

Strukturelle Modelle in der Bildverarbeitung

Markovsche Ketten II

D. Schlesinger – TUD/INF/KI/IS

- Stationäre Verteilung
- Verborgene Markovsche Ketten (HMM)
- Erkennung stochastisches Automaten

Kontrahierende Abbildungen (Kontraktionen)

Sei M ein metrischer Raum, d.h. es existiert ein Abstand $d : M \times M \rightarrow \mathbb{R}$

Eine Abbildung $\varphi : M \rightarrow M$ heißt **kontrahierend**, wenn ein $\lambda < 1$ existiert, so dass

$$d(\varphi(x), \varphi(y)) \leq \lambda \cdot d(x, y)$$

für alle $x, y \in M$ erfüllt ist, d.h. die Abstände zwischen zwei beliebigen Elementen werden nach der Anwendung der Abbildung kleiner.

Spezialfall (relevant für Markovsche Ketten):

der Raum $M = \mathbb{R}^n$, der Abstand d ist der quadratische Abstand $d(x, y) = \|x - y\|^2$,

die Abbildung φ ist linear, d.h. durch eine $n \times n$ Matrix A angegeben: $\varphi(x) = A \cdot x$

Die Matrix A heißt kontrahierend, wenn

$$\|A \cdot (x - y)\|^2 < \lambda \cdot \|x - y\|^2$$

Fixpunktsatz von Banach:

Eine Kontraktion $\varphi : M \rightarrow M$ eines (nichtleeren) vollständigen metrischen Raumes M besitzt genau einen **Fixpunkt**, also einen Punkt $x^* \in M$ mit $\varphi(x^*) = x^*$.

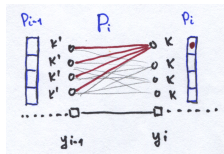
Für das obige Spezialfall heißt das $A \cdot x^* = x^*$,

d.h. x^* ist der Eigenvektor der Matrix A zum Eigenwert $\lambda = 1$.

Sei $x^0, x^1 \dots x^\infty$ eine Folge von Vektoren so, dass $x^i = \varphi(x^{i-1})$ mit einer kontrahierenden Abbildung φ . Dann $x^\infty = x^*$ unabhängig von x^0 – die Folge **konvergiert** zum Fixpunkt.

Übergangswahrscheinlichkeiten einer Markov Kette

Wie ergibt sich die Wahrscheinlichkeitsverteilung der Zustände im i -ten Zeitpunkt aus der WV im $i - 1$ -ten Zeitpunkt?



$$\begin{aligned} p(y_i = k) &= \sum_{k'} p(y_{i-1} = k', y_i = k) = \\ &= \sum_{k'} p(y_i = k | y_{i-1} = k') \cdot p(y_{i-1} = k') \end{aligned}$$

Sei $p_i = (p(y_i = 1), p(y_i = 2) \dots p(y_i = K))$ der Vektor der Zustandswahrscheinlichkeiten im i -ten Zeitpunkt und p_{i-1} analog.

Die Abbildung $p_{i-1} \mapsto p_i$ ist eine lineare Abbildung $p_i = P_i \cdot p_{i-1}$ mit der Matrix der Übergangswahrscheinlichkeiten $P_i(k, k') = p(y_i = k | y_{i-1} = k')$.

Ist die Matrix P_i kontrahierend? Ja, fast immer :-)) – eine beliebige **streng positive** Matrix der **Übergangswahrscheinlichkeiten** ist kontrahierend.

Beispiel: es gibt nur eine einzige nicht-kontrahierende 2×2 Matrix P_i (an der Tafel).

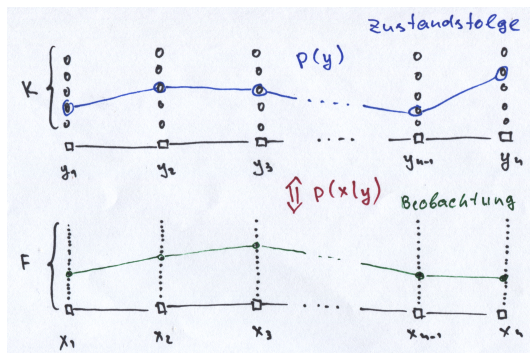
Man betrachte eine **unendliche** und **homogene** Markovsche Kette, d.h. $P_i = P$.

p_∞ (Verteilung der Zustände im Unendlichen) ist der Fixpunkt der Abbildung P und heißt **stationäre Verteilung** der Markovschen Kette.

p_∞ ist der Eigenvektor der P zum Eigenwert 1.

Matrixmultiplikation!!!

Verborgene Markovsche Ketten (HMM)



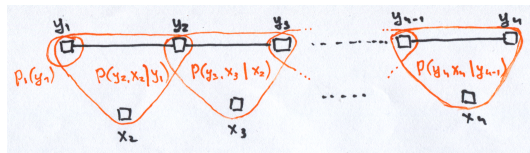
Es gibt zwei „Sorten“ der Variablen (beide sind Folgen):

$y = (y_1, y_2 \dots y_n)$, $y_i \in K$ (verborgene Variablen) und

$x = (x_1, x_2 \dots x_n)$, $x_i \in F$ (beobachtbare Variablen).

Ein Paar (x, y) ist ein elementares Ereignis.

Zwei Darstellungen für $p(x, y)$: „Zaun“ und „Kamm“ (Mealy- und Moore-Automaten).



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

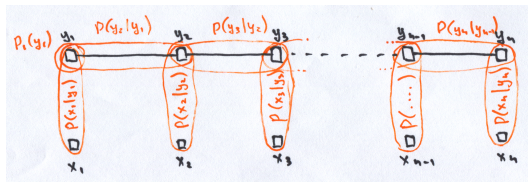
Eine Beobachtung x_i hängt explizit von den verborgenen Zuständen in i -ten und $i-1$ -ten Zeitpunkte ab (x_1 existiert nicht)

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

mit

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

Verborgene Markovsche Ketten, Kamm



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

Eine Beobachtung x_i hängt explizit nur vom verborgenen Zustand im i -ten Zeitpunkt ab

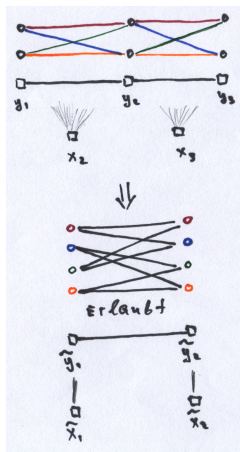
Einerseits ist es ein Spezialfall vom Zaun, d.h. wenn

$$p(x_i | y_i, y_{i-1}) \Rightarrow p(x_i | y_i)$$

Andererseits ...

Verborgene Markovsche Ketten

Andererseits kann jedes Zaun-Modell in ein äquivalentes Kamm-Modell überführt werden.



← Beispiel für drei Variablen und zwei Zustände

Allgemein:

- Man führt neue Variablen $\tilde{y}_1, \tilde{y}_2 \dots \tilde{y}_{n-1}$ ein, sie entsprechen Paaren $(y_1, y_2), (y_2, y_3) \dots (y_{n-1}, y_n)$
- Die neue Zustandsmenge $\tilde{K} = K \times K$ repräsentiert Zustandspaare des alten Modells

⇒ Jede Beobachtung \tilde{x}_i hängt explizit nur vom Zustand \tilde{y}_i

⇒ beide Darstellungen (Zaun und Kamm) sind äquivalent.

Erkennung stochastisches Automaten

Gegeben sei zwei Modelle, d.h.

$p^1(y_1)$, $p^1(y_i|y_{i-1})$ und $p^1(x_i|y_i)$ und dasselbe für das zweite Modell – $p^2(y_1) \dots$

Gegeben sei eine Beobachtung x , man entscheide, welches Modell sie generiert hat.

Einfache Vorgehensweise:

wähle das Modell, in dem die Wahrscheinlichkeit der Beobachtung maximal ist.

Die Aufgabe ist es diese Wahrscheinlichkeiten (d.h. $p^1(x)$ und $p^2(x)$) zu berechnen:

$$p(x) = \sum_y p(x, y) = \sum_y \left[p(y_1) \prod_{i=2}^n p(y_i|y_{i-1}) \prod_{i=1}^n p(x_i|y_i) \right]$$

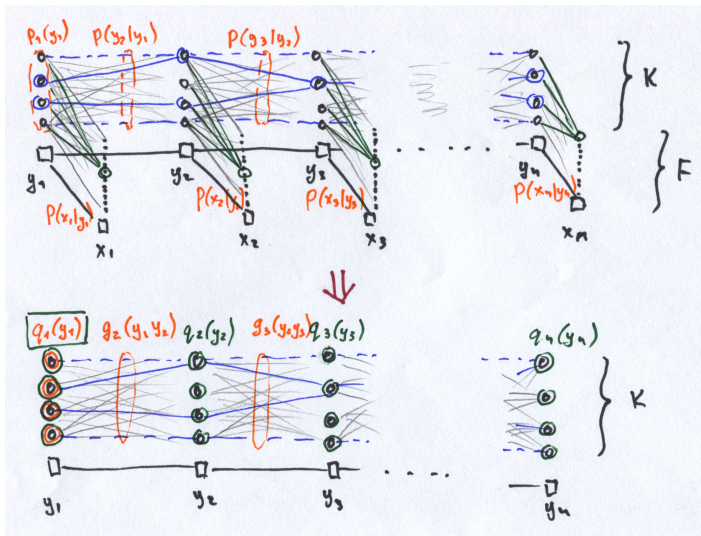
Die Berechnung erfolgt mit demselben Algorithmus, wie bei der Berechnung der Partition Funktion – die Funktionen q_i sind $p(x_i|y_i)$ und die Funktionen g_i sind $p(y_i|y_{i-1})$.

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

Die Bellmannsche Funktionen $F_i(k)$ haben die Bedeutung:

Wahrscheinlichkeit dafür, dass das Modell zum Zeitpunkt i im Zustand k war und dabei die Teilfolge $(x_1, x_2 \dots x_i)$ generiert hat.

Berechnung der Wahrscheinlichkeit der Beobachtung



Früher:

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

Heute:

- Stationäre Verteilung
- HMM (Hidden Markov Models)
- Erkennung stochastisches Automaten

Weitere Themen:

- Wahrscheinlichste Folge
- Bayessche Entscheidungstheorie, Folge wahrscheinlichster Zustände
- Lernen nach dem Maximum-Likelihood (überwacht und unüberwacht)
- Bäume, Chou Aufgabe
- Diskriminatives Lernen