Seminar: Informationsextraktion aus Texten

Informationsklassifikation mit Hidden Markov Model

Feng Wang

18.07.2003

Inhaltsverzeichnis

1 Einleitung **3**

2 Definition des HMM **3**

3 Bildliche Darstellungsmethoden **4**

3.1 Trellis 5

3.2 Automat 6

3.3 Baum 6

4 Einige wichtige Probleme beim HMM **8**

4.1 $P_{(O_3,S_3,O_2,S_2,O_1,S_1)} =?$ 8

4.2 $P_{(O_3,O_2,O_1)} =?$ 9

4.3 $P_{(S_3|O_3,O_2,O_1)} =?$ 10

4.4 $P_{(S_3,S_2,S_1|O_3,O_2,O_1)} =?$ 10

4.5 $(O_3 = b_1, O_2 = b_1, O_1 = b_1) \Longrightarrow \lambda = (I, II, III)?$ 11

5 Beschränkung des HMM **11**

6 Definition der Informationsklassifikation **12**

7 Anwendung des HMM bei der Informationsklassifikation **12**

8 Literatur **12**

1 Einleitung

Das Hidden Markov Model (HMM) ist schon sehr bekannt auf dem Gebiet der Spracherkennung. In der Systemtheorie heißt es auch "partially observed stochastic dynamical system model". Unabhängig davon, wie man es nennt, ist das HMM ein Bayessches Netz mit Zuständen und Maßen in einer diskreten Menge und in diskreter Zeit.

Das HMM wird häufig benutzt, um einen komplexen Prozess zu modellieren, mit dem man keine exakte analytische Beschreibung festlegen kann, entweder wegen seiner hohen Komplexität oder fehlender Information über sein Verhalten.

In den letzten Jahren hat man versucht, das HMM auf die Informationsklassifikation anzuwenden.

Ich erkläre zuerst, wie das HMM aufgebaut ist und wie einige wichtige Aufgaben berechnet werden. Dann stelle ich die Informationsklassifikation kurz vor und wie das HMM auf dieses Gebiet angewandt wird.

In den folgenden Abschnitten sollen die Grundlagen des HMM in einfacher Form präsentiert werden. Dafür wird aus Gründen der Übersichtlichkeit auf die Verwendung zuvieler stochastischen Formeln verzichtet. Die entsprechenden Beweise sind in der am Ende aufgeführten Literaturliste zu finden.

2 Definition des HMM

Hier betrachten wir zuerst die mathematische Definition von HMM, dann betrachten wir im nächsten Abschnitt seine verschiedenen Darstellungsmethoden.

Um ein HMM darzustellen, braucht man nur ein 3-stelliges Tupel:

$\lambda = (I, II, III)$:

I: eine Markov-Kette mit endlichen Zuständen $\{a_i\}_{1 \leq i \leq M}$ und ihren Übergangswahrscheinlichkeiten $(p_{ij})_{1 \leq i,j \leq M}$, $p_{ij} = P_{(S_t = a_j | S_{t-1} = ai)}$;

II: ein Vektor $(q_i)_{1 \leq i \leq M}$, die Anfangswahrscheinlichkeiten der Zustände in der Markov-Kette, $q_i = P_{(S_0 = a_i)}$;

III: eine endliche Menge von Beobachtungswerten $\{b_i\}_{1 \leq i \leq N}$ und ihren Beobachtungswahrscheinlichkeiten $(r_{ij})_{1 \leq i \leq M, 1 \leq j \leq N}$, $r_{ij} = P_{(O_t = b_j | S_t = a_i)}$.

Hier ist ein Beispiel, das wir später immer wieder benutzen werden:

$\lambda = (I, II, III)$:

I: eine Markov-Kette mit zwei Zuständen $\{a_1, a_2\}$ und ihren Übergangswahrscheinlichkeiten $(p_{ij})_{1 \leq i,j \leq 2}$ von Schritt $t-1$ nach Schritt t;

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} P_{(S_t=a_1|S_{t-1}=a_1)} & P_{(S_t=a_2|S_{t-1}=a_1)} \\ P_{(S_t=a_1|S_{t-1}=a_2)} & P_{(S_t=a_2|S_{t-1}=a_2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

p_{21} bedeutet: wenn die Markov-Kette im Zustand a_2 in Schritt $t-1$ ist, nimmt sie mit der Wahrscheinlichkeit $p_{21} = \frac{1}{2}$ im nächsten Schritt den Zustand a_1 an. S_t steht für die Zufallsvariable dieser Markov-Kette bei dem Schritt t.

II: die Anfangswahrscheinlichkeiten der Zustände $\{a_1, a_2\}$:

$(q_1, q_2) = (P_{(S0=a_1)}, P_{(S0=a_2)}) = (\frac{1}{2}, \frac{1}{2})$;

Um die Markov-Kette zu realisieren, braucht man einen Anfangspunkt als ersten Schritt für die Rechnung. Dafür muß man für jeden Zustand einen Wahrscheinlichkeitswert angeben. Das ist die so genannte Anfangswahrscheinlichkeit. In diesem Beispiel kann die Markov-Kette mit der Wahrscheinlichkeit $\frac{1}{2}$ entweder von Zustand a_1 oder a_2 anfangen.

III: eine Menge von Beobachtungswerten $\{b_1, b_2\}$ und ihren Beobachtungswahrscheinlichkeiten $(r_{ij})_{1 \leq i \leq 2, 1 \leq j \leq 2}$

$$\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} = \begin{pmatrix} P_{(Ot=b_1|St=a_1)} & P_{(Ot=b_2|St=a_1)} \\ P_{(Ot=b_1|St=a_2)} & P_{(Ot=b_2|St=a_2)} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

Wenn die Zufallsvariable der Markov-Kette einen Zustand annimmt, gibt die Markov-Kette in dem Zustand einen Beobachtungswert aus, und zwar mit einer bestimmten Wahrscheinlichkeit. Z.B. ist die Markov-Kette im Zustand a_2 bei dem Schritt t, gibt sie mit Wahrscheinlichkeit $\frac{1}{2}$ b_1 aus. O_t steht für die Zufallsvariable der Beobachtung bei dem Schritt t.

3 Bildliche Darstellungsmethoden

Um diese formal mathematische Definition zu veranschaulichen, hat man einige bildliche Darstellungsmethoden entwickelt.

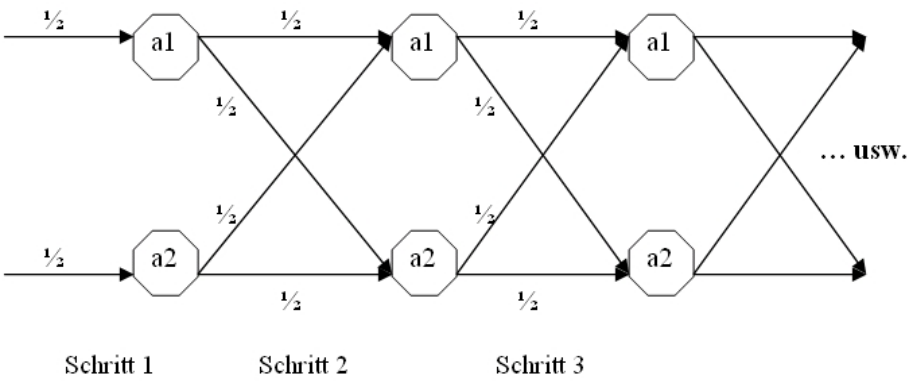

Abbildung 1: Trellis-Darstellung

3.1 Trellis

Trellis ist eine schrittweise Darstellungsmethode für die Markov-Kette, ohne die dritte Stelle III im HMM λ darzustellen. Hier nehme ich das HMM λ (Abb.1: Trellis-Darstellung) als Beispiel.

Schritt 1: Mit der Anfangswahrscheinlichkeit $(q_1, q_2) = (\frac{1}{2}, \frac{1}{2})$ nimmt die Markov-Kette die entsprechenden Zustände a_1 oder a_2 bei Schritt 1 an. Sei $S_1 = a_1$. Dann gibt die Markov-Kette einen der beiden Beobachtungswerte b_1, b_2 mit dem entsprechenden Wahrscheinlichkeitswert aus. Z.B., wenn $O_1 = b_1$, dann gibt die Kette b_1 aus, mit der Beobachtungswahrscheinlichkeit $r_{11} = (O_1 = b_1|S_1 = a_1)) = \frac{1}{2}$.

Schritt 2: Bei Schritt 2 kann die Zufallsvariable a_1 oder a_2 als Wert annehmen mit dem entsprechenden Wahrscheinlichkeitswert $p_{11} = P_{(S_2=a_1|S_1} = a_1 = \frac{1}{2}$, $p_{12} = P_{(S_2=a_2|S_1} = a_1 = \frac{1}{2}$. Sei $S_2 = a_1$. Dann gibt die Markov-Kette einen der beiden Beobachtungswerten b_1 oder b_2 mit dem entsprechenden Wahrscheinlichkeitswert aus. Z.B., wenn $O_2 = b_1$, gibt die Kette b_1 aus mit der Beobachtungswahrscheinlichkeit $r_{11} = (O_2 = b_1|S_2 = a_1) = \frac{1}{2}$.

Schritt 3: Obige Operation wird wiederholt.

Das HMM wird so schrittweise weiter laufen.

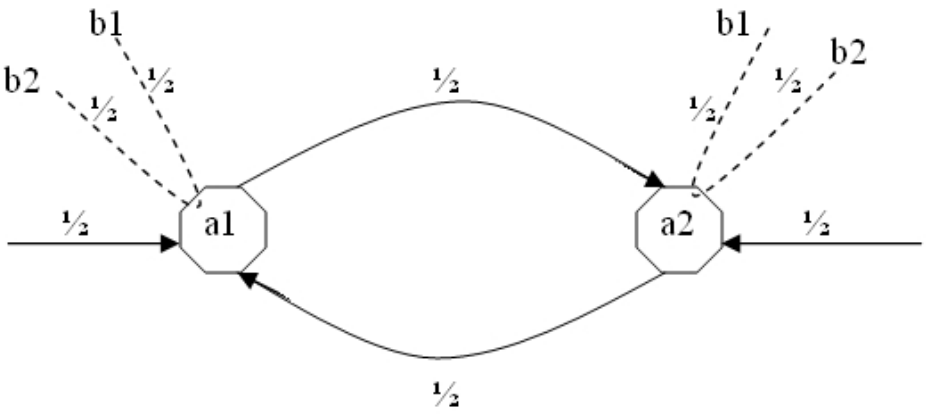

Abbildung 2: Automat-Darstellung

3.2 Automat

Das HMM entspricht einem stochastischen Automaten mit endlichen Zuständen. Daher kann man das HMM auch mit einem Automaten darstellen (Abb.2: Automat-Darstellung).

Mit der Anfangswahrscheinlichkeit $(q_1, q_2) = (\frac{1}{2}, \frac{1}{2})$ fängt der Automat an zu laufen. Bei jedem Zustandswechsel gibt der Automat einen Wert aus, und zwar mit dem entsprechenden Beobachtungswahrscheinlichkeitswert.

Bei dieser Methode kann man die Beobachtungswerte und ihre Beobachtungswahrscheinlichkeitswerte auch weglassen.

3.3 Baum

Weil das HMM ein spezielles Bayessches Netz ist, und zwar in Baumform, kann ich wie in Abb.3 (Baum-Darstellung) das HMM $\lambda = (I, II, III)$ mit einem Baum darstellen.

Zuerst zeichne ich mit einer durchgezogenen Linie eine Markov-Kette (oder einen so genannten Markov-Raum). Die Kette lesen wir von oben nach unten. Jedem Knoten ist ein Wahrscheinlichkeitsraum (Übergangsraum) zugeteilt. Eine Zufallsvariable S_i mit der Wahrscheinlichkeit $P_{(S_i)}$ ist in diesem Raum definiert. Die Variable kann a_1 oder a_2 als Wert annehmen, den wir den Zustand der Markov-Kette nennen.

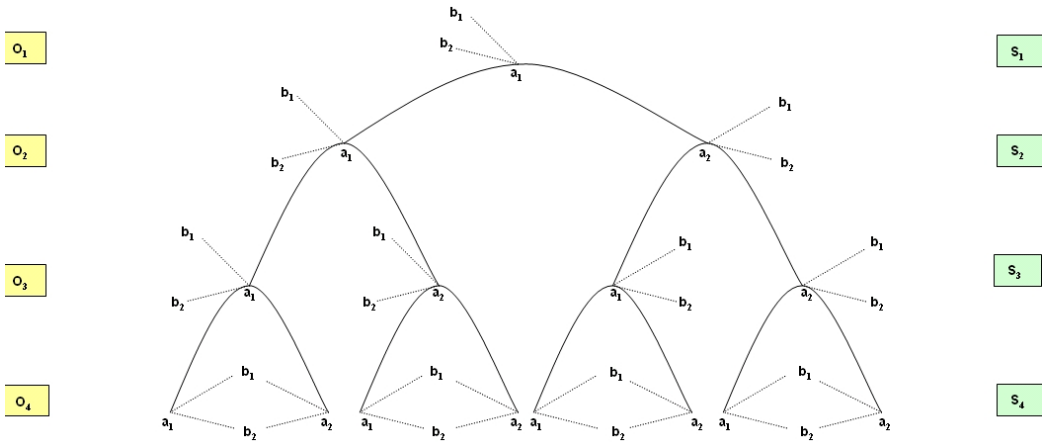

Abbildung 3: Baum-Darstellung

Jedem Knoten ist ein Beobachtungsraum zugeteilt, gezeichnet mit gestrichelten Linien. Eine Zufallsvariable O_i mit der Wahrscheinlichkeit $P_{(O_i)}$ ist in diesem Raum definiert. Die Variable kann b_1 oder b_2 als Wert annehmen, den wir Beobachtungswert nennen. Die Verteilungen in den Beobachtungsräumen können unterschiedlich sein.

Zusammen haben alle Übergangsräume und Beobachtungsräume einen gemeinsamen Wahrscheinlichkeitsraum aufgebaut. Diser ist der Wahrscheinlichkeitsraum des HMM. Eine Zufallsvariablenfolge $(S_i, O_i, S_{i-1}, O_{i-1}, ...S_1, O_1)$ mit der Wahrscheinlichkeit $P_{(S_i,O_i,S_{i-1},O_{i-1},...S_1,O_1)}$ ist in diesem Raum definiert.

Jetzt müssen wir als Beweisgrundlage für den nächsten Abschnitt die drei Beziehungen zwischen den Räumen feststellen.

Alle Beobachtungsräume sind voneinander unabhängig (independent). Die Wahrscheinlichkeit der Beobachtungswerten aus den verschiedenen Räumen kann durch Multiplikation ihrer Wahrscheinlichkeiten berechnet werden, wenn sie sich aus zwei Beobachtungen ergibt:
$P_{(O_i,Oj|Si,Sj)} = P_{(O_i|S_i)}P_{(Oj|Sj)}, i \neq j$

Ein Beobachtungsraum ist von dem Übergangsraum, auf dem er nicht steht, unabhängig. Die Wahrscheinlichkeit eines Tupels aus einem Beobachtungswert und einem Zustand kann man daher durch Multiplikation ihrer Wahrscheinlichkeiten in ihren entsprechenden Räumen berechnen, wenn der

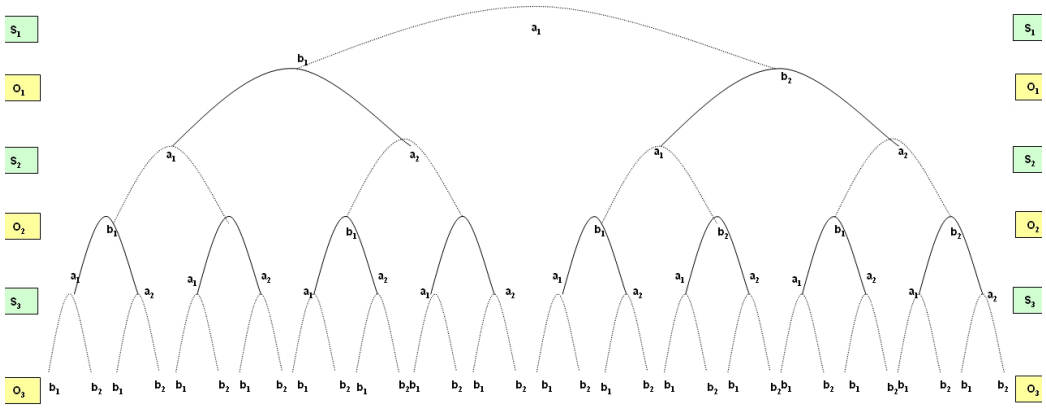

Abbildung 4: äquivalente Baum-Darstellung

Beobachtungswert nicht auf dem Zustand aufgebaut ist:
$P_{(O_i|S_i,S_j)} = P_{(O_i|S_i)}P_{(S_j)}, i \neq j$
.

Der Beobachtungsraum ist von dem Übergangsraum, auf dem er steht, abhängig. So muss man wie folgt die Wahrscheinlichkeit berechnen
$P_{(O_i,S_i)} = P_{(O_i|S_i)}P_{(S_i)}$ (Bedingungswahrscheinlichkeit).

Unter diesen Voraussetzungen ist die Baumstruktur äquivalent zu dieser Zeichnung in Abb.4: äquivalente Baum-Darstellung.

Trellis, Automat, Baum und dessen äquivalente Form sind alle äquivalent zueinander. Aber die äquivalente Form des Baums ist anschaulicher als die anderen. Anhand dieses Baums kann ich alle Rechnungen auf die Multiplikations- und die Additionsregel in der Wahrscheinlichkeitstheorie reduzieren.

4 Einige wichtige Probleme beim HMM

Um den Beweis kürzer zu machen, nehmen wir an, die Versuchslänge ist 3.

4.1 $P_{(O_3,S_3,O_2,S_2,O_1,S_1)} =?$

Wahrscheinlichkeit einer Beobachtungs- und Zustandsfolge, $(O_3, S_3, O_2, S_2, O_1, S_1)$

Z.B. eine Folge mit allen Gliedern von a_1 und b_1. So eine Folge stellt einen Pfad im HMM Baum dar (Abb.4). Man kann einfach mit der Multiplikationsregel die Wahrscheinlichkeit der Folge feststellen. Oder wenn man

es formal ausrechnen möchte, sind hier die Begründungen für jeden Schritt. Die Begründung basiert auf unseren Annahmen.

$$
\begin{aligned}
& P_{(O_3=b_1,S_3=a_1,O_2=b_1,S_2=a_1,O_1=b_1,S_1=a_1)} \\
&= P_{(O_3=b_1,O_2=b_1,O_1=b_1,S_3=a_1S_2=a_1,S_1=a_1)} \quad \textit{Bedingungswahrscheinlichkeit} \\
&= P_{(O_3=b_1,O_2=b_1,O_1=b_1|S_3=a_1,S_2=a_1,S_1=a_1)}P_{(S_3=a_1S_2=a_1,S_1=a_1)} \quad \textit{Unabhängigkeit} \\
&= P_{(O_3=b_1|S_3=a_1)}P_{(O_2=b_1|S_2=a_1)}P_{(O_1=b_1|S_1=a_1)}P_{(S_3=a_1S_2=a_1,S_1=a_1)} \quad \textit{Markov-Kette} \\
&= P_{(O_3=b_1|S_3=a_1)}P_{(O_2=b_1|S_2=a_1)}P_{(O_1=b_1|S_1=a_1)}P_{(S_3=a_1|S_2=a_1,S_1=a_1)}P_{(S_2=a_1,S_1=a_1)} \\
&= P_{(O_3=b_1|S_3=a_1)}P_{(O_2=b_1|S_2=a_1)}P_{(O_1=b_1|S_1=a_1)}P_{(S_3=a_1|S_2=a_1,S_1=a_1)}P_{(S_2=a_1|S_1=a_1)}P_{(S_1=a_1)} \\
&\textit{Markov-Kette} \\
&= P_{(O_3=b_1|S_3=a_1)}P_{(O_2=b_1|S_2=a_1)}P_{(O_1=b_1|S_1=a_1)}P_{(S_3=a_1|S_2=a_1)}P_{(S_2=a_1|S_1=a_1)}P_{(S_1=a_1)} \\
&= P_{(O_3=b_1|S_3=a_1)}P_{(S_3=a_1|S_2=a_1)}P_{(O_2=b_1|S_2=a_1)}P_{(S_2=a_1|S_1=a_1)}P_{(O_1=b_1|S_1=a_1)}P_{(S_1=a_1)}
\end{aligned}
$$

Man kann auch durch das Ergebnis feststellen, daß die Folge einen Pfad in diesem Baum definiert hat, nämlich $(b_1, a_1, b_1, a_1, b_1, a_1)$.

4.2 $P_{(O_3,O_2,O_1)} =?$

Wahrscheinlichkeit einer Folge von Beobachtungswerten

So eine Folge stellt alle Pfade dar, die die Folge erzeugen können. Die Wahrscheinlichkeit dafür ist die Summe der Wahrscheinlichkeiten aller Pfade, die die Folgen erzeugen können.

Z.B. die Beobachtungsfolge (b_1, b_1, b_1). Es gibt acht Kombinationen von Zuständen, die diese Beobachtungsfolge erzeugen können. Den Summand kann man mit der Methode berechnen, die das Problem (1) zeigt. Dann benutzt man die Additionsregel, um die gesuchte Wahrscheinlichkeit zu finden.

$$
\begin{aligned}
& P_{(O_3=b_1,O_2=b_1,O_1=b_1)} \\
&= P_{(O_3=b_1,S_3=a_1,O_2=b_1,S_2=a_1,O_1=b_1,S_1=a_1)} \\
&+P_{(O_3=b_1,S_3=a_2,O_2=b_1,S_2=a_1,O_1=b_1,S_1=a_1)} \\
&+P_{(O_3=b_1,S_3=a_1,O_2=b_1,S_2=a_2,O_1=b_1,S_1=a_1)} \\
&+P_{(O_3=b_1,S_3=a_1,O_2=b_1,S_2=a_1,O_1=b_1,S_1=a_2)} \\
&+P_{(O_3=b_1,S_3=a_2,O_2=b_1,S_2=a_2,O_1=b_1,S_1=a_1)} \\
&+P_{(O_3=b_1,S_3=a_2,O_2=b_1,S_2=a_1,O_1=b_1,S_1=a_2)} \\
&+P_{(O_3=b_1,S_3=a_1,O_2=b_1,S_2=a_2,O_1=b_1,S_1=a_2)} \\
&+P_{(O_3=b_1,S_3=a_2,O_2=b_1,S_2=a_2,O_1=b_1,S_1=a_2)}
\end{aligned}
$$

Bei der Programmierung benutzt man häufig den forward-backward Algorithmus mit geringerer Komplexität.

4.3 $P_{(S_3|O_3,O_2,O_1)} = ?$

Wenn man eine Folge von Beobachtungswerten weiß, möchte man den letzten Zustand in der Markov-Kette feststellen. Das ist die Wahrscheinlichkeit aller Pfade, die die Beobachtungsfolge erzeugen können und S_3 als letzten Zustand haben.

Z.B. die Beobachtungsfolge ist 3-mal b_1, wie hoch ist die Wahrscheinlichkeit, daß der letzten Zustand a_1 ist? Zuerst kann man mit der Bedingungswahrscheinlichkeit die Frage umschreiben. Den Nenner, $P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$, können wir schon berechnen.

Dann benutzen wir die Unabhängigkeit. Damit ist der Zähler, $P_{(S_3=a_1,O_3=b_1,O_2=b_1,O_1=b_1)}$, zerlegt. $P_{(O_2=b_1,O_1=b_1)}$ kennen wir schon und $P_{(O_3=b_1|S_3=a_1)}$ ist vorgegeben (Beobachtungswahrscheinlichkeit). Für $P_{(S_3=a_1)}$ wenden wir wieder die Additionsregel an. Es steht für vier Kombinationen von Zuständen.

$P_{(S_3=a_1|O_3=b_1,O_2=b_1,O_1=b_1)}$ *Bedingungswahrscheinlichkeit/Bayes*
$= P_{(S_3=a_1,O_3=b_1,O_2=b_1,O_1=b_1)}/P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$ *Unabhängigkeit*
$= P_{(S_3=a_1,O_3=b_1)}P_{(O_2=b_1,O_1=b_1)}/P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$ *Bedingungswahrscheinlichkeit/Bayes*
$= P_{(O_3=b_1|S_3=a_1)}P_{(S_3=a_1)}P_{(O_2=b_1,O_1=b_1)}/P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$
$= P_{(O_3=b_1|S_3=a_1)} \cdot [P_{(S_3=a_1,S_2=a_1,S_1=a_1)} + P_{(S_3=a_1,S_2=a_2,S_1=a_1)} + P_{(S_3=a_1,S_2=a_1,S_1=a_2)} + P_{(S_3=a_1,S_2=a_2,S_1=a_2)}] \cdot P_{(O_2=b_1,O_1=b_1)}/P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$

Auf dieser Basis kann man anhand einer Beobachtungsfolge vorhersagen, welcher Zustand auftauchen kann. Dafür benutzt man die maximale Erwartung.

Zuerst berechnet man die Wahrscheinlichkeit mit verschiedenen letzten Zuständen.

Hier $p_1 = P_{(S_3=a_1|O_3=b_1,O_2=b_1,O_1=b_1)}$ und $p_2 = P_{(S_3=a_2|O_3=b_1,O_2=b_1,O_1=b_1)}$.

Die größte Wahrscheinlichkeit steht für den wahrscheinlichsten letzten Zustand: $S_3 = \max_{1 \leq i \leq 2}(p_i)$.

4.4 $P_{(S_3,S_2,S_1|O_3,O_2,O_1)} = ?$

Wenn man eine Folge von Beobachtungswerten weiß, möchte man die wahrscheinlichste Zustandsfolge feststellen, die die Beobachtungsfolge erzeugt hat.
Die gesuchte Zustandsfolge stellt den Pfad dar, der mit maximaler Wahrscheinlichkeit die Beobachtungsfolge erzeugen kann.

Zuerst berechnet man die Wahrscheinlichkeit aller wahrscheinlichste Zustandsfolgen wie folgt:

$p_1 = P_{(S_3=a_1,S_2=a_1,S_1=a_1|O_3=b_1,O_2=b_1,O_1=b_1)}$ *Bedingungswahrscheinlichkeit*
$= P_{(S_3=a_1,S_2=a_1,S_1=a_1,O_3=b_1,O_2=b_1,O_1=b_1)}/P_{(O_3=b_1,O_2=b_1,O_1=b_1)}$;
$p_2 = P_{(S_3=a_2,S_2=a_1,S_1=a_1|O_3=b_1,O_2=b_1,O_1=b_1)}$;
$p_3 = P_{(S_3=a_1,S_2=a_2,S_1=a_1|O_3=b_1,O_2=b_1,O_1=b_1)}$;
$p_4; p_5; p_6; p_7; p_8$

Dann benutzt man die maximale Erwartung um sie festzustellen. Die Folge, die die größte Wahrscheinlichkeit hat, ist die gesuchte:

$$(a_j)_{1\leq j\leq 3} = \max_i(P_i), i = 1, 2, 3, 4, 5, 6, 7, 8$$

Wenn man direkt die Wahrscheinlichkeit berechnet, hat der Algorithmus die Komplexität $O(M^T)$. M ist die Anzahl der Zustände, T die Anzahl der Beobachtungen. Um diese hohe Komplexität zu vermeiden, wird der Viterbi-Algorithmus benutzt. Seine Komplexität beträgt $O[(M^2)T]$. Der Viterbi-Algorithmus ist eine Anwendung von "dynamic programming". Er funktioniert wie der "kürzeste Weg"-Algorithmus.

4.5 $(O_3 = b_1, O_2 = b_1, O_1 = b_1) \Longrightarrow \lambda = (I, II, III)$?

Wenn man eine Folge von Beobachtungswerten weiß, möchte man das Tupel $\lambda = (I, II, III)$ feststellen.

Das ist eine Optimierungsfrage mit Bedingung.
Lösung: Baum-Welch (schrittweise Annäherung)
$S_1 : \lambda_1$
$S_2 : \lambda_2 = f(O_1, \lambda_1)$
$S_3 : \lambda_3 = f(O_2, \lambda_2)$
usw.

Auf der Basis einer Beobachtungsfolge und eines vorher festgelegten λ berechnet man ein neues λ, dann auf der Basis von einer neuen Beobachtungsfolge und dem neuen λ berechnet man wieder ein neues Lambda. Und so weiter.

5 Beschränkung des HMM

Bis hier hat man alles unter 3 Annahmen beobachtet:

1) eine Markov-Kette erster Ordnung:

$$P_{(X_i|X_{i-1},\ldots,X_1)} = P_{(X_i|X_{i-1})}$$

2) der Prozess ist zeit-invariant (constant):

$$P_{(X_i|X_{i-1})} = P_{(X_k|X_{k-1})}$$

3) Unabhängigkeit zwischen Beobachtungen:

$$P_{(O_i,\ldots,O_1|X_i,\ldots,X_1)} = \prod_t P_{(O_t|X_t)}$$

Daher paßt das HMM nicht unbedingt zu der realen Welt. Bei der Anwendung muß es durch andere Methoden ergänzt werden.

6 Definition der Informationsklassifikation

Für den Begriff Informationsklassifikation gibt es zwei unterschiedliche Definitionen:
a) Zuordnung eines Textes zu einer bestimmten Textkategorie;
b) Erkennung der Glieder eines Textes.

zu a): gegeben ist ein Text. Man möchte durch die Reihenfolge der Schlüsselwörter festlegen, welche Textsorte vorliegt, z.B. eine mathematische Arbeit, Email oder Anleitung.

zu b): gegeben ist ein Text, man möchte durch die Reihenfolge der Schlüsselwörter die Glieder des Textes erkennen. Liegt z.B. ein Email vor, möchte man Subjekt, Adresse oder Anhang erkennen. Danach kann man aus den Glieder die gesuchte Information extrahieren.

7 Anwendung des HMM bei der Informationsklassifikation

Wenn man das HMM auf die Informationsklassifikation anwendet, sind die Schlüsselwörter die Beobachtungswerte.

Nach Definition a) sind die Kategorien die Zustände.

Nach Definition b) sind die verschiedenen Glieder in einem Text die Zustände.

So kann man die vorher entwickelten Algorithmen im HMM anwenden.

8 Literatur

R.J.Elliott/ L.Aggoun/J.B. Moore, Hidden Markov Models. Estimation and Control, New York 1995.