in K$ | $\sum_{k=1}^{M} b_{ijk} = 1$ | Wahrscheinlichkeiten der Symbolemissionen |



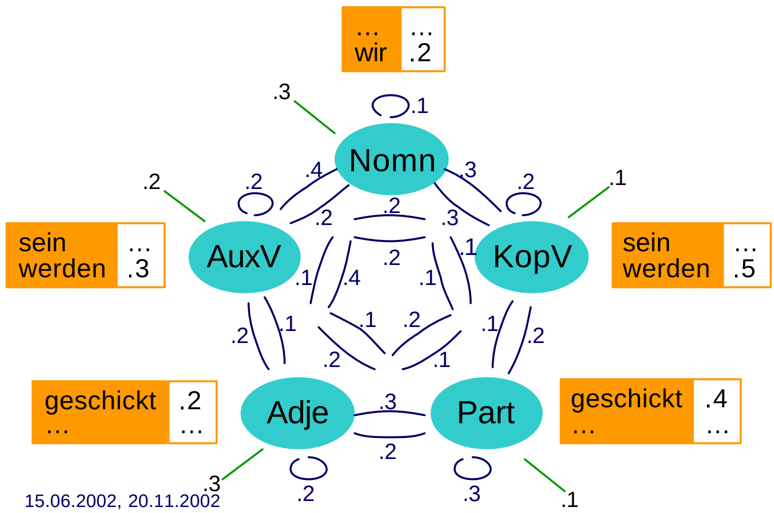
# Notationskonventionen

$X = (X_1, ..., X_{T+1})$ Sequenz der Zustände

$X_t : S \rightarrow \{1, ..., N\}$

$O = (o_1, \ldots, o_T), \quad o_T \in K$ Sequenz der Beobachtungen

# Hidden Markov Model: Graph






# Hidden Markov Model: Matrizen

ein Modell $m = (A, B, \Pi)$

Übergangs-Matrix A | Emissions-Matrix B | Start-Tabelle $\Pi$

| | Adje | AuxV | KopV | Nomn | Part | g‘schickt | werden | wir | $p$ |
|---|---|---|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | .2 | 0 | 0 | .3 |
| AuxV | .2 | .2 | .2 | .2 | .2 | 0 | .3 | 0 | .2 |
| KopV | .2 | .2 | .2 | .3 | .1 | 0 | .5 | 0 | .1 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 | 0 | .2 | .3 |
| Part | .3 | .1 | .2 | .1 | .3 | .4 | 0 | 0 | .1 |

# Hidden Markov Model: Matrizen

Modellvariante:
mit Startsymbol
statt mit Tabelle der Startwahrscheinlichkeiten

| | Adje | AuxV | KopV | Nomn | Part | Ω |
|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | 0 |
| AuxV | .2 | .2 | .2 | .2 | .2 | 0 |
| KopV | .2 | .2 | .2 | .3 | .1 | 0 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 |
| Part | .3 | .1 | .2 | .1 | .3 | 0 |
| Ω | .3 | .2 | .1 | .3 | .1 | 0 |

| g'schickt | werden | wir | . |
|---|---|---|---|
| .2 | 0 | 0 | 0 |
| 0 | .3 | 0 | 0 |
| 0 | .5 | 0 | 0 |
| 0 | 0 | .2 | 0 |
| .4 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 |



# Aufgabe: Wahrscheinlichkeit einer Beobachtung finden

gegeben eine Sequenz von Beobachtungen $O = (o_1, ..., o_T)$

O=(*wir,werden,geschickt*)

ein Modell $m = (A, B, \Pi)$

| | Adje | AuxV | KopV | Nomn | Part | g'schickt | werden | wir | *p* |
|---|---|---|---|---|---|---|---|---|---|
| Adje | .2 | .1 | .1 | .4 | .2 | .2 | 0 | 0 | .3 |
| AuxV | .2 | .2 | .2 | .2 | .2 | 0 | .3 | 0 | .2 |
| KopV | .2 | .2 | .2 | .3 | .1 | 0 | .5 | 0 | .1 |
| Nomn | .1 | .4 | .3 | .1 | .1 | 0 | 0 | .2 | .3 |
| Part | .3 | .1 | .2 | .1 | .3 | .4 | 0 | 0 | .1 |

gesucht die Wahrscheinlichkeit $P(O \mid m)$

$$P(wir, werden, geschickt \mid m)$$

# Forward-Lösung

- Kompakte Darstellung der Pfade als Gitter (Trellis)
- Wiederverwendung partieller Ergebnisse statt wiederholter Neuberechnung




# Spezifikation des Algorithmus

Variante mit arc-emission HMM

1. Initialisierung

$$a_i(1) = p_i, \quad 1 \le i \le N$$

2. Induktion

$$a_j(t+1) = \sum_{i=1}^{N} a_i(t)\, a_{ij}\, b_{ijo_t}, \quad 1 \le i \le N, 1 \le j \le N$$

3. Gesamt

$$P(O \mid m) = \sum_{i=1}^{N} a_i(T+1)$$


Manning/Schütze, 2000: 327

# Spezifikation des Algorithmus

Variante mit state-emission HMM

1. Initialisierung

$$a_i(1) = p_i, \quad 1 \le i \le N$$

2. Induktion

$$a_j(t+1) = \sum_{i=1}^{N} a_i(t)\, a_{ij}\, b_{jo_t}, \quad 1 \le i \le N, 1 \le j \le N$$

3. Gesamt

$$P(O \mid m) = \sum_{i=1}^{N} a_i(T+1)$$


Manning/Schütze, 2000: 327

# Gitter mit Zellen ($s_i$,t)

Variable $a_i(t)$ wird gespeichert in $s_i(t)$

repräsentiert die Wahrscheinlichkeit,
zum Zeitpunkt t im Zustand $s_i$ zu sein
bei gegebenen Beobachtungen $o_1 \ldots o_{t-1}$

# Gitter für das Beispiel

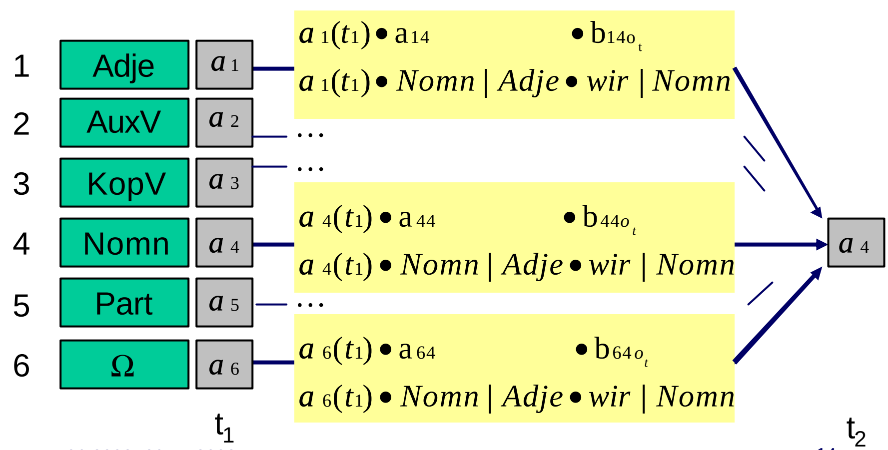
| | | | | | |
|---|---|---|---|---|---|
| 1 | Adje | 0 | 0 | 0 | .000648 |
| 2 | AuxV | 0 | 0 | .0072 | 0 |
| 3 | KopV | 0 | 0 | .0090 | 0 |
| 4 | Nomn | 0 | .06 | 0 | 0 |
| 5 | Part | 0 | 0 | 0 | .000936 |
| 6 | Ω | 1 | 0 | 0 | 0 |
| | | . | wir | werden | geschickt |
| | | $t_1$ | $t_2$ | $t_3$ | $t_4$ |

0.001584



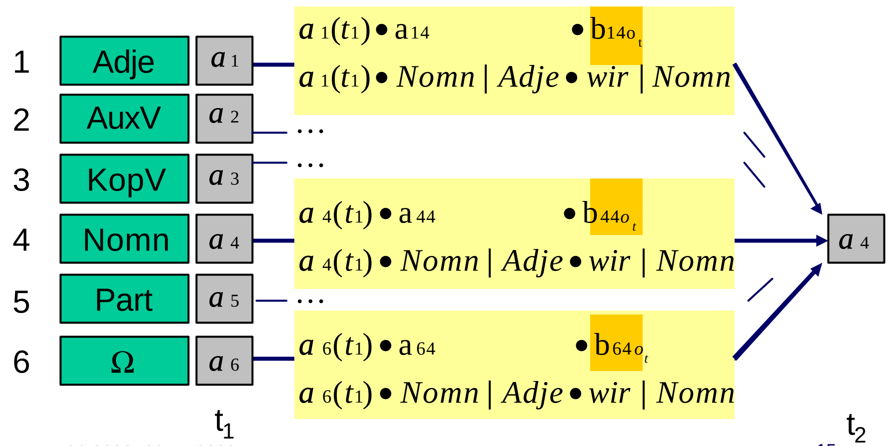
# Berechnung für Zelle $a_4(t_2)$

mit arc-emission HMM



# Berechnung für Zelle $\alpha_4(t_2)$

mit state-emission HMM






# Tracing für ausgewählte Zellen ($t_2$)

$t_2$

| $\alpha_j(t+1)=\sum_{i=1}^{N}$ = | | | | $\alpha_i(t)$ | | x $a_{ij}$ | | x $b_{j\,ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 1 | Adje | 1 | Adje | $a_1(1)$ | 0 | Adje\|Adje | .2 | wir\|Adje | 0 | 0 |
| | | 2 | AuxV | $a_2(1)$ | 0 | Adje\|AuxV | .2 | wir\|Adje | 0 | 0 |
| | | 3 | KopV | $a_3(1)$ | 0 | Adje\|KopV | .2 | wir\|Adje | 0 | 0 |
| | | 4 | Nomn | $a_4(1)$ | 0 | Adje\|Nomn | .1 | wir\|Adje | 0 | 0 |
| | | 5 | Part | $a_5(1)$ | 0 | Adje\|Part | .3 | wir\|Adje | 0 | 0 |
| | | 6 | Ω | $a_6(1)$ | 1 | Adje\| Ω | .3 | wir\|Adje | 0 | 0 |
| **$a_{12}$ =** | | | | | | | | | | **0** |

...

| $\alpha_j(t+1)=\sum_{i=1}^{N}$ = | | | | $\alpha_i(t)$ | | x $a_{ij}$ | | x $b_{ij\,ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 4 | Nomn | 1 | Adje | $a_1(1)$ | 0 | Nomn\|Adje | .4 | wir\|Nomn | .2 | 0 |
| | | 2 | AuxV | $a_2(1)$ | 0 | Nomn\|AuxV | .2 | wir\|Nomn | .2 | 0 |
| | | 3 | KopV | $a_3(1)$ | 0 | Nomn\|KopV | .3 | wir\|Nomn | .2 | 0 |
| | | 4 | Nomn | $a_4(1)$ | 0 | Nomn\|Nomn | .1 | wir\|Nomn | .2 | 0 |
| | | 5 | Part | $a_5(1)$ | 0 | Nomn\|Part | .1 | wir\|Nomn | .2 | 0 |
| | | 6 | Ω | $a_6(1)$ | 1 | Nomn\| Ω | .3 | wir\|Nomn | .2 | .06 |
| **$a_{42}$ =** | | | | | | | | | | **.06** |

# Tracing für ausgewählte Zellen ($t_3$)

| $a_j(t+1)=\sum_{i=1}^{N}$ = | | | | $a_i(t)$ | | x $a_{ij}$ | | x $b_{ij\ ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 1 | Adje | 1 | Adje | $a_1(2)$ | 0 | Adje\|Adje | .2 | werden\|Adje | 0 | 0 |
| | | 2 | AuxV | $a_2(2)$ | 0 | Adje\|AuxV | .2 | werden\|Adje | 0 | 0 |
| | | 3 | KopV | $a_3(2)$ | 0 | Adje\|KopV | .2 | werden\|Adje | 0 | 0 |
| | | 4 | Nomn | $a_4(2)$ | .06 | Adje\|Nomn | .1 | werden\|Adje | 0 | 0 |
| | | 5 | Part | $a_5(2)$ | 0 | Adje\|Part | .3 | werden\|Adje | 0 | 0 |
| | | 6 | Ω | $a_6(2)$ | 0 | Adje\| Ω | .3 | werden\|Adje | 0 | 0 |
| **$a_{13}$** = | | | | | | | | | | **0** |

| $a_j(t+1)=\sum_{i=1}^{N}$ = | | | | $a_i(t)$ | | x $a_{ij}$ | | x $b_{ij\ ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 2 | AuxV | 1 | Adje | $a_1(2)$ | 0 | AuxV\|Adje | .1 | werden\|AuxV | .3 | 0 |
| | | 2 | AuxV | $a_2(2)$ | 0 | AuxV\|AuxV | .2 | werden\|AuxV | .3 | 0 |
| | | 3 | KopV | $a_3(2)$ | 0 | AuxV\|KopV | .2 | werden\|AuxV | .3 | 0 |
| | | 4 | Nomn | $a_4(2)$ | .06 | AuxV\|Nomn | .4 | werden\|AuxV | .3 | .0072 |
| | | 5 | Part | $a_5(2)$ | 0 | AuxV\|Part | .1 | werden\|AuxV | .3 | 0 |
| | | 6 | Ω | $a_6(2)$ | 0 | AuxV\| Ω | .2 | werden\|AuxV | .3 | 0 |
| **$a_{23}$** = | | | | | | | | | | **.0072** |



# Tracing für ausgewählte Zellen ($t_3$)

| $a_j(t+1)=\sum_{i=1}^{N}$ = | | | | $a_i(t)$ | | x $a_j$ | | x $b_{ij\ ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 3 | KopV | 1 | Adje | $a_1(2)$ | 0 | KopV\|Adje | .1 | werden\|KopV | .5 | 0 |
| | | 2 | AuxV | $a_2(2)$ | 0 | KopV\|AuxV | .2 | werden\|KopV | .5 | 0 |
| | | 3 | KopV | $a_3(2)$ | 0 | KopV\|KopV | .2 | werden\|KopV | .5 | 0 |
| | | 4 | Nomn | $a_4(2)$ | .06 | KopV\|Nomn | .3 | werden\|KopV | .5 | .009 |
| | | 5 | Part | $a_5(2)$ | 0 | KopV\|Part | .2 | werden\|KopV | .5 | 0 |
| | | 6 | Ω | $a_6(2)$ | 0 | KopV\| Ω | .1 | werden\|KopV | .5 | 0 |
| **$a_{33}$** = | | | | | | | | | | **.009** |

# Tracing für ausgewählte Zellen ($t_4$)

| $a_j(t+1) = \sum_{i=1}^{N}$ = | | | | $a_i(t)$ | | x $a_j$ | | x $b_{ij\ ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 1 | Adje | 1 | Adje | $a_1(3)$ | 0 | Adje\|Adje | .2 | geschickt\|Adje | .2 | 0 |
| | | 2 | AuxV | $a_2(3)$ | .0072 | Adje\|AuxV | .2 | geschickt\|Adje | .2 | .000288 |
| | | 3 | KopV | $a_3(3)$ | .0090 | Adje\|KopV | .2 | geschickt\|Adje | .2 | .000360 |
| | | 4 | Nomn | $a_4(3)$ | 0 | Adje\|Nomn | .1 | geschickt\|Adje | .2 | 0 |
| | | 5 | Part | $a_5(3)$ | 0 | Adje\|Part | .3 | geschickt\|Adje | .2 | 0 |
| | | 6 | Ω | $a_6(3)$ | 0 | Adje\| Ω | .3 | geschickt\|Adje | .2 | 0 |
| **$a_{14}$ =** | | | | | | | | | | **.000648** |

…

| $a_j(t+1) = \sum_{i=1}^{N}$ = | | | | $a_i(t)$ | | x $a_{ij}$ | | x $b_{ij\ ot}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | vorige | | aktuelle Beobachtung | | | | |
| j | | i | | | | Übergang | | Emission | | |
| 5 | Part | 1 | Adje | $a_1(3)$ | 0 | Part\|Adje | .2 | geschickt\|Part | .4 | 0 |
| | | 2 | AuxV | $a_2(3)$ | .0072 | Part\|AuxV | .2 | geschickt\|Part | .4 | .000576 |
| | | 3 | KopV | $a_3(3)$ | .0090 | Part\|KopV | .1 | geschickt\|Part | .4 | .000360 |
| | | 4 | Nomn | $a_4(3)$ | 0 | Part\|Nomn | .1 | geschickt\|Part | .4 | 0 |
| | | 5 | Part | $a_5(3)$ | 0 | Part\|Part | .3 | geschickt\|Part | .4 | 0 |
| | | 6 | Ω | $a_6(3)$ | 0 | Part\| Ω | .1 | geschickt\|Part | .4 | 0 |
| **$a_{54}$ =** | | | | | | | | | | **.000936** |



# Gesamtsumme

$$P(O \mid m) = \sum_{i=1}^{N} a_i(T+1)$$

P(*wir werden geschickt* | µ) = 0.001584

# Berechnungsaufwand

$2\,N^2T$ Multiplikationen


Manning/Schütze, 2000: 327

# Literatur


- Manning, Christopher D.; Schütze, Hinrich (1999): *Foundations of Statistical Natural Language Processing.* Cambridge, Mass., London: The MIT Press. (vgl.: http://www.sultry.arts.usyd.edu.au/fsnlp)