Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Sprachverarbeitung I / 13 HS 2016

Spracherkennung: Statistischer Ansatz

Buch: Kapitel 13.1 bis 13.5

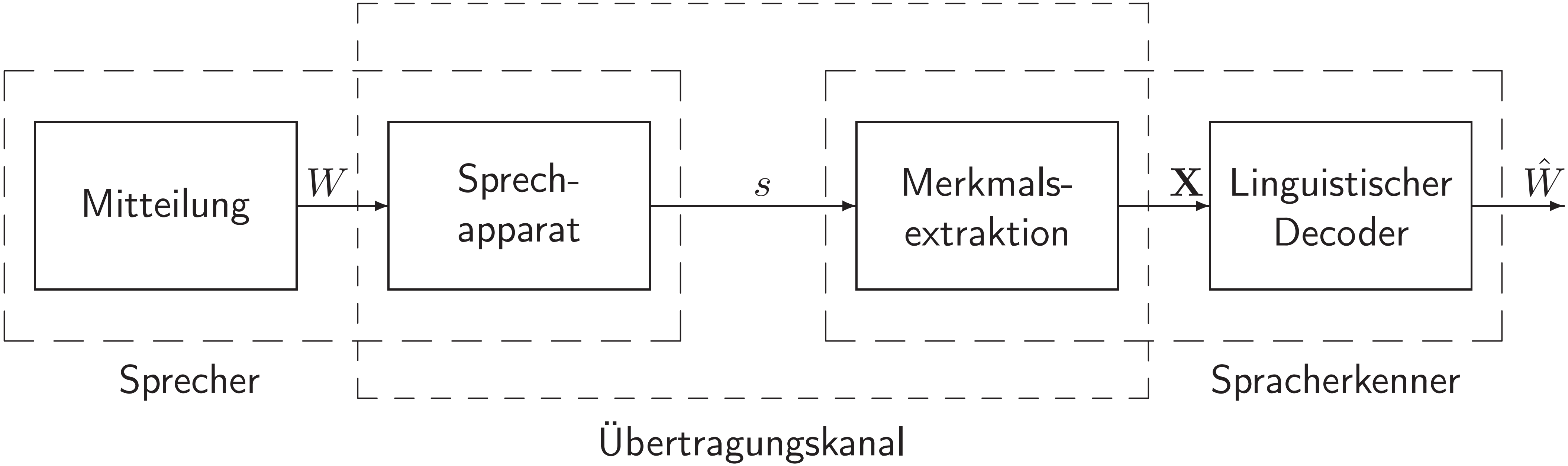
Beat Pfister

Programm heute:

Vorlesung:
- Statistischer Ansatz der Spracherkennung
- Hidden-Markov-Modelle (HMM)
- Verschiedene Spracherkenner mit HMM

Übung:
- ⋆ Spracherkennung mit DTW

Informationstheoretische Sicht der Spracherkennung



Decodierungsproblem

Gegeben: Merkmalssequenz $\mathbf{X} = \mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_T$

Gesucht: Optimale Schätzung $\widehat{W}$ für die geäusserte Wortfolge

Ziel (formuliert aus statistischer Sicht)

Bestimmen der optimalen Wortfolge $\widehat{W} = w_1 w_2 \ldots w_K$,
sodass Wahrscheinlichkeit eines Fehlentscheides minimal

Maximum-a-posteriori Regel (MAP-Regel)

$\boxed{\text{Wahrscheinlichkeit für Fehlentscheid ist minimal, wenn die Wortfolge mit der höchsten A-posteriori-Wahrscheinlichkeit } P(W|\mathbf{X}) \text{ gewählt wird}}$

$$\longrightarrow \quad \hat{W} = \underset{W}{\operatorname{argmax}}\ P(W|\mathbf{X})$$

Problem: $P(W|\mathbf{X})$ ist praktisch nicht ermittelbar!

Grund: Anzahl der verschiedenen $\mathbf{X}$ ist viel zu gross!

MAP-Regel

Original: $\hat{W} = \underset{W}{\text{argmax}}\ P(W|\mathbf{X})$

mit: $P(A,B) = P(A|B)\,P(B) = P(B|A)\,P(A)$

$$\Rightarrow P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

äquivalent: $\hat{W} = \underset{W}{\text{argmax}}\ \frac{P(\mathbf{X}|W)\,P(W)}{P(\mathbf{X})} = \underset{W}{\text{argmax}}\ P(\mathbf{X}|W)\,P(W)$

$P(\mathbf{X}|W)$: akustisches Modell

$P(W)$: A-priori-Wissen (Sprachmodell)

Sprachmodell $P(W)$

- Gibt Auskunft über die Wahrscheinlichkeit (relative Häufigkeit) von W

- Hilft insbes. bei akustisch nicht oder schlecht unterscheidbaren Wortfolgen:

$$P(\text{“Gestern fiel viel Schnee.”}) \gg P(\text{“Gestern viel fiel Schnee.”})$$

(Behandlung in SPV II)

Akustisches Modell $P(\mathbf{X}|W)$

Wahrscheinlichkeit der Merkmalssequenz $\mathbf{X}$
gegeben die Wortfolge W

Notationen:	W	Wortfolge $w_1 w_2 \ldots w_K$
	w_k	das k-te Wort einer Wortfolge W
	V	Vokabular des Erkenners
	v_i	das i-te Wort des Vokabulars V

Spezialfall: W ist nur 1 Wort lang $\longrightarrow P(\mathbf{X}|v_i)$

Akustisches Modell $P(\mathbf{X}|v_i)$

Wahrscheinlichkeit der Merkmalssequenz $\mathbf{X}$, gegeben das Wort v_i

$\longrightarrow P(\mathbf{X}|v_i)$ ist eine statistische Beschreibung der Merkmalssequenz von v_i

Merke: Variabilität des Sprachsignals schlägt sich in der Merkmalssequenz $\mathbf{X} = \mathbf{x}_1\,\mathbf{x}_2 \ldots \mathbf{x}_t \ldots \mathbf{x}_T$ nieder.

Sowohl die Länge T als auch die $\mathbf{x}_t$ variieren!

Frage: Wie lassen sich diese beiden Arten von Variabilität beschreiben?

Variabilität des Merkmals $\mathbf{x}_t$

Annahmen: – alle Merkmalssequenzen von v_i seien gleich lang

– die Merkmale seien diskret (z.B. aus Vektorquantisierung):

$\mathbf{x}_t \in \{1, 2, 3, \ldots, M\}$ für $1 \leq t \leq T$

Folge: Für den Zeitpunkt t kann $\mathbf{x}_t$ als diskrete Zufallsvariable betrachtet werden und ist somit beschreibbar als diskrete Wahrscheinlichkeitsverteilung $P(\mathbf{x}_t)$

>>>

Variabilität der Länge T und der Merkmale $\mathbf{x}_t$

Ansatz zur statistischen Beschreibung: Hidden-Markov-Modell (HMM)

Zwei gekoppelte Zufallsprozesse:

a) Markov-Prozess mit N (verdeckten) Zuständen $S_1, S_2, \ldots, S_N$

spezifiziert durch Zustandsübergangswahrscheinlichkeiten a_{ij} >>>

b) zustandsabhängiger Zufallsprozess, der zu jedem diskreten Zeitpunkt eine Beobachtung $\mathbf{x}_t$ erzeugt

spezifiziert durch Beobachtungswahrscheinlichkeiten $b_j(\mathbf{x})$ >>>

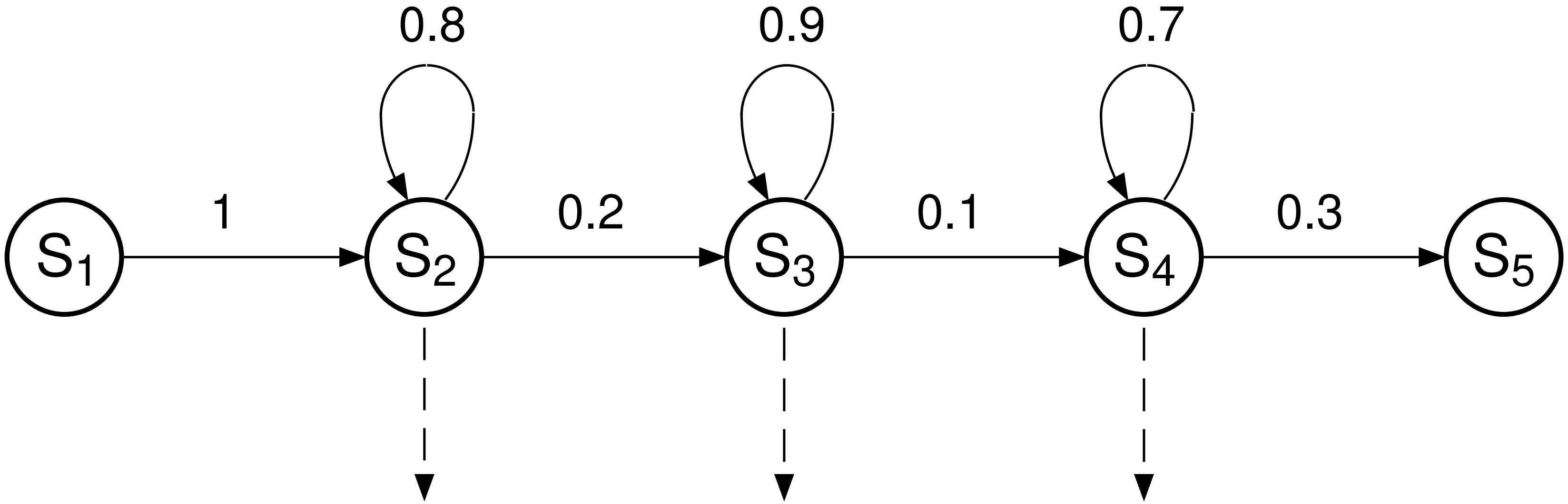
Spezifikation eines HMM

Ein HMM mit N Zuständen und M diskreten Beobachtungen wird vollständig beschrieben durch:

- Zustandsübergangswahrscheinlichkeitsmatrix $A = \{a_{ij}\}$
  mit $1 \leq i, j \leq N \quad \longrightarrow \quad N \times N$-Matrix
- Beobachtungswahrscheinlichkeitsverteilung $B = \{b_j(k)\}$
  mit $1 < j < N$ und $1 \leq k \leq M \quad \longrightarrow \quad (N-2) \times M$-Matrix

Kurzform: $\lambda = (A, B)$

Beispiel: HMM mit $N = 5$ Zuständen und $M = 4$ diskreten Beobachtungen



x	$P(x\|S_2)$
1	0.04
2	0.12
3	0.08
4	0.76

x	$P(x\|S_3)$
1	0.03
2	0.91
3	0.04
4	0.02

x	$P(x\|S_4)$
1	0.68
2	0.11
3	0.05
4	0.16

HMM als Generator von Zufallssequenzen

Beim Generieren einer Beobachtungssequenz $\mathbf{X}$ durchläuft das HMM die Zustandsfolge $Q = S_1 q_1 \ldots q_T S_N$

Diskrete Zeit t:	0	1	2	3	4	5	6	7	8
Zustand q_t:	S_1								
Beobachtung $\mathbf{x}_t$:	–								

HMM als Generator von Zufallssequenzen

Beim Generieren einer Beobachtungssequenz $\mathbf{X}$ durchläuft das HMM die Zustandsfolge $Q = S_1 q_1 \ldots q_T S_N$

Diskrete Zeit t:	0	1	2	3	4	5	6	7	8
Zustand q_t:	S_1	S_2							
Beobachtung $\mathbf{x}_t$:	–								

>>>

HMM als Generator von Zufallssequenzen

Beim Generieren einer Beobachtungssequenz $\mathbf{X}$ durchläuft das HMM die Zustandsfolge $Q = S_1 q_1 \ldots q_T S_N$

Diskrete Zeit t:	0	1	2	3	4	5	6	7	8
Zustand q_t:	S_1	S_2							
Beobachtung $\mathbf{x}_t$:	–	3							

HMM als Generator von Zufallssequenzen

Beim Generieren einer Beobachtungssequenz $\mathbf{X}$ durchläuft das HMM die Zustandsfolge $Q = S_1 q_1 \ldots q_T S_N$

Diskrete Zeit t:	0	1	2	3	4	5	6	7	8
Zustand q_t:	S_1	S_2	S_2						
Beobachtung $\mathbf{x}_t$:	–	3							

HMM als Generator von Zufallssequenzen

Beim Generieren einer Beobachtungssequenz $\mathbf{X}$ durchläuft das HMM die Zustandsfolge $Q = S_1 q_1 \dots q_T S_N$

Diskrete Zeit t:	0	1	2	3	4	5	6	7	8
Zustand q_t:	S_1	S_2	S_2	S_3	S_4	S_4	S_5		
Beobachtung $\mathbf{x}_t$:	–	3	1	1	4	2	–		

HMM-Rechenbeispiele

1. Mit welcher Wahrscheinlichkeit durchläuft das HMM die Zustandssequenz $Q = S_1\, S_2\, S_2\, S_3\, S_4\, S_4\, S_5$?

$$\begin{aligned} P(Q) &= P(q_1{=}S_2 \,|\, q_0{=}S_1)\, P(q_2{=}S_2 \,|\, q_1{=}S_2)\, \ldots\, P(q_6{=}S_5 \,|\, q_5{=}S_4) \\ &= a_{12}\, a_{22}\, a_{23}\, a_{34}\, a_{44}\, a_{45} = 1 \cdot 0.8 \cdot 0.2 \cdot 0.1 \cdot 0.7 \cdot 0.3 = 0.00336 \end{aligned}$$

2. Mit welcher Wahrscheinlichkeit erzeugt das HMM die Beobachtungssequenz $\mathbf{X} = 2\,4\,2\,1\,4$, wenn es diese Zustandssequenz Q durchläuft?

$$\begin{aligned} P(\mathbf{X}|Q) &= P(x{=}2|S_2)\, P(x{=}4|S_2)\, P(x{=}2|S_3)\, P(x{=}1|S_4)\, P(x{=}4|S_4) \\ &= b_2(2)\; b_2(4)\; b_3(2)\; b_4(1)\; b_4(4) \\ &= 0.12 \cdot 0.76 \cdot 0.91 \cdot 0.68 \cdot 0.16 \approx 0.00903 \end{aligned}$$

HMM-Rechenbeispiele (Fortsetzung)

3. Mit welcher Wahrscheinlichkeit durchläuft das HMM die Zustandssequenz $Q = S_1\,S_2\,S_2\,S_3\,S_4\,S_4\,S_5$? **und** erzeugt dabei die Beobachtungssequenz $\mathbf{X} = 2\,4\,2\,1\,4$?

   $P(Q, \mathbf{X}) = P(Q)\,P(\mathbf{X}|Q)$

4. Mit welcher Wahrscheinlichkeit erzeugt das HMM die Beobachtungssequenz $\mathbf{X} = 2\,4\,2\,1\,4$?

   Produktionswahrscheinlichkeit $P(\mathbf{X}|\lambda) =?$

   $\longrightarrow$ Aufsummieren der $P(Q, \mathbf{X})$ über alle möglichen Zustandssequenzen

Die grundlegenden HMM-Probleme

1. Evaluationsproblem:

Gegeben:	HMM λ, Beobachtungssequenz $\mathbf{X} = \mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_T$
Gesucht:	Produktionswahrscheinlichkeit $P(\mathbf{X}\vert\lambda)$
Lösung:	Forward-Algorithmus

2. Decodierungsproblem:

Gegeben:	HMM λ, Beobachtungssequenz $\mathbf{X} = \mathbf{x}_1\mathbf{x}_2\ldots\mathbf{x}_T$
Gesucht:	optimale Zustandssequenz $\widehat{Q} = S_1\widehat{q}_1\ldots\widehat{q}_T S_N$
Lösung:	Viterbi-Algorithmus

3. Schätzproblem:

Gegeben:	Satz von Beobachtungssequenzen $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_S\}$
Gesucht:	HMM $\lambda = (A, B)$, mit $\rightarrow P(\mathcal{X}\vert\lambda) = \text{maximal}$
Lösung:	Baum-Welch-Algorithmus

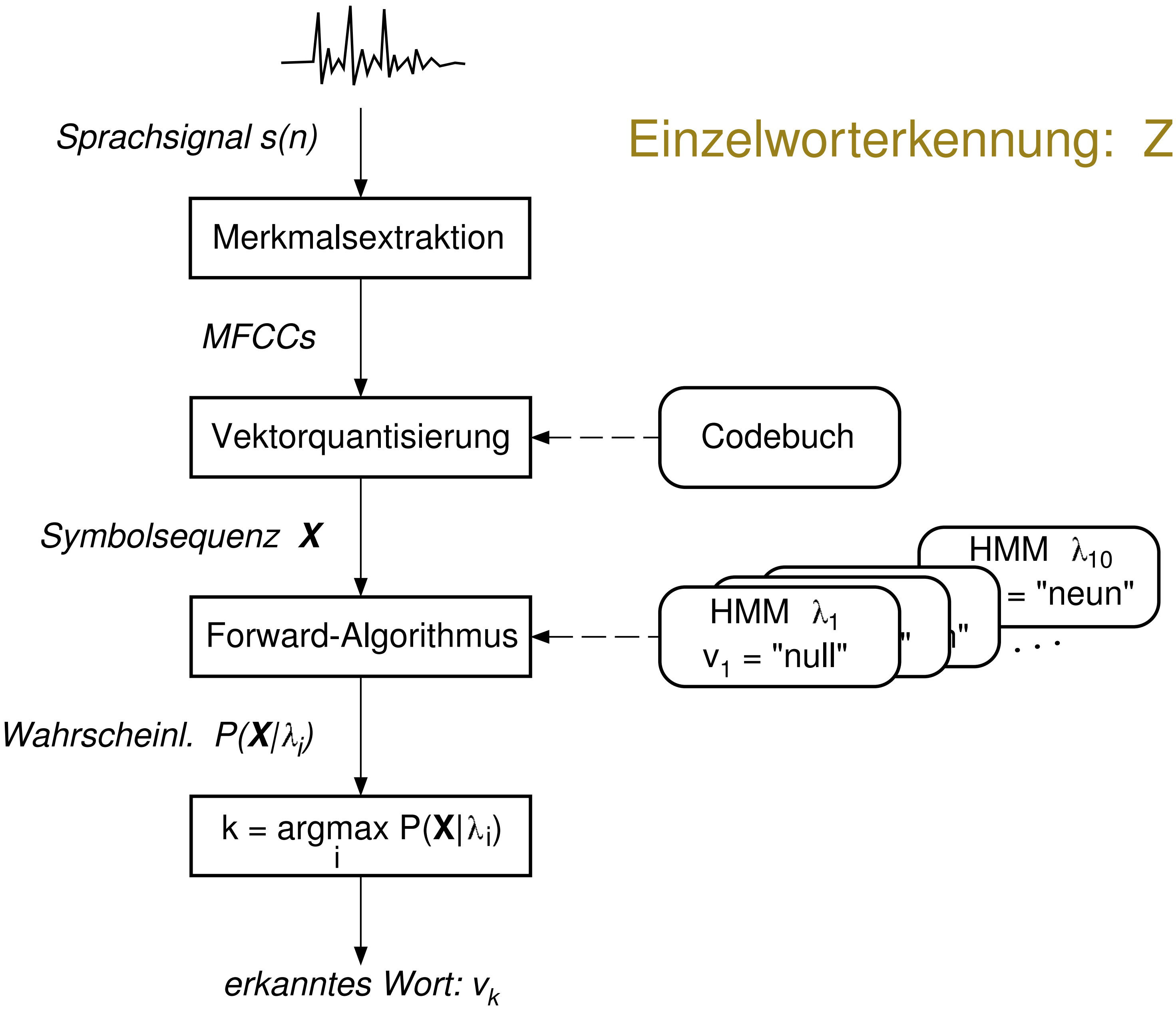
Wortmodelle für einen Spracherkenner

Ziel: Wort-HMM λ_i für jedes Wort des Erkennervokabulars $V = \{v_1, v_2, \ldots, v_{|V|}\}$

Vorgehen für jedes Wort v_i:

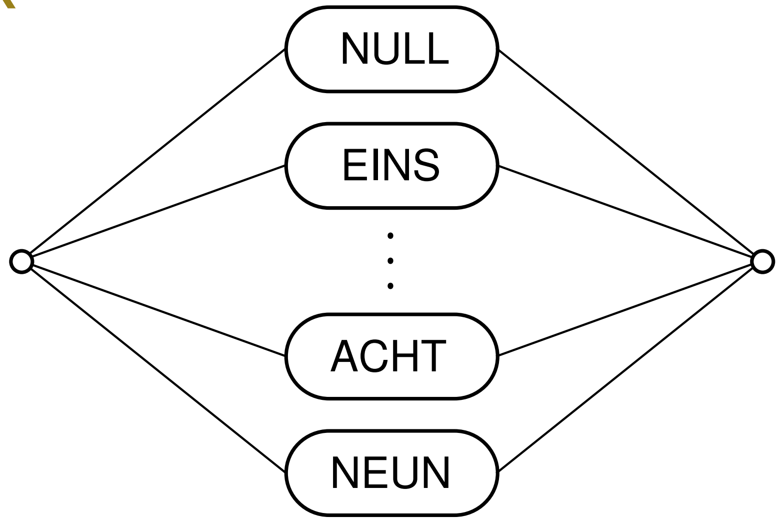
- Sprachsignale des Wortes v_i von vielen Personen aufnehmen
- Merkmalsextraktion und Vektorquantisierung $\rightarrow \mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_S\}$
- HMM-Training mit $\mathcal{X} \longrightarrow \lambda_i$
  (Baum-Welch-Algorithmus)

Einzelworterkennung: Ziffern



Erkennungsnetzwerk

Spracherkenner für die
Ziffern "NULL" . . . "NEUN"

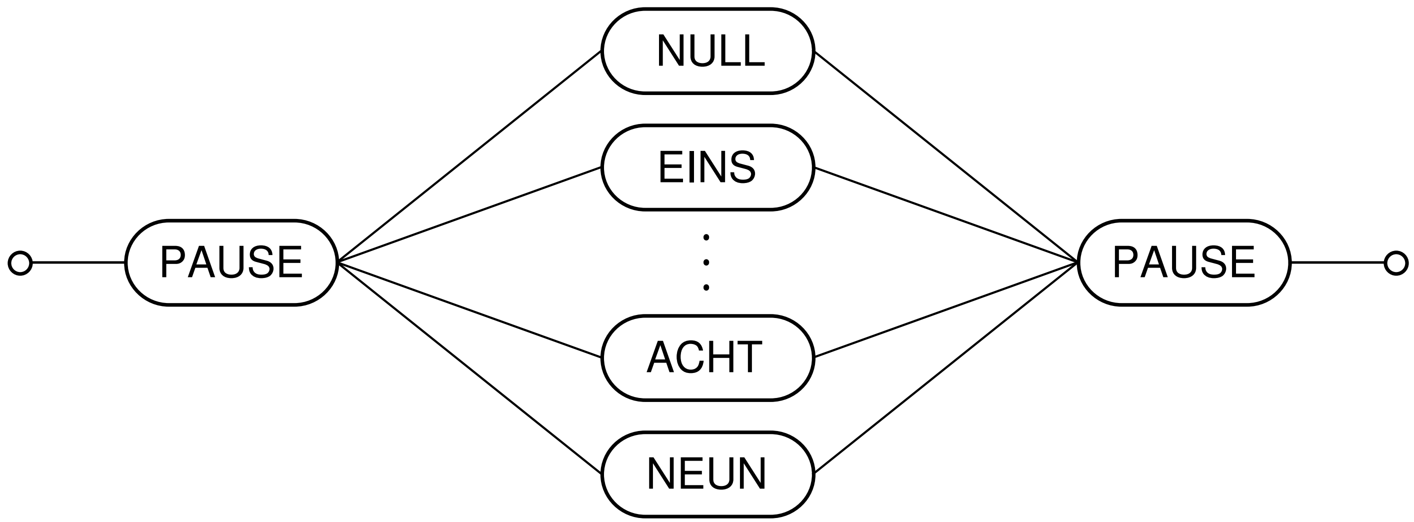


Parallelschaltung mehrerer HMM ergibt wiederum ein HMM >>>

Anwendung des Viterbi-Algorithmus: $\longrightarrow$ optimale Zustandssequenz
$\longrightarrow$ erkanntes Wort

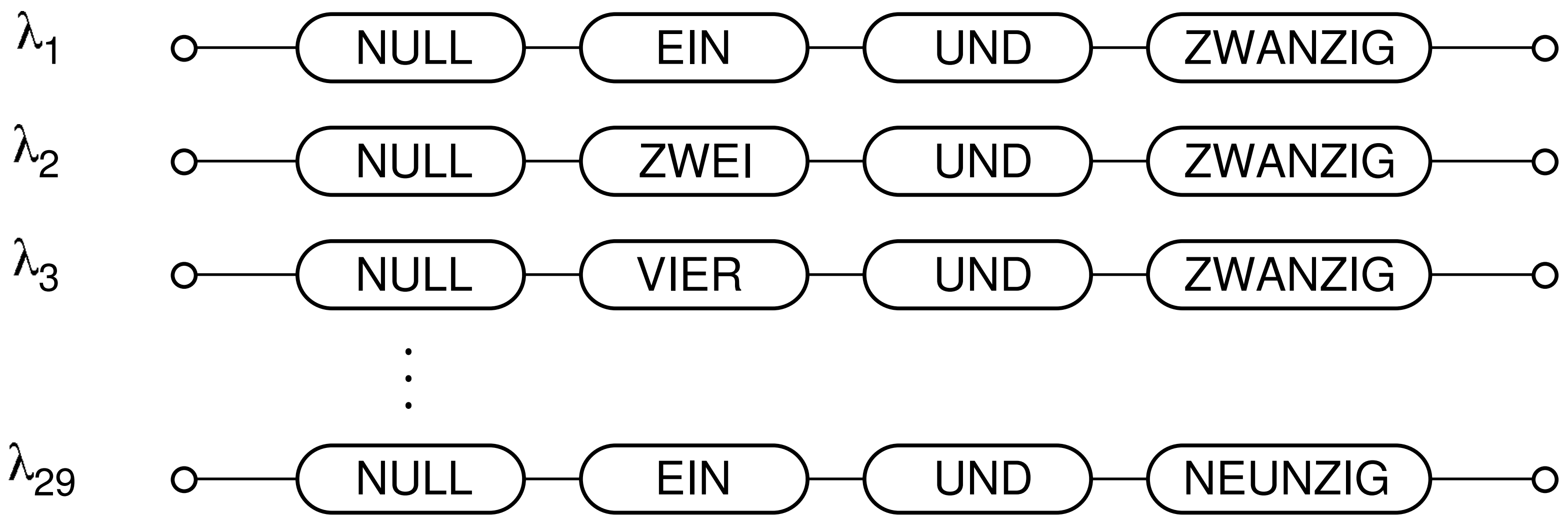
Anfangs- und Endpunktdetektion

Detektieren und Erkennen des Wortes finden gleichzeitig statt!



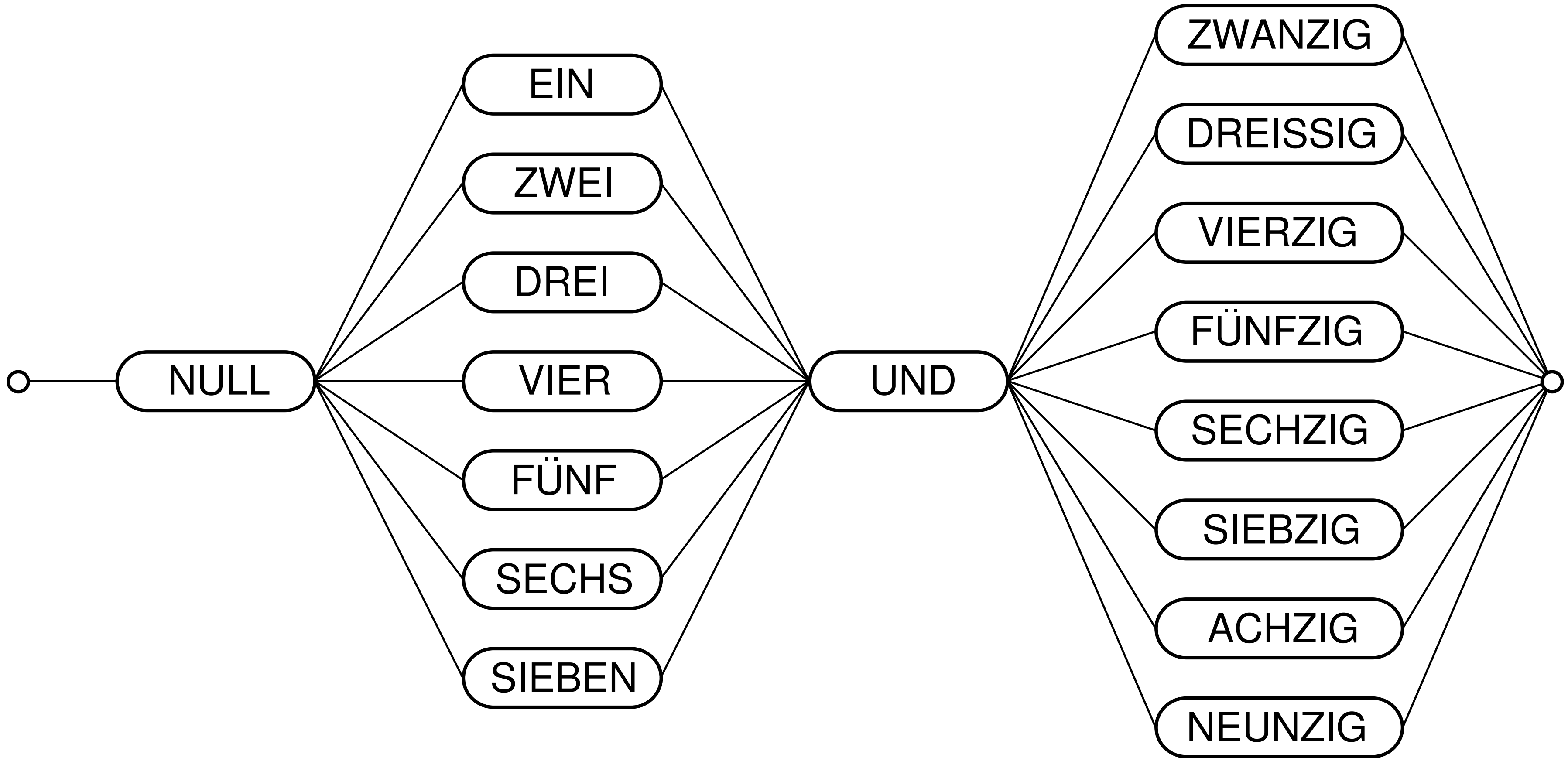
Voraussetzung: HMM für Pause vorhanden
(Pause = leise Hintergrundgeräusche)

Verbundworterkenner

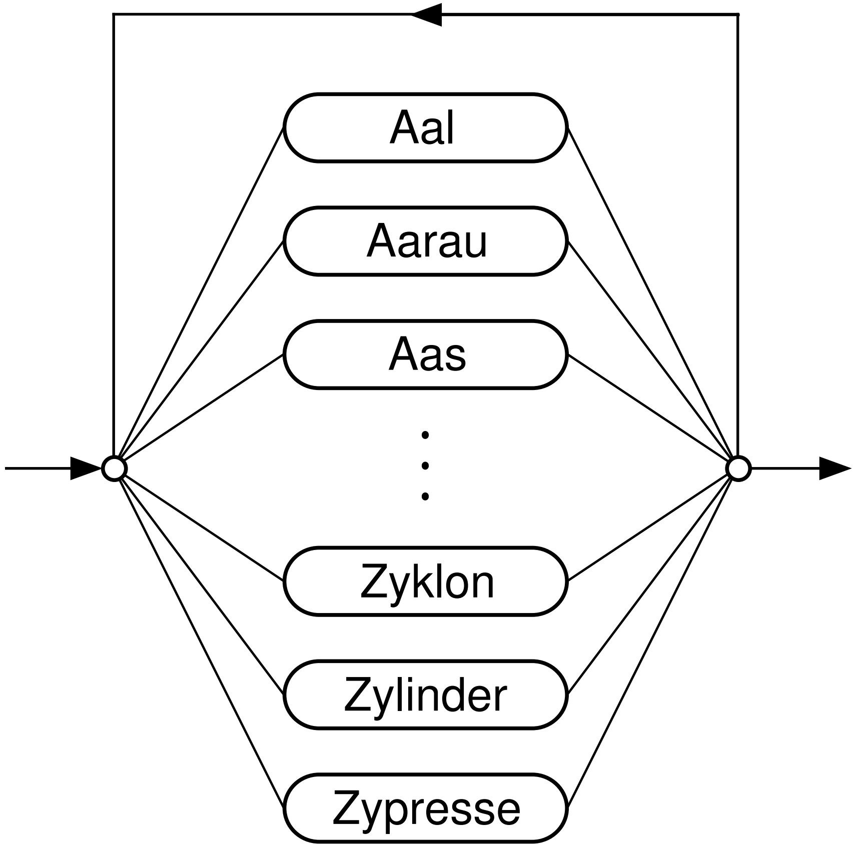


Wörter werden gleichzeitig detektiert und erkannt!

Verbundworterkenner (effizienter)



Erkennung kontinuierlich gesprochener Sprache



Ausblick auf Sprachverarbeitung II

Sprachsynthese: Transkription (5)

- ⋆ Grundlagen: formale Sprachen/Automaten
- ⋆ morphologische/syntaktische Analyse

Spracherkennung: statistischer Ansatz (5)

- ⋆ Grundlagen der Hidden-Markov-Modelle
- ⋆ Modellierung (Training/Einsatz von HMMs)

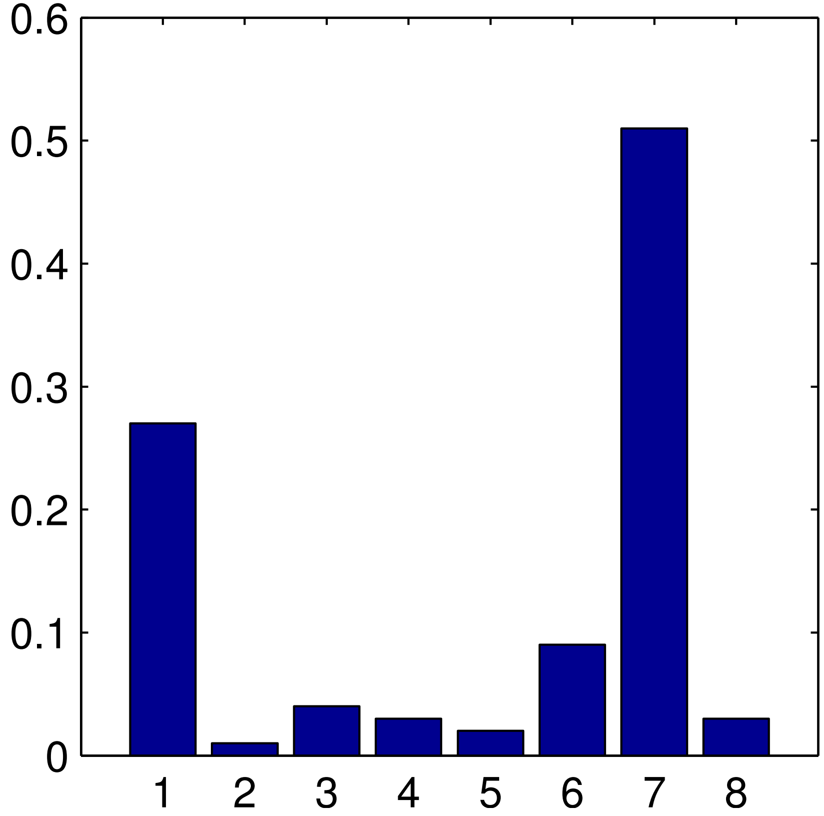
Sprachmodellierung: (*language modelling*) (2)

- ⋆ statistischer Ansatz
- ⋆ wissensbasierter Ansatz

Zur Übersicht der Vorlesung *Sprachverarbeitung I* >>>

Diskrete Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsverteilung des diskreten Merkmals $\mathbf{x}_t$ zum Zeitpunkt t:



$\mathbf{x}$	$P(\mathbf{x})$
1	0.27
2	0.01
3	0.04
4	0.03
5	0.02
6	0.09
7	0.51
8	0.03

<<<

Markov-Prozess (zeitdiskrete Markov-Kette)

Annahme für Markov-Prozess 1. Ordnung:

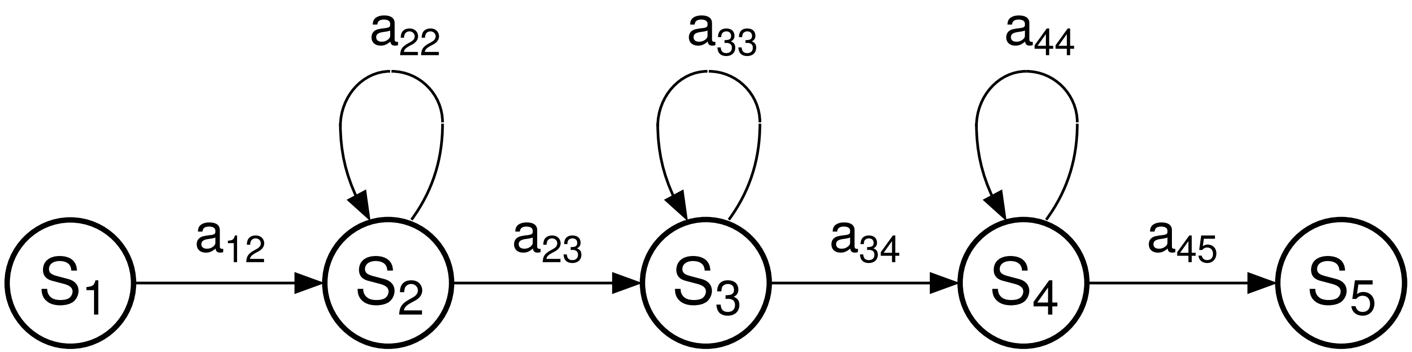
$$P(q_{t+1}{=}S_j \,|\, q_t{=}S_i,\, q_{t-1}{=}S_k,\, \ldots) \;=\; P(q_{t+1}{=}S_j \,|\, q_t{=}S_i)$$

Zustandsübergangswahrscheinlichkeiten a_{ij}

$$a_{ij} \;=\; P(q_{t+1}{=}S_j \,|\, q_t{=}S_i)\;, \qquad\qquad 1 \leq i,j \leq N$$

Lineares Markov-Modell mit $N = 5$ Zuständen

Zustandsdiagramm:

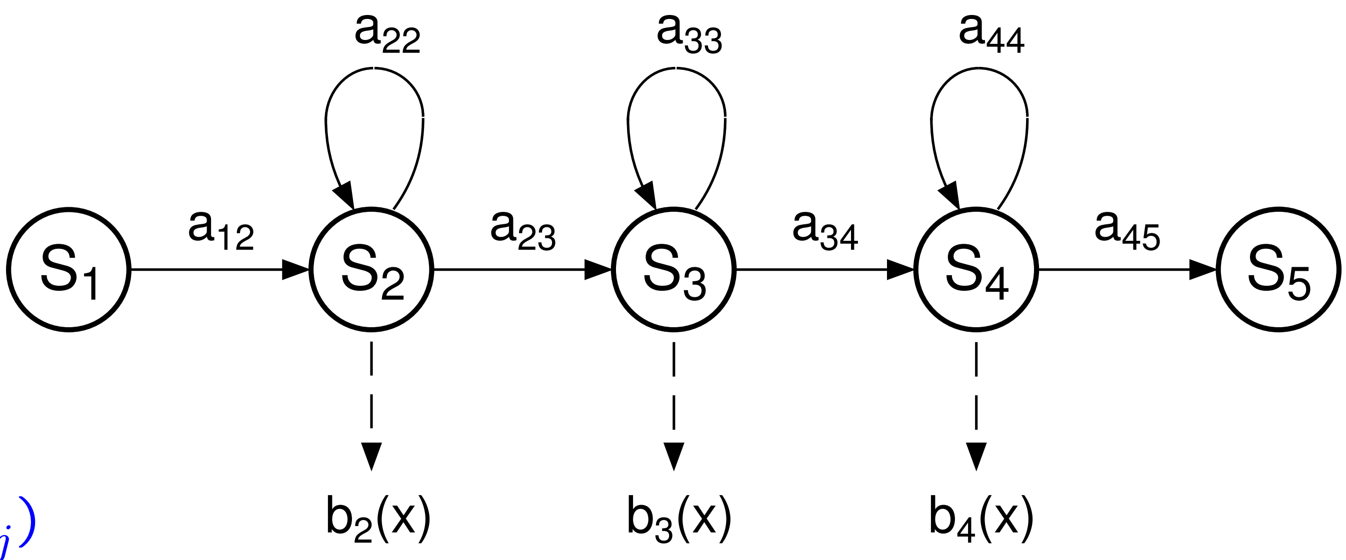


Zustandsübergangswahrscheinlichkeitsmatrix: $N \times N$-Matrix

$$A = \{a_{ij}\} = \begin{bmatrix} 0 & a_{12} & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

<<<

Lineares HMM mit $N = 5$ Zuständen und $M = 4$ Beobachtungen



$P(\mathbf{x}|S_j)$

Beobachtungswahrscheinlichkeitsverteilungen: $(N-2) \times M$-Matrix

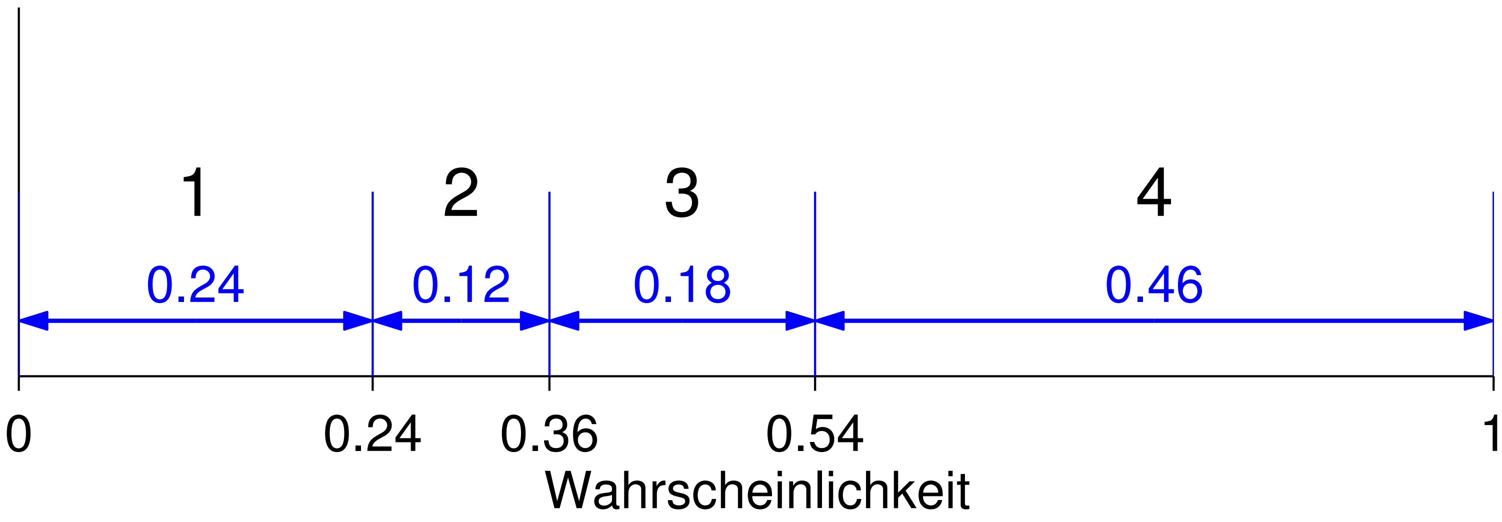
$$B = \{b_j(k)\} = \begin{bmatrix} b_2(1) & b_2(2) & b_2(3) & b_2(4) \\ b_3(1) & b_3(2) & b_3(3) & b_3(4) \\ b_4(1) & b_4(2) & b_4(3) & b_4(4) \end{bmatrix}$$

<<<

Generieren diskreter Zufallswerte

diskr. Verteilung:

x	$P(x)$
1	0.24
2	0.12
3	0.18
4	0.46

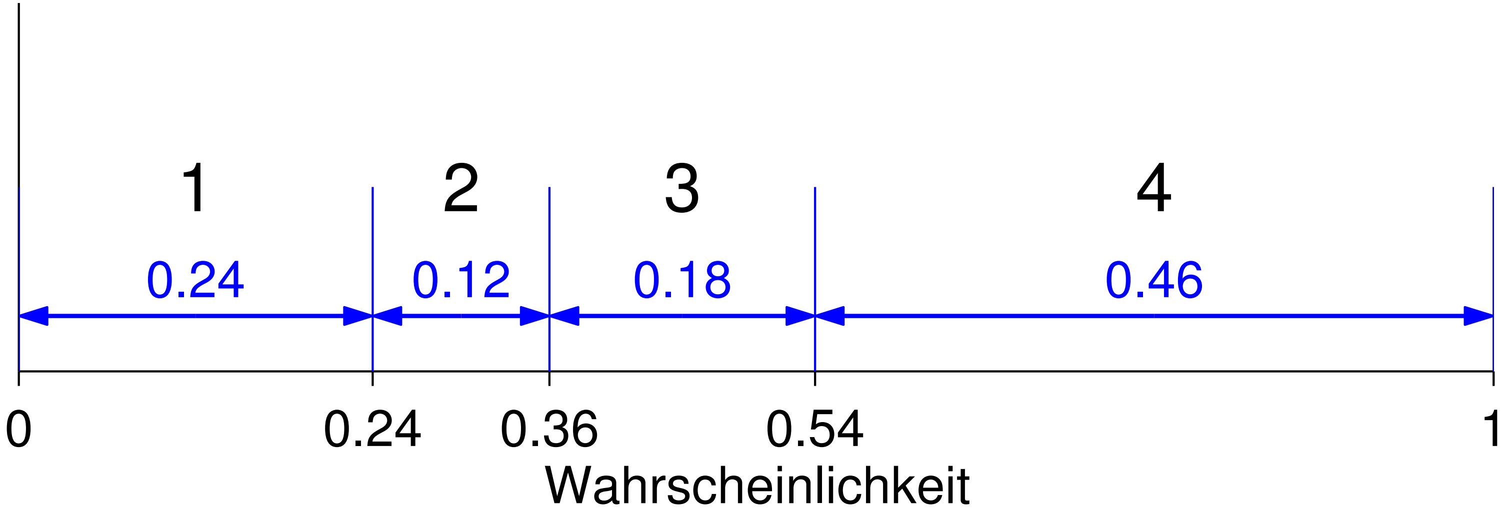


Generieren diskreter Zufallswerte

diskr. Verteilung:

x	$P(\mathrm{x})$
1	0.24
2	0.12
3	0.18
4	0.46

Zufallszahlengenerator $[0 \ldots 1]$

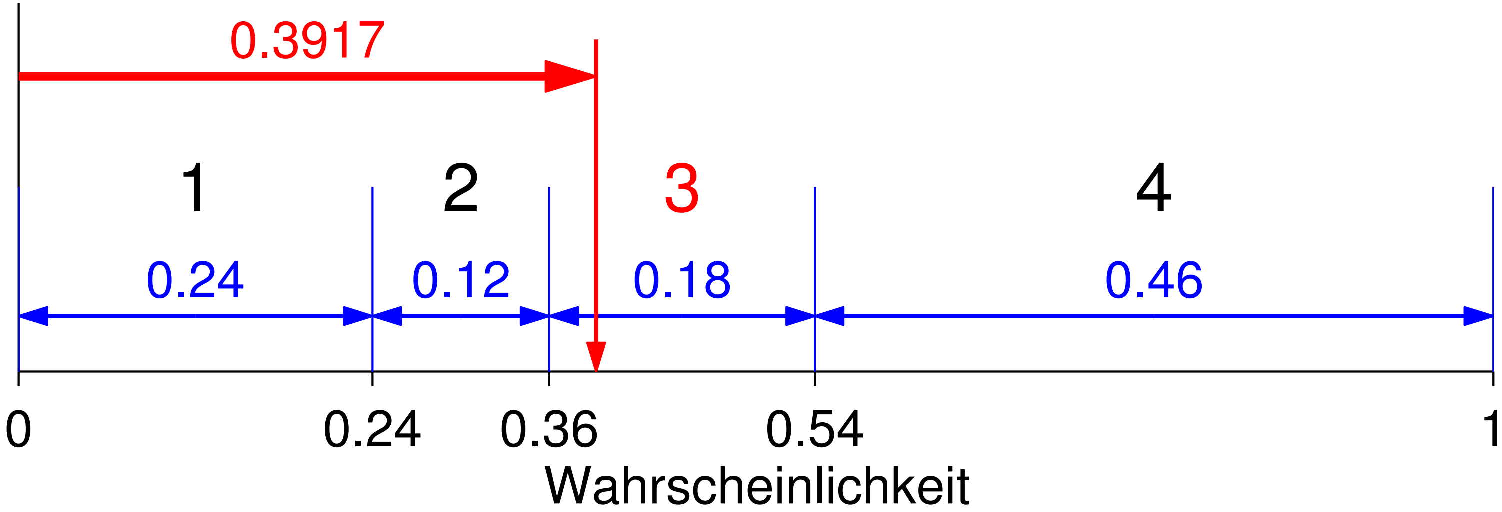


Generieren diskreter Zufallswerte

diskr. Verteilung:

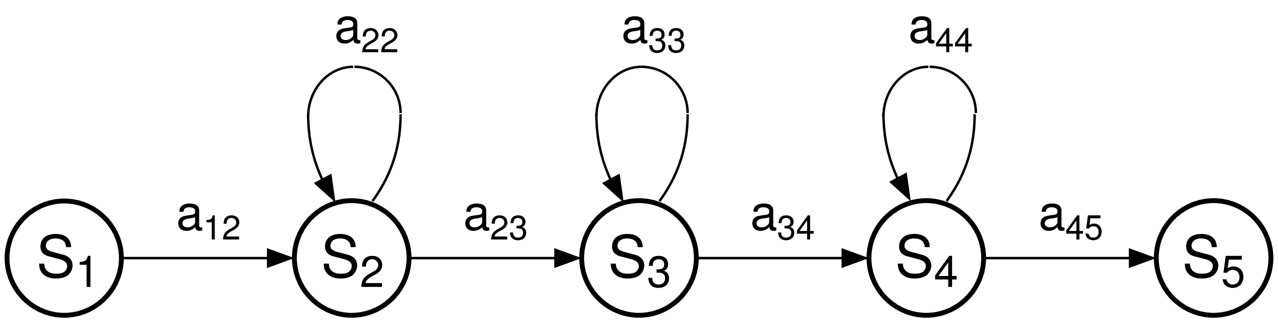
x	$P(\mathrm{x})$
1	0.24
2	0.12
3	0.18
4	0.46

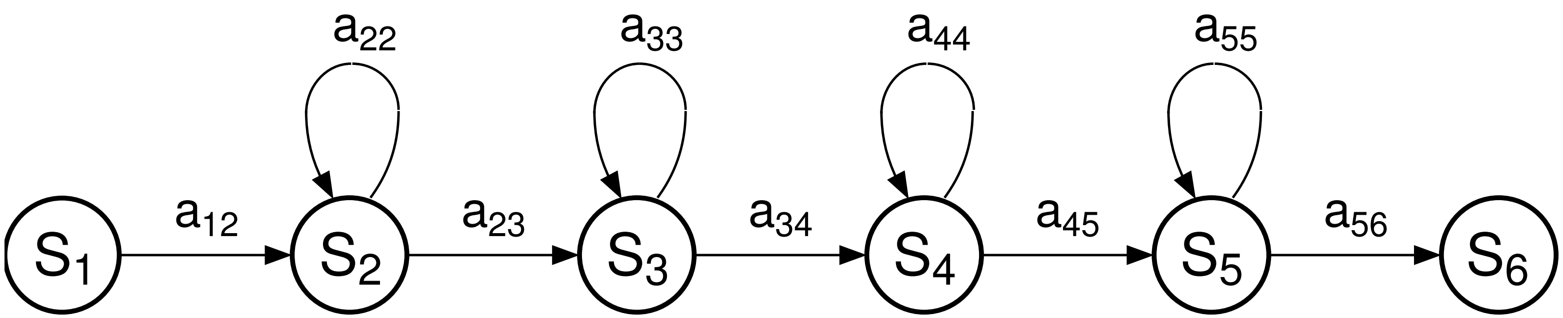
Zufallszahlengenerator $[0 \ldots 1]$ $\rightarrow$ 0.3917



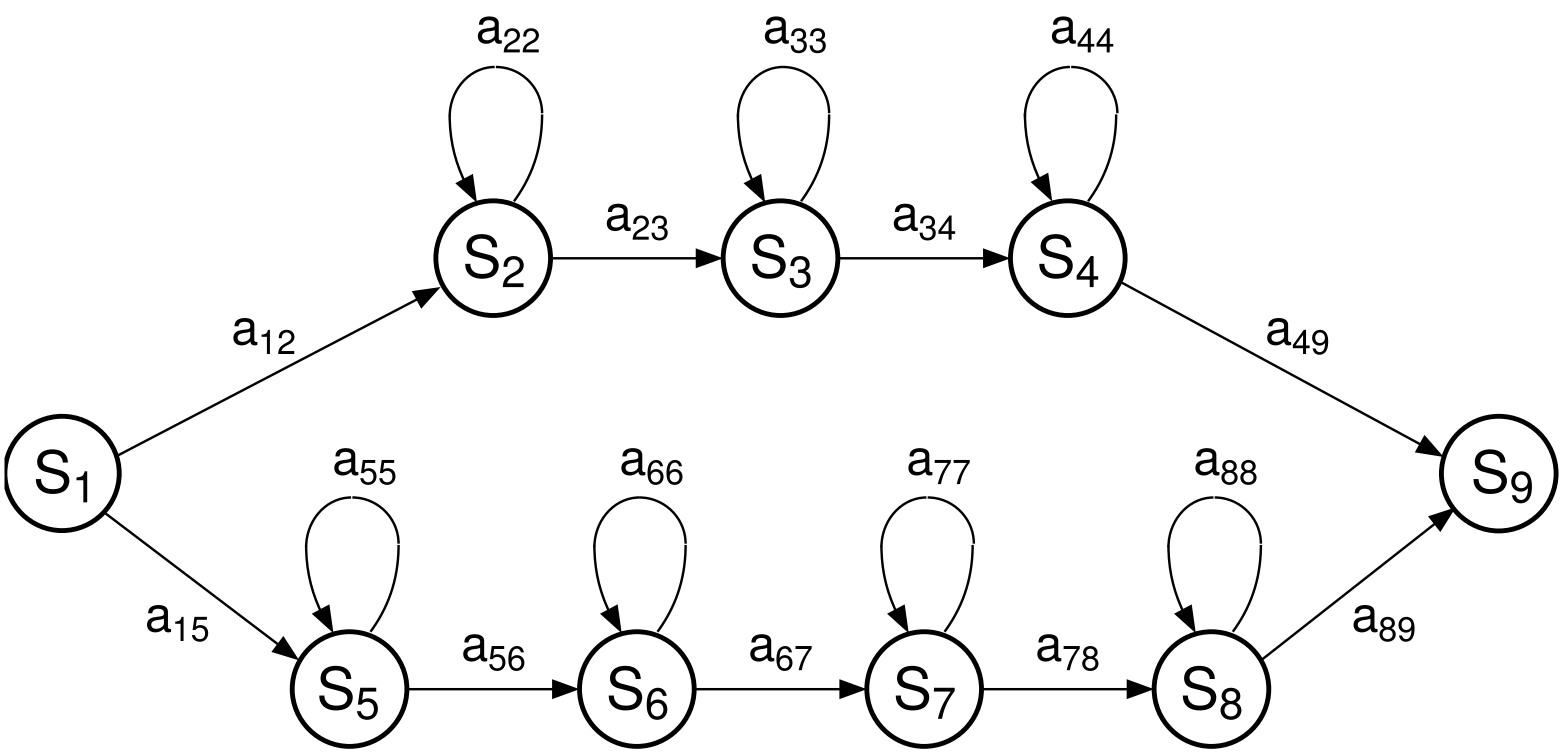
<<<

Parallelschaltung von HMM





Parallelschaltung von HMM



<<<

Parallelschaltung von HMM

$$\lambda_1 = (A_1, B_1) \qquad \text{mit} \qquad A_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\lambda_2 = (A_2, B_2) \qquad \text{mit} \qquad A_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 & 0 \\ 0 & 0 & 0 & a_{44} & a_{45} & 0 \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Parallelschaltung von HMM

$$\lambda_p = (A_p, B_p) \qquad \text{mit} \qquad A_p = \begin{bmatrix} 0 & a_{12} & 0 & 0 & a_{15} & 0 & 0 & 0 & 0 \\ 0 & a_{22} & a_{23} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33} & a_{34} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_{44} & 0 & 0 & 0 & 0 & a_{49} \\ 0 & 0 & 0 & 0 & a_{55} & a_{56} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{66} & a_{67} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & a_{77} & a_{78} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{88} & a_{89} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

<<<