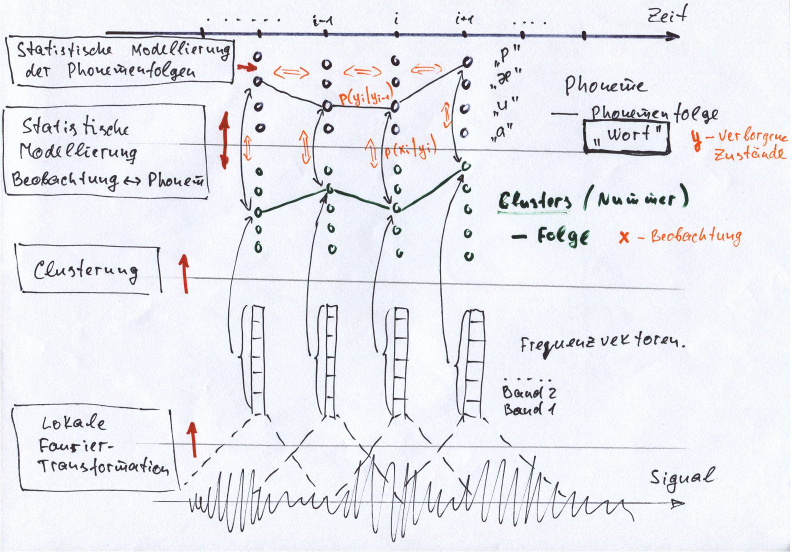
Strukturelle Modelle in der Bildverarbeitung
Markovsche Ketten I

D. Schlesinger – TUD/INF/KI/IS

- Anwendung: Sprachverarbeitung
- Wahrscheinlichkeitsmodell
- Unterschiedliche Parametrisierungen
- Berechnung der Partition Funktion



Zeit
i-1
i
i+1
Statistische Modellierung der Phonemenfolgen
„p"
„æ"
„u"
„a"
Phoneme
— Phonemenfolge
„Wort"
y – Verborgene Zustände
p(yi|yi-1)
p(xi|yi)
Statistische Modellierung Beobachtung ↔ Phonem
Clusters (Nummer)
— Folge
x – Beobachtung
Clusterung
Frequenzvektoren.
Band 2
Band 1
Lokale Fourier-Transformation
Signal

Anwendung: Sprachverarbeitung

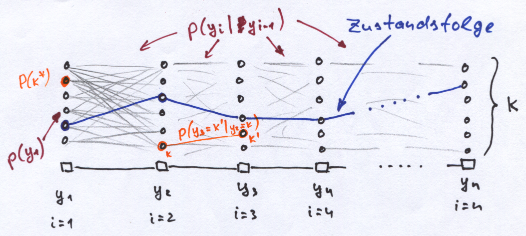
Signal $\Rightarrow$ Lokale Fouriertransformation $\Rightarrow$ Spektralvektoren
Eine Koordinate entspricht einem Frequenzbereich

Die Menge aller Frequenzvektoren ist (im Voraus) auf Clusters aufgeteilt
Frequenzvektoren (Kontinuierlich) werden durch Clusternummern (diskret) ersetzt
So entsteht **Beobachtung**

Die „Interpretation“ des Signals ist eine Phonemenfolge – **Zustandsfolge**

Es gibt einen statistischen Zusammenhang zwischen Beobachtungen und Phonemenfolgen
(Wahrscheinlichkeit der Beobachtung unter der Bedingung einer Phonemenfolge)

Es gibt ein statistisches Modell für Phonemenfolgen – **Markovsche Kette**



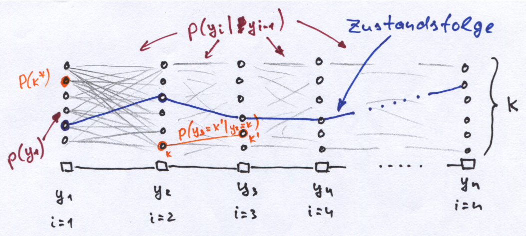
Sei $I = \{1, 2, \ldots, n\}$ die Menge der Zeitpunkte

Für jeden Zeitpunkt $i \in I$ gibt es eine **statistische Variable** y_i, die die Werte aus einer endlichen diskreten Menge K (**Zustandsmenge**, Labelmenge, Phonemenmenge etc.) annehmen kann, d.h. $y_i \in K$

$y = (y_1, y_2, \ldots, y_n)$ mit $y_i \in K$ ist eine **Zustandsfolge**

Die Menge aller Zustandsfolgen $\mathcal{Y} = K^n$ ist die Menge der elementaren Ereignisse

$p(y) = p(y_1, y_2, \ldots, y_n)$ ist die Wahrscheinlichkeit dafür



Zunächst allgemein: $p(y_1, y_2, \ldots, y_n) = p(y_1, y_2, \ldots, y_{n-1}) p(y_n | y_1, y_2, \ldots, y_{n-1})$

Markovsche Eigenschaft (etwas vereinfacht): $p(y_n | y_1, y_2, \ldots, y_{n-1}) = p(y_n | y_{n-1})$

Somit gilt:

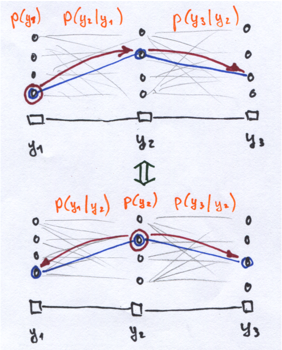
$$p(y) = p(y_1) \cdot \prod_{i=2}^{n} p(y_i | y_{i-1})$$

Die Parameter des Modells sind:

- Die Wahrscheinlichkeitsverteilung $p(y_1)$ der Zustände im ersten Zeitpunkt (Vektor aus $|K|$ Elementen)
- Die Übergangswahrscheinlichkeiten $p(y_i | y_{i-1})$ für alle Zeitpunkte $i = 2 \ldots n$ ($|K| \times |K|$ Matrizen)

Andere Parametrisierungen

Ein Beispiel:



$$
\begin{aligned}
p(y) = p(y_1, y_2, y_3) = \\
p(y_1)p(y_2|y_1)p(y_3|y_2) = \\
p(y_1)\frac{p(y_1|y_2)p(y_2)}{p(y_1)}p(y_3|y_2) = \\
p(y_1|y_2)p(y_2)p(y_3|y_2)
\end{aligned}
$$

Etwas allgemeiner:

$$
p(y) = p(y_1) \cdot \prod_{i=2}^{n} p(y_i|y_{i-1}) = \prod_{i=2}^{i^*} p(y_{i-1}|y_i) \cdot p(y_{i^*}) \cdot \prod_{i=i^*+1}^{n} p(y_i|y_{i-1})
$$

⇒ der „Anfang“ ist in der Mitte.

Andere Parametrisierungen

Noch anders:

$$p(y) = p(y_1) \cdot \prod_{i=2}^{n} p(y_i | y_{i-1}) = p(y_1) \cdot \prod_{i=2}^{n} \frac{p(y_i, y_{i-1})}{p(y_{i-1})} = \prod_{i=2}^{n} p(y_i, y_{i-1}) \cdot \prod_{i=2}^{n-1} p(y_i)^{-1}$$

$\Rightarrow$ Es gibt kein „Anfang“ und „Ende“ – die Parametrisierung ist symmetrisch.

Und noch etwas anders:

$$p(y) = \frac{1}{Z} \prod_{i=1}^{n} q(y_i) \cdot \prod_{i=2}^{n} g_i(y_{i-1}, y_i)$$

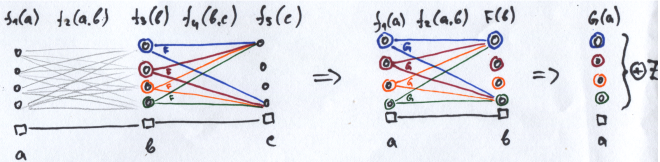
mit der **Partition Funktion**:

$$Z = \sum_{y \in \mathcal{Y}} \left[\prod_{i=1}^{n} q_i(y_i) \cdot \prod_{i=2}^{n} g_i(y_{i-1}, y_i) \right]$$

$q_i : K \to \mathbb{R}_+$ und $g_i : K \times K \to \mathbb{R}_+$ sind Parameter
– Funktionen, die keine Wahrscheinlichkeiten (explizit) repräsentieren

Berechnung der Partition Funktion

Beispiel: $f(a,b,c) = f_1(a) f_2(a,b) f_3(b) f_4(b,c) f_5(c)$



$$Z = \sum_{abc} f(a,b,c) = \sum_a \sum_b \sum_c \Big[f_1(a) f_2(a,b) f_3(b) f_4(b,c) f_5(c) \Big] =$$

$$= \sum_a f_1(a) \cdot \left[\sum_b f_2(a,b) f_3(b) \cdot \left[\sum_c f_4(b,c) f_5(c) \right] \right] =$$

$$\left| \text{bezeichne} \quad F(b) = f_3(b) \sum_c f_4(b,c) f_5(c) \right|$$

$$= \sum_a f_1(a) \cdot \left[\sum_b f_2(a,b) F(b) \right] =$$

$$\left| \text{bezeichne} \quad G(a) = f_1(a) \sum_b f_2(a,b) F(b) \right| = \sum_a G(a)$$

$$Z = \sum_{y \in \mathcal{Y}} \left[\prod_{i=1}^{n} q_i(y_i) \cdot \prod_{i=2}^{n} g_i(y_{i-1}, y_i) \right] =$$

$$= \sum_{y_n \in K} \sum_{y_{n-1} \in K} \dots \sum_{y_2 \in K} \sum_{y_1 \in K} \left[\prod_{i=1}^{n} q_i(y_i) \cdot \prod_{i=2}^{n} g_i(y_{i-1}, y_i) \right] =$$

$$= \sum_{y_n} q_n(y_n) \sum_{y_{n-1}} g_n(y_{n-1}, y_n) q_{n-1}(y_{n-1}) \dots$$

$$\dots \sum_{y_2} g_3(y_2, y_3) q_2(y_2) \sum_{y_1} g_2(y_1, y_2) q_1(y_1) =$$

$$= \sum_{y_n} q_n(y_n) \sum_{y_{n-1}} g_n(y_{n-1}, y_n) q_{n-1}(y_{n-1}) \dots \sum_{y_2} g_3(y_2, y_3) F_2(y_2) =$$

$$\left| \text{mit} \quad F_2(y_2) = q_2(y_2) \sum_{y_1} g_2(y_1, y_2) q_1(y_1) \right|$$

usw.

Berechnung der Partition Funktion

Algorithmus:

for ($k = 1 \dots K$) $F_1(k) = q_1(k)$
for ($i = 2 \dots n$)
  for ($k = 1 \dots K$)
    $F_i(k) = 0$
    for ($k' = 1 \dots K$)
      $F_i(k) = F_i(k) + F_{i-1}(k') g_i(k', k)$
    $F_i(k) = F_i(k) \cdot q_i(k)$
$Z = \sum_k F_n(k)$

Dynamische Programmierung im $(\mathbb{R}, +, \times)$-Semiring (Matrix-Multiplikation)

Funktionen $F_i(k)$ sind Bellmannsche Funktionen
– repräsentieren (wie bei der normalen Dynamischen Optimierung) partielle Lösungen.

Zeitkomplexität – $O(nK^2)$

Weitere Themen

- HMM (Hidden Markov Models)
- Erkennung stochastisches Automaten
- Wahrscheinlichste Folge
- Bayessche Entscheidungstheorie, Folge wahrscheinlichster Zustände
- Lernen nach dem Maximum-Likelihood (überwacht und unüberwacht)
- Bäume, Chou Aufgabe
- Diskriminatives Lernen