Hidden Markov Modelle

in der Sprachverarbeitung

Paul Gabriel

paul@pogo.franken.de

Überblick

- Merkmalsvektoren
- Stochastischer Prozess
- Markov-Ketten
- Hidden Makrov Modelle (HMMs)
- Worterkennung
- Aufgaben und grundlegende Fragen zu und über HMMs

Ausgangssituation

Wir haben:

- Mel-Cepstrum Merkmalsvektoren

Wir wollen:

- Wörter erkennen

Merkmalsvektoren

Problem:

- zeitliche Struktur (kurz - lang gesprochene Wörter)
- spektrale Zusammensetzung (Sprecher abhängig)

müssen vom Worterkennungssystem berücksichtigt werden

Bis in die achziger Jahre: Dynamic Time Warping (DTW)
Hidden Markov Modelle (HMMs) haben sich durchgesetzt

Prinzip

- Für jedes bekannte Wort ω_λ wird ein Wortmodell HMM_λ definiert
- Für jedes Wort wird die Wahrscheinlichkeit berechnet, dass das zugehörige HMM_λ aktiv war, wenn eine bestimmte Beobachtung gemacht wurde

Warum ‚hidden‘?

HMMs sind doppelt *stochastische Prozesse:* Die Dynamik der Zustände ist stochastisch (markovsch) und die Erzeugung der beobachteten Meßwerte für die jeweiligen Zustände ebenfalls

HMM postulieren zweistufigen Zufallsprozess für die Datenprodution während der Artikulation eines Wortes:

- unsichtbare Zustände: zeitliche Strukur in Form der Markovkette
- sichtbare Ausgabe: Sequenz von Symbolen (diskrete HMMs) oder Vektoren (kontinuierliche HMMs) gemäß einer zustandsbezogenen Ausgabeverteilung

Stochastischer Prozess

(auch *Zufallsprozess*) beschreibt eine Folge von Zufallsereignissen,

- die nicht voneinander unabhängig sind
- deren Werte von vorangegangenen Elementen der Sequenz abhängen

für die vollständige Beschreibung notwendig:

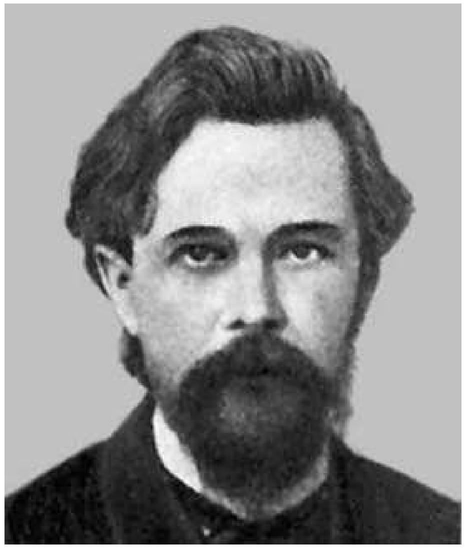
- Anfangswahrscheinlichkeiten eines jeden Zustands
- Übergangswahrscheinlichkeiten

Stochastischer Prozess

Beispiel

- Ein Textgenerator hat ein Lexikon mit drei Wörtern
- von denen an jeder Position jedes auftreten kann $(\Omega = \{fertig, wir, sind\})$
- 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 Zufallsvariabler und diskretem Zeitparameter

Andrei Markov



Andrei Andreyevich Markov
*14. Juni 1856 †20. Juli 1922
Professor in Sankt Petersburg

Markov-Ketten

Definition:
Eine *Markov-Kette* ist ein spezieller stochastischer Prozess, bei dem zu jedem Zeitpunkt die Wahrscheinlichkeit aller zukünftigen Zustände nur vom momentanen Zustand abhängen *(Markov-Eigenschaft)*

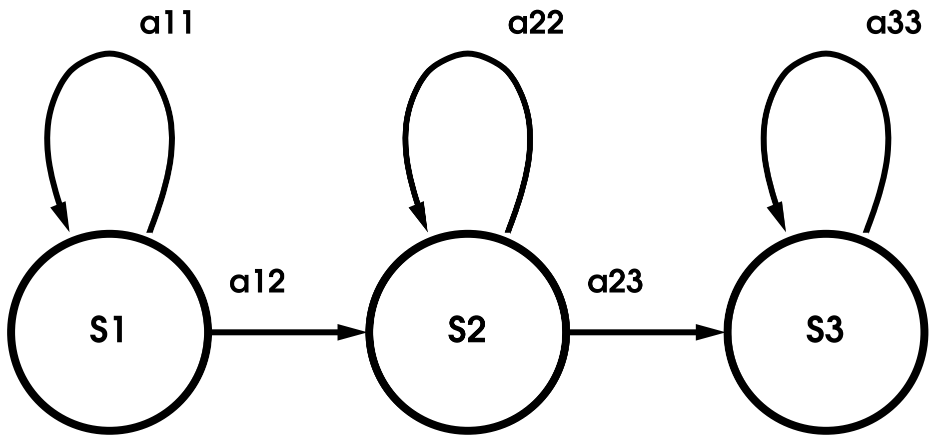
Es gilt:
$$P(X_{t+1} = x_{t+1} \mid X_1 = x_1, X_2 = x_2, \ldots, X_t = x_t) \approx$$
$$P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

HMMs sind Wahrscheinlichkeitsmodelle, die auf Markov-Ketten beruhen

Markov-Ketten

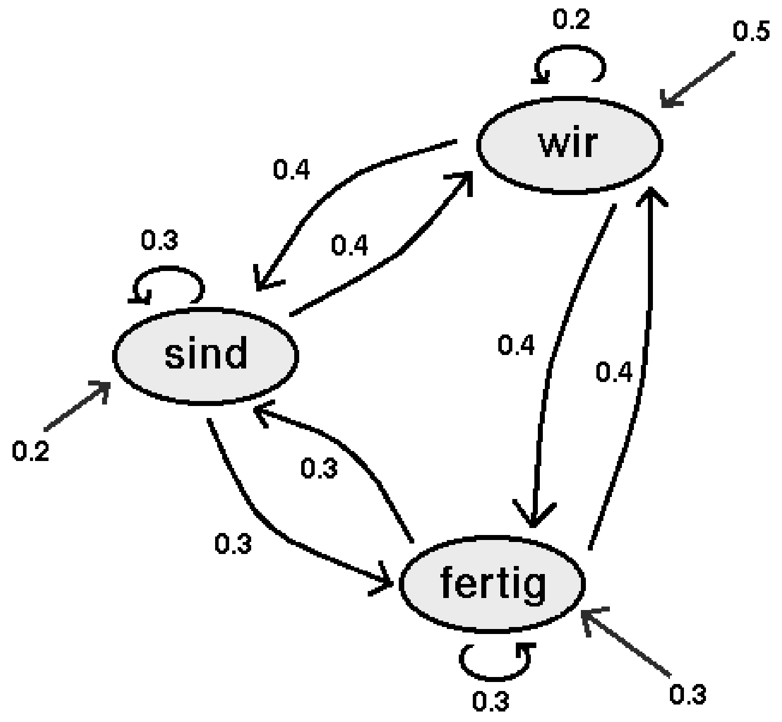
für eine *endliche Markov-Kette* gibt es endlich viele Zustände und die Kette muss sich zu jedem Zeitpunkt in einem dieser Zustände befinden

Somit haben Markov-Ketten die Eigenschaft eines **endlichen Automaten**



Markov-Ketten

können beschrieben werden durch einen Zustandsübergangsgraphen



Markov-Ketten

werden beschrieben durch:

- Stochastische Übergangsmatrix A

  $a_{ij} = P(X_{t+1} = s_j \mid X_t = s_i)$

  $a_{ij} \geq 0 \ \forall i, j$

  $\sum_{j=1}^{N} a_{i,j} = 1 \ \forall i$

	fertig	wir	sind
fertig	0,3	0,4	0,3
wir	0,4	0,2	0,4
sind	0,3	0,4	0,3

Markov-Ketten

- Anfangswahrscheinlichkeiten Π
  $\pi_i = P(X_1 = s_i)$
  $\sum_{i=1}^{N} \pi_i = 1$

X_t	π
fertig	0,3
wir	0,5
sind	0,2

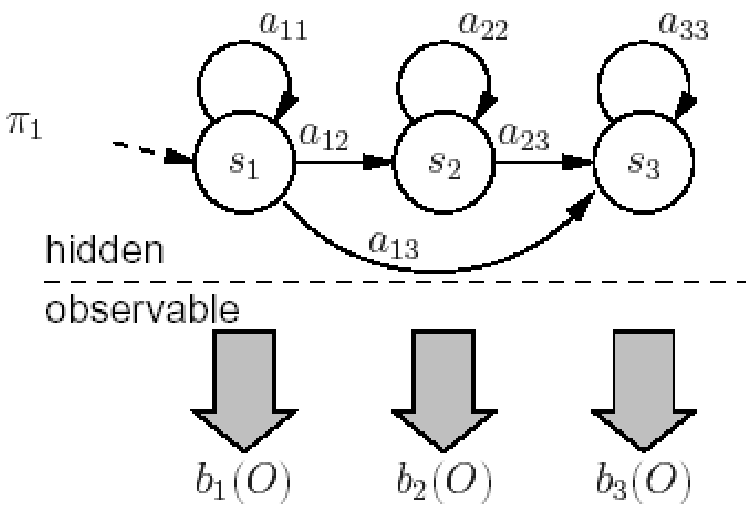
Markov-Modell (MM)

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

Wird auch Visible Markov Modell (VMM) genannt

Warum ‚hidden‘? - Teil 2

- Ein *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 Zustadnssequenzen geben, die die selbe Ausgabe erzeugen



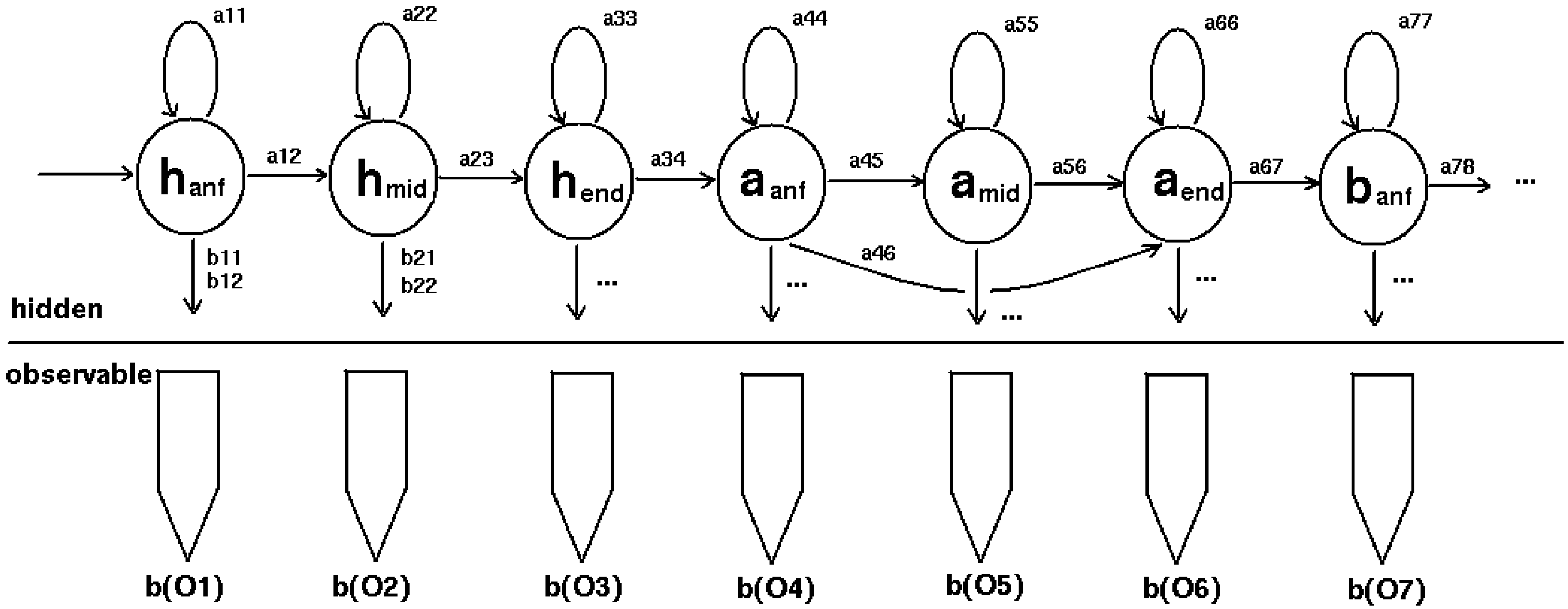
Definition

Ein *Hidden Markov Modell* ist ein Tupel $\lambda = (S, V, \Pi, A, B)$
Für eine Folge von Zuständen $q_1, q_2, \ldots, q_T$ und den dazugehörigen Ausgaben $O_1, O_2, \ldots, O_T$ definiert man:

- $S = (S_1, S_2, \ldots, S_N)$: endliche Menge von Zuständen
- $V = (v_1, v_2, \ldots, v_M)$: Ausgabealphabet
- $\Pi = (\pi_i)_{1 \le i \le N}$: Anfangswahrscheinlichkeiten der Zustände $\pi_i = P(q_1 = S_i)$
- $A = (a_{ij})_{1 \le i,j \le N}$: Übergangswahrscheinlichkeiten der Zustände $a_{ij} = P(q_t = S_j \mid q_{t-1} = S_i)$
- $B = (b_{ik})_{1 \le i \le N, 1 \le k \le M}$: Ausgabewahrscheinlichkeiten $b_{ik} = P(O_t = v_k \mid q_t = S_i)$ für $1 \le t \le T$

Wortbeispiel

Das Wort „haben“



Worterkennung

1. wähle Zustandsmenge S: *Wortuntereinheiten*, z.B. Phoneme
2. baue je Wort ω_λ ein *Wortmodell* HMM_λ, das M Zustände hat
3. *segmentiere* das Sprachsignal einer Äußerung in Wortuntereinheiten
4. betrachte die Segmentierung als *Beobachtung* $\mathbf{o}$
5. bestimme das HMM, das *am wahrscheinlichsten* diese Beobachtung generierte
6. *klassifiziere* das Sprachsignal als das Wort, das durch das wahrscheinlichste HMM modelliert wurde

Grundlegende Aufgaben

- Wahrscheinlichkeit einer Beobachtung finden
  - brute force
  - Vorwärts-Algorithmus, Rückwärts-Alogrithmus
- Dekodierung: Bestimmung der wahrscheinlichsten Zustandsfolge
  - brute force
  - Viterbi-Algorithmus
- Training: Aufbau eines optimalen Modells aus Trainigsdaten
  - Baum-Welch-Algorithmus
  - Viterbi-Training

Produktionswahrscheinlichkeit

Welcher einzelne Zustand zur Zeit t passt am besten zur Beobachtungsfolge?

Gegeben: Folge von Meßwerten $O = O_1, O_2, \ldots, O_T$ und das HMM $\lambda = (S, V, \Pi, A, B)$
Gesucht: Ein Algorithmus zur effizienten Berechnung von $P(O \mid \lambda)$ *(Produktionswahrscheinlichkeit)*

Lösung 1: direkte Berechnug
Aufaddieren der Wahrscheinlichkeiten aller möglichen Statusfolgen

Lösung 2: Vorwärts- und Rückwärts-Algorithmus

Produktionswahrscheinlichkeit

Lösung 1: direkte Berechnug
Aufaddieren der Wahrscheinlichkeiten aller möglichen Statusfolgen

Übergangsmatrix

$$A = \{a_{ij}\} = \begin{pmatrix} 0,3 & 0,4 & 0,3 \\ 0,4 & 0,2 & 0,4 \\ 0,3 & 0,4 & 0,3 \end{pmatrix}$$

Anfangswahrscheinlichkeiten

$$\Pi = \{\pi_i\} = (0,3 \quad 0,5 \quad 0,2)$$

$$O = (wir, sind, fertig) \qquad \Rightarrow O = (S_2, S_3, S_1)$$

$$P(O \mid (\Pi, A)) = \pi_2 \cdot a_{23} \cdot a_{31} = 0,06$$

Produktionswahrscheinlichkeit

direkte Berechnung

- Start in jedem Zustand möglichen
- Jeder Zustand kann auf jeden folgen

Aufaddieren der Wahrscheinlichkeiten aller möglichen Statusfolgen

$P(O \mid \lambda) = \sum_{q \in \mathcal{Q}^T} P(O, q \mid \lambda) =$

$\sum_{q \in \mathcal{Q}^T} \pi_{q_1} b_{q_1}(O_1) \cdot \prod_{t=2}^{T} a_{q_{t-1} q_t} b_{q_t}(O_t)$

$\Rightarrow 2TN^T$ Multiplikationen $\Rightarrow$ *exponentielle Komplexität*

5 Zustände, 12 Zeiteinheiten Dauer $\Rightarrow 5,8 \cdot 10^9$ Multiplikationen

Produktionswahrscheinlichkeit

Lösung 2: Vorwärts- und Rückwärts-Algorithmus
Vorwärtswahrscheinlichkeit: $\alpha_t(j) = P(O_1 \ldots O_t, q_t = j \mid \lambda)$

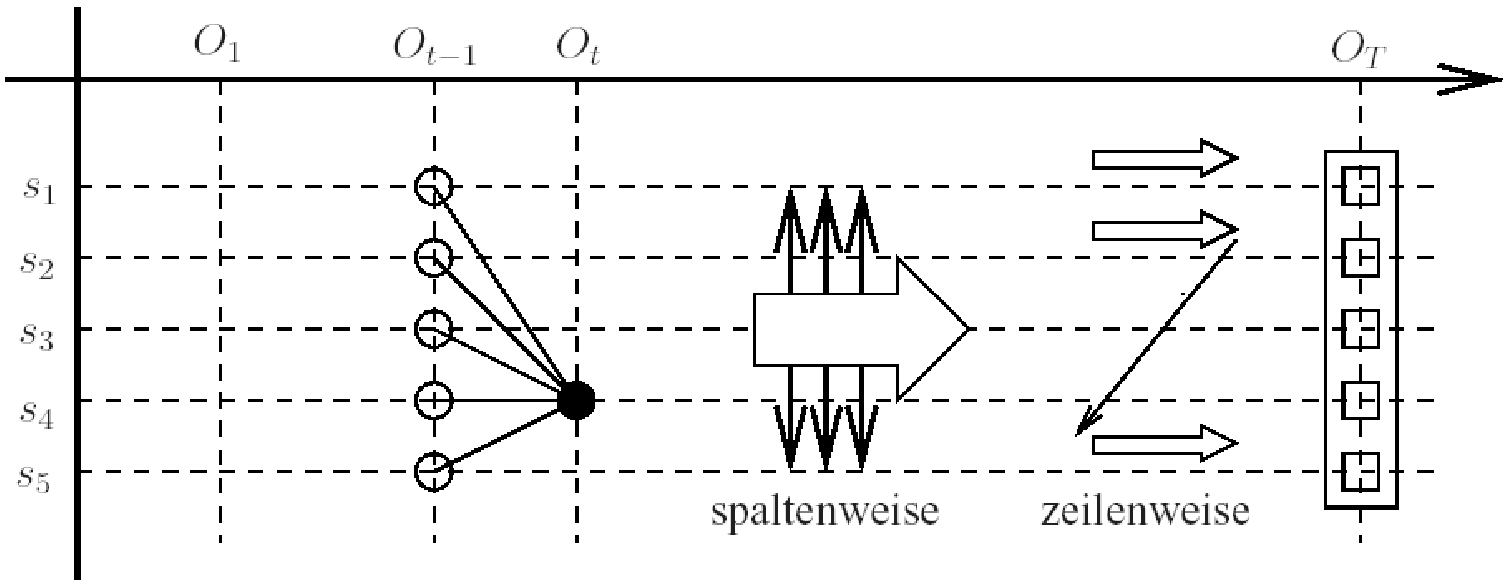
Vorwärtswahrscheinlichkeiten:

- Initialisierung: $\forall j = 1, \ldots, N$ setze $\alpha_1(j) = \pi_j b_j(O_1)$
- Rekursion: für $t < 1$ und alle $j = 1, \ldots, N$ setze

$$\alpha_t(j) = \left(\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij} \right) b_j(O_t)$$

- Terminierung: Berechne $P(O \mid \lambda) = \sum_{j=1}^{N} \alpha_T(j)$

Produktionswahrscheinlichkeit



$\Rightarrow N^2T$ Operationen

Rückwärtsalgorithmus analog mit Rückwärtswahrscheinlichkeiten mit gleichem Rechenaufwand

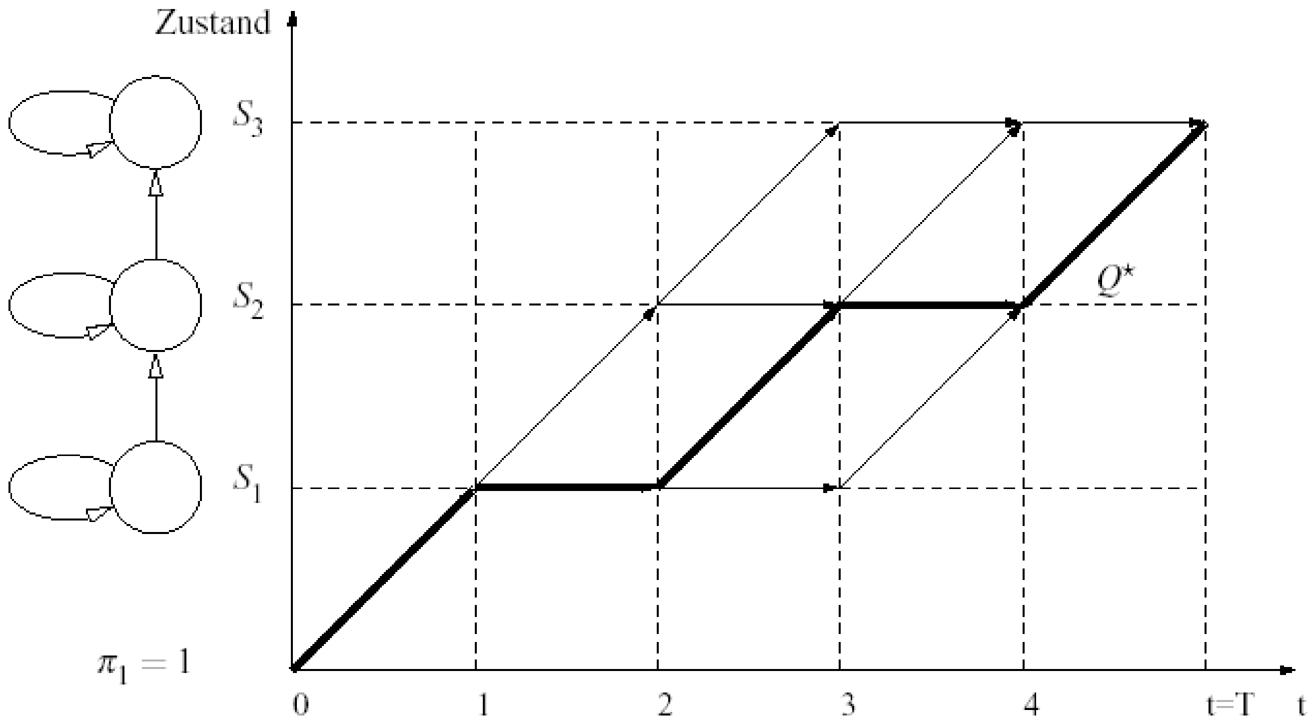
wahrscheinlichste Zustandsfolge

Welche komplette Zustandsfolge passt am besten zur Beobachtungsfolge?

Gegeben: Die Parameter λ eines HMM und eine Beobachtungsfolge $O = (o_1, o_2, \ldots, o_T)$ der Dauer T
Gesucht: die a posteriori wahrscheinlichsten HMM-Zustände q_t^* und die a posteriori wahrscheinlichste Zustandsfolge q^* $P(O \mid \lambda)$

Die Folge $\hat{q}(o)$ von Zuständen
$\hat{q}_t(o) := argmax \quad P(q_t = j \mid o, \lambda)$ heißt *Maximum a posteriori* Zustandsfolge von o

wahrscheinlichster Pfad



Mögliche Zustandssequenzen als Pfade in einem *Trellis-Diagramm*: Der hervorgehobene Pfad repräsentiert die wahrscheinlichste Zustandssequenz Q^*

optimales Modell

Wie läßt sich λ bestimmen, wenn nur die Meßwerte bekannt sind?

Gegeben: die Struktur (N Zustände, M Symbole) eines HMM und eine Stichprobe ω von Beobachtungsfolgen o
Gesucht: das bestpassende Parameterfeld $\lambda = (\pi, A, B)$

$\Rightarrow$ Baum-Welsh-Algorithmus
$\Rightarrow$ Viterbi-Training (entscheidungsüberwachte Variante des Baum-Welsh-Alogrithmus, effizienter aber unzuverlässiger bei kleinen Lernstichproben)

Baum-Welch-Algorithmus

- **Idee:** Wenn die Zustände der Beobachtungen bekannt wären, ließen sich die Parameter durch Abzählen der Häufigkeiten in der Trainingsmenge schätzen
- Instanz des EM-Algorithmus
- Beginne mit zufälligen Parametern und iteriere zwei Schritte bis zur Konvergenz
  - Berechne die Zustände durch Forward-Backward-Algorithmus auf Grundlage des aktuellen Modells
  - Schätze die Parameter des Modells auf Grundlage berechneter Zustände

Anwendungsbereiche

- Spracherkennung (Sprachverarbeitung)
- Gestenerkennung in Videos (Bildverarbeitung)
- sensitive Suche in Datenbanken (Bioinformatik)
- DNA-Sequenzanalyse (Genetik)
- Reiz-Reaktions-Analysen (Neurologie)
- Messung bei korreliertem Rauschen
- Drift- und Volatilitätseinflüsse

Fragen?

Fragen?

Literatur

- Schukat-Talamazzini, E.G (1995): Automatische Spracherkennung, Vieweg Verlag
- Niemann, H. (1996): Mustererkennung 2, Friedrich-Alexander-Universität Erlangen-Nürnberg
- Nöth, E. (2004): Folien zur Vorlesung „Sprachverstehen“, Friedrich-Alexander-Universität Erlangen-Nürnberg
- Müller, S. (2001): Segmentierung und Klassifikation von Bildern und Bildsequenzen mit Hidden-Markov-Modellen, Gerhard-Mercator-Universität

Literatur Forts.

- Schukat-Talamazzini, E.G (2002): Folien zur Vorlesung „Mustererkennung“, Friedrich-Schiller-Universität Jena
- Haenelt, K. (2004): Hidden Markov-Modelle, Kursfolien

Diese Folien: http://www.pogo.franken.de/files/HMM.pdf
Druckgeeignet: http://www.pogo.franken.de/files/HMM.ps