

Hidden Markov Modelle

Dr. Michaela Geierhos

Centrum für Informations- und Sprachverarbeitung
Ludwig-Maximilians-Universität München
Symbolische Programmiersprache WS 2020/2021

02. February 2021

Überblick

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Aufgaben die mit HMMs bearbeitet werden

Viterbi-Algorithmus

Formen von HMMs

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Aufgaben die mit HMMs bearbeitet werden

Viterbi-Algorithmus

Formen von HMMs

Hidden Markov Modelle

Hidden Markov Modelle

sind stochastische Modelle, die auf **Markov-Ketten** beruhen

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Viterbi-Algorithmus

Formen von HMMs

Stochastischer Prozess I

Stochastischer Prozess

ein **stochastischer Prozess** oder **Zufallsprozess** ist eine Folge von elementaren Zufallsereignissen $X_1, X_2, \dots, X_i \in \Omega, i = 1, 2, \dots$

Prozesszustand

die möglichen Zufallswerte in einem stochastischen Prozess heißen **Zustände des Prozesses**.

Man sagt, dass sich der Prozess zum Zeitpunkt t in Zustand X_t befindet

Stochastischer Prozess II

Beispiel

- ein Textgenerator hat ein Lexikon mit drei Wörtern
- von denen an jeder Position jedes auftreten kann ($\Omega = \{\textit{geschickt}, \textit{werden}, \textit{wir}\}$)
- wir beobachten an jeder Position, welches Wort generiert wurde
- Sei:
 - X_1 das Wort zum ersten Beobachtungszeitpunkt
 - X_2 das Wort zum zweiten Beobachtungszeitpunkt
 - ...
- dann ist die Folge der Wörter ein stochastischer Prozess mit diskreter Zufallsvariable und diskretem Zeitparameter

Stochastischer Prozess III

für die vollständige Beschreibung eines Zufallsprozesses mit diskretem Zeitparameter benötigt man

- **Anfangswahrscheinlichkeit**

gibt für jeden Zustand an mit welcher Wahrscheinlichkeit er als Zustand X_1 beobachtet werden kann (d.h. den Startzustand bildet

$$\pi_i = P(X_1 = s_i)$$

- **Übergangswahrscheinlichkeit**

gibt für jeden Zustand an, mit welcher Wahrscheinlichkeit er in einer Zustandsfolge auftritt

$$P(X_{t+1} = x_{t+1} | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t)$$

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Viterbi-Algorithmus

Formen von HMMs

Markov-Kette

Markov-Kette

eine **Markov-Kette** ist ein spezieller stochastischer Prozess, bei dem zu jedem Zeitpunkt die Wahrscheinlichkeiten aller zukünftigen Zustände nur vom momentanen Zustand abhängen
(= Markov-Eigenschaft)

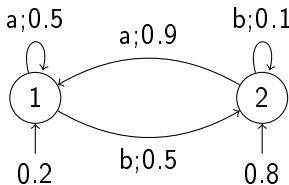
d.h. es gilt:

$$\begin{aligned} P(X_{t+1} = x_{t+1} | X_1 = x_1, X_2 = x_2, \dots, X_t = x_t) &= \\ P(X_{t+1} = x_{t+1} | X_t = x_t) \end{aligned}$$

endliche Markov-Kette

endliche Markov-Kette

für eine **endliche Markov-Kette** gibt es endlich viele Zustände, und die Kette muss sich zu jedem Zeitpunkt in einem dieser endlich vielen Zustände befinden



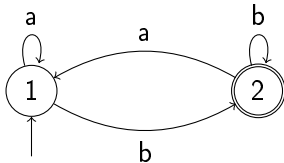
Unterschiede zum endlichen Automaten

folgende Modifikationen (in Fett) ergeben sich in der formalen Spezifikation zu den endlichen Automaten:

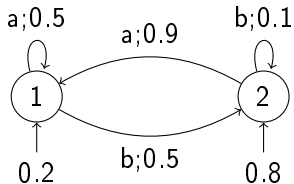
Spezifikation	Beschreibung	Im Beispiel
$S = \{s_1 \dots s_n\}$	Menge der Zustände	$\{1,2\}$
$K = \{k_1 \dots k_m\}$	Menge der Ausgabesymbole	$\{a,b\}$
$A = \{a_{ijx} \dots a_{kly}\} \text{ } i,j,k,l \in S;$ $x,y \in K$	Wahrscheinlichkeiten der Übergänge zwischen Zuständen mit den emittierten Symbolen	$a_{12b} = 0,5$ (beim Wechsel von Zustand 1 zu 2 wird das Symbol b erkannt bzw. emittiert; die Wahrscheinlichkeit des Übergangs ist 0,5)
$P = \{p_i \dots\} \text{ mit } i \in S$	der Startzustand entfällt, an seine Stelle treten die Wahrscheinlichkeiten, in einem Zustand s zu beginnen	$1 \rightarrow 0,2$
$e_1, e_2, \dots \in S$	die Menge der Endzustände entfällt	-

endlicher Automat vs. endliche Markov-Kette

endlicher Automat



endliche Markov-Kette



Markov-Kette

Matrix-Darstellung I

Eine endliche Markov-Kette kann beschrieben werden durch die Angaben einer statistischen Übergangsmatrix A :

$$\begin{aligned}
 a_{ij} &= P(X_{t+1} = s_j | X_t = s_i) \\
 \forall_{i,j} a_{ij} &\geq 0 \\
 \forall_i \sum_{j=1}^N a_{i,j} &= 1
 \end{aligned}$$

$X_t = s_i$	$X_{t+1} = s_j$		
	geschickt	werden	wir
geschickt	0.3	0.4	0.3
werden	0.4	0.2	0.4
wir	0.3	0.4	0.3

Markov-Kette

Matrix-Darstellung II

... und durch Angabe der Anfangswahrscheinlichkeiten Π :

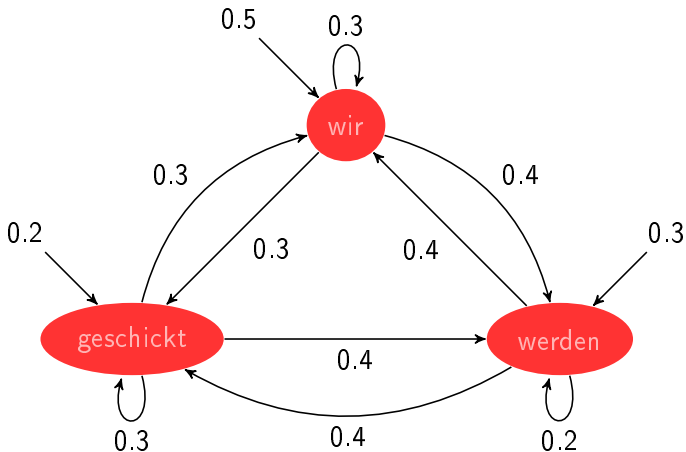
$$\begin{aligned}\pi_i &= P(X_1 = s_i) \\ \sum_{i=1}^N \pi_i &= 1\end{aligned}$$

X_t	π
geschickt	0.2
werden	0.3
wir	0.5

Markov-Kette

Graph-Darstellung

...oder durch einen Zustandsübergangsgraphen:



Markov-Kette

Berechnung einer Sequenz-Wahrscheinlichkeit

Wahrscheinlichkeit der Sequenz der Zustände $X_1 \dots X_T$

$$\begin{aligned} P(X_1, \dots, X_T) \\ = P(X_1)P(X_2|X_1)P(X_3|X_1, X_2)\dots P(X_T|X_1, \dots, X_{T-1}) \end{aligned}$$

für eine Markov-Kette gilt:

$$\begin{aligned} &= P(X_1)P(X_2|X_1)P(X_3|X_2)\dots P(X_T|X_{T-1}) \\ &= \pi_{X_1} \prod_{t=1}^{T-1} a_{X_t X_{t+1}} \end{aligned}$$

Markov-Kette

Berechnungsbeispiel

Wahrscheinlichkeit der Sequenz

$P(X_1 = \text{wir}, X_2 = \text{werden}, X_3 = \text{geschickt})$

X_t	π
geschickt	0.2
werden	0.3
wir	0.5

$X_t = s_i$	$X_{t+1} = s_j$		
	geschickt	werden	wir
geschickt	0.3	0.4	0.3
werden	0.4	0.2	0.4
wir	0.3	0.4	0.3

$$\begin{aligned}
 &= P(X_1 = \text{wir}) \times \\
 &\quad P(X_2 = \text{werden} | X_1 = \text{wir}) \times \\
 &\quad P(X_3 = \text{geschickt} | X_2 = \text{werden}) \times \\
 &= (0.5 \times 0.4 \times 0.4) = 0.08
 \end{aligned}$$

Markov-Modell (MM)

- ein Markov-Modell ordnet jedem Zustand (andere Variante: jedem Zustandsübergang) eine **Ausgabe** zu, die ausschließlich vom aktuellen Zustand (bzw. Zustandsübergang) abhängig ist
- Ausgabe: Sequenz von Ereignissen, die die Beobachtungen in der Beobachtungssequenz repräsentieren

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Viterbi-Algorithmus

Formen von HMMs

Hidden Markov Modell (HMM)

Beschreibung

Hidden Markov Modell ist ein Markov-Modell

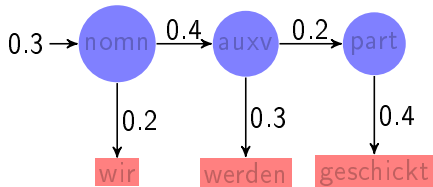
- bei dem nur die Sequenz der Ausgaben beobachtbar ist
- die Sequenz der Zustände verborgen bleibt

es kann mehrere Zustandssequenzen geben, die dieselbe Ausgabe erzeugen

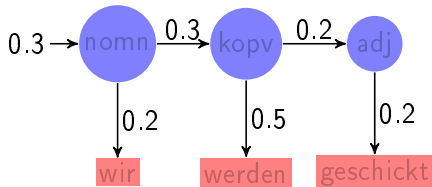
Hidden Markov Modell

Beispiel

- in einem Text lassen sich nur die **Ausgaben** (= produzierte Wörter) beobachten (visible)
- die Sequenz von **Zuständen** (= Wortarten), die die Wörter ausgeben, (Satzmuster) lässt sich nicht beobachten (hidden)
- mehrere Sequenzen können dieselbe Ausgabe erzeugen



$$0.3 \times 0.2 \times 0.4 \times 0.3 \times 0.2 \times 0.4 = 0.000576)$$



$$0.3 \times 0.2 \times 0.3 \times 0.5 \times 0.2 \times 0.5 = 0.00036)$$

Hidden Markov Modell

Definition

Ein HMM wird spezifiziert durch ein Fünf-Tupel (S, K, Π, A, B)

$$(S , K , \Pi , A , B)$$

- Menge der Zustände
 - $S = \{s_1, \dots, s_N\}$
- Menge der Ausgangssymbole
 - $K = \{k_1, \dots, k_M\}$
- Wahrscheinlichkeiten der Startzustände
 - $\pi_i = P(X_1 = s_i)$
 - $\sum_{i=1}^N \pi_i = 1$
- Wahrscheinlichkeiten der Zustandsübergänge
 - $a_{ij} = P(X_{t+1} = s_j | X_t = s_i)$, mit $1 \leq i, j \leq N$
 - $\sum_{i=1}^N a_{ij} = 1$
- Wahrscheinlichkeiten der Symbolemissionen in Zustand j
 - $b_i(k) = P(K_k \text{ in } t | X_t = s_i)$, mit $1 \leq i \leq N, 1 \leq k \leq M$
 - $\sum_{i=1}^M b_i(k) = 1$

Hidden Markov Modell

Beispiel

	Übergangsmatrix					Emissionsmatrix				Startwahrscheinlichkeit
X_t	X_{t+1}					O_t				π
	Adje	AuxV	KopV	Nomn	Part	geschickt	werden	wir	...	
Adje	.2	.1	.1	.4	.2	.2	0	0	.8	.3
AuxV	.2	.3	.1	.2	.2	0	.3	0	.7	.2
KopV	.2	.2	.1	.4	.1	0	.5	0	.5	.1
Nomn	.1	.4	.3	.1	.1	0	0	.2	.8	.3
Part	.3	.1	.2	.1	.3	.4	0	0	.6	.1

Hidden Markov Modell: Gewinnung der Daten (1)

Annotation eines Korpus

wir	werden	geschickt	vom	König	.
nomn	auxv	part	Ω

wir	werden	geschickt	durch	Übung	.
nomn	kopv	adje	Ω

Hidden Markov Modell: Gewinnung der Daten (2)

Auszählung der Sequenzen

	Adje	AuxV	KopV	Nomn	Part	Ω	geschickt	werden	wir	.
Adje	–	–	–	–	–	1	1	–	–	–
AuxV	–	–	–	–	1	–	–	1	–	–
KopV	1	–	–	–	–	–	–	1	–	–
Nomn	–	1	1	–	–	–	–	–	2	–
Part	–	–	–	–	–	1	1	–	–	–
Ω	–	–	–	1	–	–	–	–	–	2

Tabelle: Ausgezählte Sequenzen

Hidden Markov Modell: Gewinnung der Daten (3)

Umrechnung der Häufigkeiten in prozentuale Anteile

	Adje	AuxV	KopV	Nomn	Part	Ω	geschickt	werden	wir	.
Adje	—	—	—	—	—	1.0	1.0	—	—	—
AuxV	—	—	—	—	1.0	—	—	1.0	—	—
KopV	1.0	—	—	—	—	—	—	1.0	—	—
Nomn	—	0.5	0.5	—	—	—	—	—	1.0	—
Part	—	—	—	—	—	1.0	1.0	—	—	—
Ω	—	—	—	1.0	—	—	—	—	—	1.0

Tabelle: Umgerechnete Wahrscheinlichkeiten

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Aufgaben die mit HMMs bearbeitet werden

Viterbi-Algorithmus

Formen von HMMs

Drei grundlegende Aufgaben, die mit HMMs bearbeitet werden

- Filtern/Evaluierung: Wahrscheinlichkeit einer Beobachtung finden
 - brute force
 - Forward-Algorithmus / Backward-Algorithmus
- Dekodierung: Beste Pfad-Sequenz finden
 - brute force
 - Viterbi-Algorithmus
- Lernen: Aufbau des besten Modells aus Trainingsdaten
 - Baum-Welch-Algorithmus

Wahrscheinlichkeit einer Beobachtung finden

gegeben

- eine Sequenz von Beobachtungen

$$O = (o_1, \dots, o_n)$$

$$O = (\text{wir, werden, geschickt})$$

- ein Modell

$$\mu = (A, B, \Pi)$$

	Adje	AuxV	KopV	Nomn	Part	g'schickt	werden	wir	..	π
Adje	.2	.1	.1	.4	.2	.2	0	0	.8	.3
AuxV	.2	.3	.1	.2	.2	0	.3	0	.7	.2
KopV	.2	.2	.1	.4	.1	0	.5	0	.5	.1
Nomn	.1	.4	.3	.1	.1	0	0	.2	.8	.3
Part	.3	.1	.2	.1	.3	.4	0	0	.6	.1

gesucht

- die Wahrscheinlichkeit

$$P(O|\mu)$$

$$P(\text{wir, werden, geschickt}|\mu)$$

Wahrscheinlichkeit einer Beobachtung finden

Lösungsweg 1: bruteforce

für alle möglichen Zustandsfolgen

- Berechnung der Wahrscheinlichkeit der Beobachtungen
- Summierung der Wahrscheinlichkeiten

$$\begin{aligned}
 P(O|\mu) &= \sum_X P(O|X, \mu) P(X|\mu) \\
 &= \sum_{X_1 \dots X_T} \pi_{X_1} b_{X_1 O_1} \prod_{t=1}^{T-1} a_{X_1 X_{t+1}} b_{X_{t+1} O_{t+1}}
 \end{aligned}$$

- state transition
- symbol emission



Wahrscheinlichkeit einer Beobachtung finden

Lösungsweg 1: bruteforce

Beispiel

$$P(O|\mu) = \sum_{X_1 \dots X_T} \pi_{X_1} b_{X_1 O_1} \prod_{t=1}^{T-1}$$

	$P(\text{wir,werden,geschickt} \text{Adje,Adje,Adje}, \mu)$		=	0.0
+	$P(\text{wir,werden,geschickt} \text{Adje,Adje,AuxV}, \mu)$		=	0.0
...			=	0.0
+	$P(\text{wir,werden,geschickt} \text{Nomn,AuxV,Part}, \mu)$	$0.3 \times 0.2 \times 0.4 \times 0.3 \times 0.2 \times 0.4$	=	0.000576
...			=	0.0
+	$P(\text{wir,werden,geschickt} \text{Nomn,KopV,Adj}, \mu)$	$0.3 \times 0.2 \times 0.3 \times 0.5 \times 0.2 \times 0.2$	=	0.000360
...			=	0.0
+	$P(\text{wir,werden,geschickt} \text{Part,Part,Part}, \mu)$		=	0.0
<hr/>				
=	...		=	0.00936

Wahrscheinlichkeit einer Beobachtung finden

Lösungsweg 1: bruteforce

Effizienz

$$P(O|\mu) = \sum_{x_1 \dots x_T} \pi_{x_1} b_{x_1 o_1} \prod_{t=1}^{T-1}$$

Lösungsweg ist hoffnungslos ineffizient
benötigt im allgemeinen Fall, wobei

- Start in jedem Zustand möglich
- jeder Zustand kann auf jeden folgen

$(2T - 1) \times N^T$ Multiplikationen

T Anzahl der Beobachtungen

N Anzahl der Zustände

Wahrscheinlichkeit einer Beobachtung finden

Lösungsweg 2: Vorwärts-und Rückwärts-Verfahren

- Forward procedure
- Backward procedure

Merken partieller Ergebnisse statt wiederholter Berechnung

Beste Pfadsequenz finden

Lösungsweg 2: Vorwärts-und Rückwärts-Verfahren

gegeben

- eine Sequenz von Beobachtungen

$$O = (o_1, \dots, o_n)$$

$$O = (\text{wir}, \text{werden}, \text{geschickt})$$

- ein Modell

$$\mu = (A, B, \Pi)$$

	Adje	AuxV	KopV	Nomn	Part	g'schickt	werden	wir	..	π
Adje	.2	.1	.1	.4	.2	.2	0	0	.8	.3
AuxV	.2	.3	.1	.2	.2	0	.3	0	.7	.2
KopV	.2	.2	.1	.4	.1	0	.5	0	.5	.1
Nomn	.1	.4	.3	.1	.1	0	0	.2	.8	.3
Part	.3	.1	.2	.1	.3	.4	0	0	.6	.1

gesucht

- die wahrscheinlichste Pfadsequenz

$$P(\text{wir}, \text{werden}, \text{geschickt} | \mu)$$

$$\arg_X \max P(X | O, \mu)$$

Beste Pfadsequenz finden

- Lösungsweg 1: brute force
 - alle Varianten berechnen
 - die wahrscheinlichste auswählen
 - hoffnungslos ineffizient
- Lösungsweg 2: beste Einzelzustände
 - für jeden Zeitpunkt t :
 - Zustand mit höchster Ausgabewahrscheinlichkeit auswählen
 - Zusammensetzung kann unwahrscheinliche Sequenzen erkennen

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Aufgaben die mit HMMs bearbeitet werden

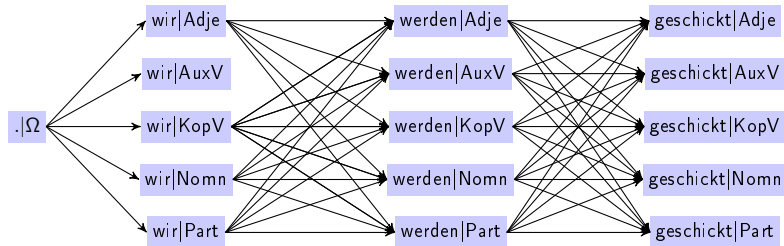
Viterbi-Algorithmus

Formen von HMMs

Beste Pfadsequenz finden

Lösungsweg 3: Viterbi-Algorithmus

Speichert für jeden Zeitpunkt t die Wahrscheinlichkeit des wahrscheinlichsten Pfades, der zu einem Knoten führt



Training der Modellparameter

gegeben

- eine Sequenz von Beobachtungen in einem Trainingscorpus
- $O = (o_1, \dots, o_T)$

gesucht

- ein Modell
 - $\mu = (A, B, \Pi)$
 - das für die beobachteten Sequenzen im Trainingscorpus die maximalen Wahrscheinlichkeiten erzeugt

$$\arg_{\mu} \max P(O_{Training} | \mu)$$

Lösung

- Baum-Welch-Algorithmus
- Forward-backward-Algorithmus

Definition

Stochastischer Prozess

Markov-Kette

Hidden Markov Modell

Viterbi-Algorithmus

Formen von HMMs

Formen von Hidden Markov Modellen: Emissionen

- auf den vorangehenden Folien wurde ein **State Emission Modell** verwendet
- den allgemeinen Fall stellt ein **Arc Emission Modell** dar
- ein State Emission Modell kann in ein Arc Emission Modell überführt werden, umgekehrt ist dies nicht immer möglich
- auf den folgenden Folien wird ein **Arc Emission Modell** beschrieben

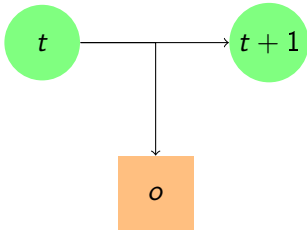
Formen von Hidden Markov Modellen: Emissionen

Allgemeine Form

Arc Emission Modell

zur Zeit t emittiertes Symbol
hängt ab von:

- Zustand zur Zeit t
- Zustand zur Zeit $t + 1$

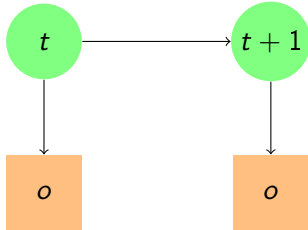


Spezielle Form

State Emission Modell

zur Zeit t emittiertes Symbol
hängt ab von:

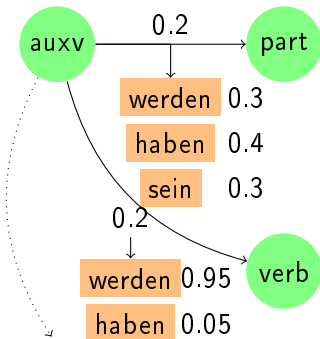
- Zustand zur Zeit t



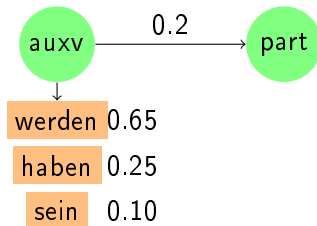
Formen von Hidden Markov Modellen: Emissionen

Beispiel

Arc Emission Modell



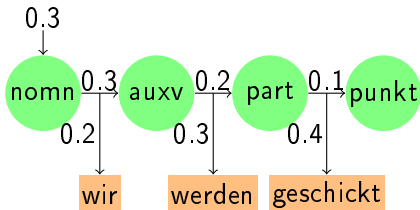
State Emission Modell



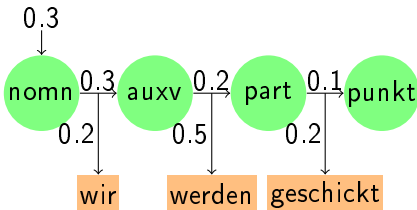
Arc Emission Modell

Beispiel

- in einem Text lassen sich nur die Ausgabe ● (= produzierte Wörter) beobachten (visible)
- die Sequenz von Zuständen ■ (= Wortarten), die die Wörter ausgeben, (Satzmuster) lässt sich nicht beobachten (hidden)
- mehrere Sequenzen können dieselbe Ausgabe erzeugen:



$$0.3 \times 0.3 \times 0.2 \times 0.2 \times 0.3 \times 0.1 \times 0.4 = 0.0000432$$



$$0.3 \times 0.3 \times 0.2 \times 0.2 \times 0.5 \times 0.1 \times 0.2 = 0.000036$$

Arc Emission Modell

Darstellung als Wahrscheinlichkeitsmatrix

	Übergangsmatrix								Start	
X_t	X_{t+1}									
	Adje				AuxV	KopV	Nomn	Part	Punkt	π
Adje	.2				.1	.1	.4	.1	.1	.3
	Emissionsmatrix									
	o_t									
	geschickt	werden	wir	...						
	.2	0	0	.8						
AuxV	.2				.3	.1	.1	.2	.1	.2
KopV	.2				.1	.1	.4	.1	.1	.1
	Emissionsmatrix									
	o_t									
	geschickt	werden	wir	...						
	0.05	.5	.05	.4						
Nomn	.05				.4	.3	.05	.1	.1	.3
Part	.3				.1	.1	.1	.3	.1	.1
Punkt	.2				.2	.1	.3	.1	.1	.1

Arc Emission Modell

Spezialfall: State Emission Modell

	Übergangsmatrix									Start
X _t	X _{t+1}									
	Adje				AuxV	KopV	Nomn	Part	Punkt	π
Adje	.2				.1	.1	.4	.1	.1	.3
	Emissionsmatrix									
	o _t									
	geschickt	werden	wir	...						
	.2	0	0	.8						
AuxV	.2				.3	.1	.1	.2	.1	.2
KopV	.2				.1	.1	.4	.1	.1	.1
	Emissionsmatrix									
	o _t									
	geschickt	werden	wir	...						
	0.05	.5	.05	.4						
Nomn	.05				.4	.3	.05	.1	.1	.3
Part	.3				.1	.1	.1	.3	.1	.1
Punkt	.2				.2	.1	.3	.1	.1	.1

Wenn die Emissionsverteilungen für alle Übergänge aus einem

Arc Emission Modell

Definition

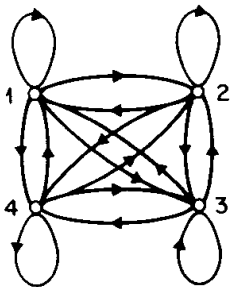
(S , K , Π , A , B)

- Menge der Zustände
 - $S = \{S_1, \dots, S_N\}$
- Menge der Ausgangssymbole
 - $K = \{k_1, \dots, k_M\}$
- Wahrscheinlichkeiten der Startzustände
 - $\pi = P(X_1 = S_i)$
 - $\sum_{i=1}^N \pi_i = 1$
- Wahrscheinlichkeiten der Zustandsübergänge
 - $a_{ij} = P(X_{t+1} = S_j | X_t = S_i)$, mit $1 \leq i, j \leq N$
 - $\sum_{i=1}^N a_{ij} = 1$
- Wahrscheinlichkeiten der Symbolemissionen
 - $b_{ijk} = P(K_k \text{ bei Übergang von } X_t \text{ zu } X_{t+1} | X_t = S_j, X_{t+1} = S_j)$, mit $1 \leq j \leq N, 1 \leq k \leq M$
 - $\sum_{k=1}^M b_{ijk}(k) = 1$



Formen von Hidden Markov Modellen

- ergodisches Modell
- jeder Zustand kann von jedem in einer endlichen Anzahl von Schritten erreicht werden



- andere Arten z.B. in der Verarbeitung gesprochener Sprache verwendet :
- links-rechts-Modell

