Konzeptionelle Beschreibung des Corticalen Algorithmus und seine Verwendung in der Automatischen SprachVerarbeitung

Ronald Römer, Tobias Herbig

Harman-Becker Automotive Systems GmbH
Ronald.Roemer@harman.com; Tobias.Herbig@harman.com

Abstract: Im vorliegenden Beitrag wird das Konzept für einen corticalen Algorithmus vorgestellt, der auf der HMM-Technologie basiert. Dabei wird zunächst eine verallgemeinerte Darstellung eines HMM's verwendet, um einen einschichtigen corticalen Algorithmus zu beschreiben. Dabei zeigt es sich, dass die Dynamik eines HMM durch verallgemeinerte Kalmangleichungen beschrieben werden kann. Unter Verwendung dieser Gleichungen gelingt es, die Dynamik eines mehrschichtigen bidirektionalen HMM's als corticalen Algorithmus zu formulieren. Konventionelle HMM's erscheinen dann lediglich als unidirektionaler Sonderfall des corticalen Algorithmus.

1 Einführung

Die Automatische Sprachverarbeitung lässt sich in verschiedene hierarchisch organisierte Abstraktionsebenen einteilen, z.B. können auf einem sehr hohen Abstraktionsniveau Dialog und Sprecher modelliert werden, weiter unten in der Hierarchie finden sich dagegen die linguistische- und die akustische Ebene wieder. Bei der Lösung der jeweiligen Probleme in diesen Ebenen kommen eine Vielzahl von lokalen Ansätzen zur Anwendung, welche nur selten einen Informationsaustausch zwischen den verschiedenen Abstraktionsebenen berücksichtigen [1]. Ein typisches Problem, welches über alle Abstraktionsebenen hinweg gemeinsam auftritt, besteht darin, aktuell sinntragende Symbolfolgen der jeweiligen Schicht zu identifizieren. Von Interesse sind dabei z.B. Dialogschrittfolgen, Sprecherfolgen, Wortfolgen oder auch Phonem- bzw. Modellfolgen. Da Kommunikation auf der Sinnhaftigkeit der auszutauschenden Information basiert, schränken die identifizierten Symbolfolgen der höheren Abstraktionsebenen die Anzahl der möglichen Symbolfolgen der unteren Abstraktionsfolgen deutlich ein.

Mit der Sprachmodellierung durch die ***H**idden **M**arkov **M**odel* (HMM)-Technologie liegt ein mächtiges Werkzeug vor, mit dem die Symbolfolgen der verschiedenen Abstraktionsebenen verarbeitet werden können. Daher ist es ein ebenso erstrebenswertes wie ambitionierte Ziel, unter Verwendung der HMM-Technologie eine vereinheitlichte Sicht auf die verschiedenen Abstraktionsebenen der automatischen Sprachverarbeitung zu gewinnen. Die Motivation zur Lösung dieser Aufgabe hat ihre Wurzeln in der biologischen Informationsverarbeitung. Die Idee, dass die Auswahl der möglichen Symbolfolgen in einer tiefer liegenden Schicht von einer höheren Schicht gesteuert wird und andererseits die untere Schicht die höhere Schicht durch aktuelle Messungen beeinflusst, liegt dem sogenannten corticalen Algorithmus zugrunde. Dieser Algorithmus ist auf ein theoretisches Modell in [2] zurückzuführen, wonach die neuronalen Signale in den verschiedenen Sinnesregionen der Großhirnrinde (Neocortex) grundsätzlich nach dem gleichen Prinzip verarbeitet werden. Im Kern beschreibt dieses Modell den bidirektionalen Informationsfluss zwischen den hierarchisch organisierten Schichten des Neocortex, wobei afferente (aufsteigende) und efferente (absteigende) Informationen in jeder Schicht derart fusioniert werden, dass Konflikte oder Mehrdeutigkeiten aufgelöst werden können. In jüngeren Arbeiten wurden nun einige Vorschläge unterbreitet, wie ein solcher Algorithmus implementiert werden könnte. Häufig wurde dabei auf den ***B**elief-**P**ropagation-**A**lgorithmus* (BPA) zurückgegriffen, welcher auf graphbasierten Modellen basiert, die auch als Bayes'sche Netze bekannt geworden sind [3]. In der Literatur werden hinsichtlich ihrer Struktur zwei Arten von BPA unterschieden, den kettenförmig und den baumartig strukturierten BPA [4]. Konkrete Anwendungen von BPA beider Arten sind bspw. aus der Bildverarbeitung bekannt. Dagegen ist die

Formulierung des corticalen Algorithmus für die Sprachverarbeitung unter Verwendung der HMM-Technologie noch unbekannt. Mit dem diesjährigen Beitrag wird daher das Ziel verfolgt, in einem ersten Schritt den corticalen Algorithmus für kettenförmig strukturierte HMM's zu formulieren.

2 Biologisch motivierte Informationsverarbeitung

Zu den Organen der Wahrnehmung werden vor allem die Sinnesorgane (Auge, Ohr usw.) gezählt, zweifellos gehören aber auch diejenigen Teile des Gehirns dazu, die Signale von diesen Sinnesorganen erhalten. Den Zugang für solche Signale bildet das sogenannte thalamo-corticale System bestehend aus dem Thalamus, der dünnen äußeren Hüllschicht des Gehirns (Neocortex, Großhirnrinde) und den reziproken thalamo-corticalen Leitungsbahnen. Der strukturelle Aufbau des Neocortex kann folgendermaßen beschrieben werden.

Die Großhirnrinde besteht aus sechs Ebenen. Die einzelnen Regionen der sechs übereinander liegenden Ebenen lassen sich in sogenannte "Säulen" aus übereinander liegenden Zellen einteilen, die durch Axone miteinander verbunden sind. Diese Säulen können als Basis-Einheiten des Neocortex verstanden werden. Die Einteilung des Neocortex in verschiedene Regionen ist hierarchisch organisiert. Dabei findet der Informationsfluss bidirektional statt, d.h. sowohl von niedrigeren Regionen zu den höheren als auch umgekehrt (Feedback). Unter der Voraussetzung, dass das Gehirn über ein dynamisches Modell der Realität und demnach über Kontextwissen verfügt, kann es in jeder Schicht Vorhersagen über die zu erwartenden Sinnesempfindungen treffen und diese ständig mit den tatsächlichen Sinnesempfindungen vergleichen. Das Vorgehen des Neocortex kann dabei in drei Schritte unterteilt werden.

Bottom-Up. Hier wird an jede nächst höhere Ebene nur diejenige sensorische Information weitergereicht, welche durch Abstraktion entsteht. Die afferente Reduktion der eingehenden Information ermöglicht damit die Bildung von Kontextwissen auf den höheren Ebenen.

Top-Down. Gleichzeitig findet ein efferentes Feedback von der obersten Ebene zu den unteren Ebenen statt. Durch das weiterreichende Kontextwissen der oberen Ebenen kann eine Prädiktion für den nächsten Zeitpunkt an die jeweils darunter liegende Ebene weitergereicht werden.

Konflikte. Mit dem Vorliegen afferenter und efferenter Information kann in jeder Schicht anschließend eine optimale Fusion beider Anteile erfolgen, so dass Konflikte oder Mehrdeutigkeiten aufgelöst werden können.

Im Folgenden werden die biologisch motivierten Begriffe für den bidirektionalen Informationsfluss durch technisch motivierte Begriffe ersetzt, d.h. für afferente Information wird der Begriff Diagnostik verwendet, für die efferente Information dagegen wird der Begriff Prädiktion gebraucht.

3 Konventionelle Sprachmodellierung basierend auf HMM's

Mit der Einführung der HMM-Technologie wurden lautsprachliche Modelle λ^m (*m=Modellindex*) eingeführt, die drei wesentliche Aspekte der Sprachmodellierung berücksichtigen: Ein zentraler Aspekt ist dabei die Einführung von *N* Zuständen, so dass in jedem Zustand eine Emissionsverteilung $\mathbf{B} = \{b_s^m(\mathbf{x})\}$ von Merkmalvektoren realisiert werden kann. Die Modellierung der Emissionsverteilung eines Zustands $S = s$ beruht auf Gaußschen Mischverteilungen, welche durch die Mixtur-Gewichte c bewertet werden. Neben der statischen Zustandsmodellierung wird der dynamische Aspekt eines HMM durch die Zustandsübergänge $\mathbf{A} = \{a_{s,s'}^m\}$ erfasst, so dass schließlich zweistufige Zufallsprozesse mit einer verdeckten Zustandsvariablen simuliert werden können. Formal kann also jeder Zustand eines Modells λ^m durch eine Mixtur von multivariaten Gaußverteilungen beschrieben werden:

$$b_s^m(\mathbf{x}) = p(\mathbf{x} \mid m, s) = \sum_K c_{s,k}^m \cdot N(\mathbf{x} \mid \boldsymbol{\mu}_{k,s}, \boldsymbol{\Sigma}_{k,s}) \tag{3.1}$$

Wenn für alle Modelle λ^m und deren Zustände S nur ein globaler Satz von K- Basisverteilungen zur Konstruktion der Emissionsdichten verwendet wird, vereinfacht sich die Berechnung der Emissionswahrscheinlichkeit. Dies führt zu semikontinuierlichen HMM's (SCHMM).

$$b_s^m(\mathbf{x}) = \sum_{k=1}^{K} c_{s,k}^m \cdot p(\mathbf{x} \mid \omega_k) \qquad \text{mit} \qquad \sum_k c_{s,k}^m = 1 \tag{3.2}$$

Man kann zeigen, dass die Emissionsdichte aus Gleichung (3.2) durch eine gewichtete Summe der weichen Quantisierungen approximiert werden kann [7].

$$b_s^m(\mathbf{x}) \approx \sum_{k=1}^{K} c_{s,k}^m \cdot p(\omega_k \mid \mathbf{x}) \tag{3.3}$$

Hier wird die Berechnung der Emissionsdichte im Zustand $S = s$ lediglich als Skalarprodukt über dem Ergebnis einer weichen Quantisierung und den Mixturgewichten aufgefasst. Die optimale Decodierung einer HMM-Folge $M = m_1, m_2, ..., m_L$ erfolgt dann gemäß dem MAP-Kriterium.

$$r(M) = \arg\max_{m_1,...,m_L} \left\{ p(m_{1:L} \mid \mathbf{x}_{0:T}) \right\} \qquad mit \quad p(m_{1:L} \mid \mathbf{x}_{0:T}) = \frac{p(\mathbf{x}_{0:T} \mid m_{1:L}) \cdot p(m_{1:L})}{p(\mathbf{x}_{0:T})} \tag{3.4}$$

Die Decodierung wird dabei im Wesentlichen von zwei Faktoren beeinflusst, der akustischen Likelihood und dem Sprachmodell. Dabei wird die Notation $\mathbf{x}_{0:T}$ für die Schreibweise für Sequenzen verwendet, Wahrscheinlichkeitsverteilungen werden als WK-Verteilungen bezeichnet.

In der Darstellung (3.4) werden die durchlaufenen Zustände nicht berücksichtigt. Ein verallgemeinertes Modell, in dem auch die Zustände auftreten können, führt auf ein Schichtenmodell.

3.1 Verallgemeinerung

Zunächst wird für die Verallgemeinerung eines HMM's von der Verbundverteilung für alle drei Zufallsvariablen (Messung- *X*, Zustand- *S*, Modell- *M*) ausgegangen, welche hierarchisch in räumlicher Richtung angeordnet sind (X: unterste-, S: mittlere-, M: oberste Schicht).

$$P(X, S, M) = P(X \mid S, M) \cdot P(S, M) = P(S \mid X, M) \cdot P(X, M) \tag{3.5}$$

Nach wiederholter Anwendung des Bayes'schen Satzes für $P(S, M) = P(S \mid M) \cdot P(M)$ bzw. $P(X, M) = P(X \mid M) \cdot P(M)$ folgt unmittelbar die WK-Verteilung für den Zustand *S*.

$$P(S \mid X, M) = \frac{P(X \mid S, M) \cdot P(S \mid M)}{P(X \mid M)} \tag{3.6}$$

Mit dieser Verallgemeinerung liegt nun für den Zustand *S* eine a-posteriori WK-Verteilung vor. Der Nenner ist unabhängig von *S* und dient daher als Normalisierung. Nimmt man darüber hinaus eine bedingte Unabhängigkeit zwischen den Zufallsvariablen *X* und *M* an, folgt wie in [5] die folgende Gleichung.

$$P(S \mid X, M) = \frac{P(X \mid S) \cdot P(S \mid M)}{P(X)} \tag{3.7}$$

Unter Berücksichtigung von $P(X \mid S) \cdot P(S) = P(S \mid X) \cdot P(X)$ folgt dann schließlich für die WK-Verteilung der Zustände:

$$P(S \mid X, M) = \frac{P(S \mid X) \cdot P(S \mid M)}{P(S)} \tag{3.8}$$

Dieser Ausdruck entspricht den Anforderungen eines corticalen Algorithmus. Die WK-Verteilung $P(S \mid M)$ repräsentiert die prädiktive WK-Verteilung für die Zustände *S*, mit $P(S \mid X)$ liegt eine dia-gnostische Verteilung für die Zustände *S* vor.

4 Einschichtiger Corticaler Algorithmus

Um den Corticalen Algorithmus auf HMM's anwenden zu können, wird sowohl eine prädiktive WK-Verteilung für den Zustand *S* als auch eine diagnostische WK-Verteilung für die Messung *X* benötigt. Die prädiktive Verteilung liegt bereits mit der Markov-Modellierung der Zustandsübergänge vor.

$$p(s_t \mid m) = \sum_{s_{t-1}} p(s_t \mid s_{t-1}, m) \cdot p(s_{t-1} \mid m) \tag{4.1}$$

Diese Gleichung kann als Prädiktionsgleichung aufgefasst werden. Die diagnostische WK-Verteilung ist durch die Emissionsdichte gegeben und stellt eine Likelihood-Funktion dar.

$$p(\mathbf{x}_t \mid s_t, m) = b_s^m(\mathbf{x}_t) = \sum_{k=1}^{K} c_{s,k}^m \cdot p(\omega_k \mid \mathbf{x}_t) \tag{4.2}$$

Um nun die a-posteriori WK-Verteilung auf den Zuständen *S* zu erhalten, erfolgt eine Multiplikation der prädiktiven WK-Verteilung mit der diagnostischen WK-Verteilung mit anschließender Normierung.

$$p(s_t \mid \mathbf{x}_t, m) = \frac{p(\mathbf{x}_t \mid s_t, m) \cdot p(s_t \mid m)}{NORM} \tag{4.3}$$

Für die Erkennung der HMM-Modellfolge wird wieder ein Viterbiverfahren vorgeschlagen, welches nun aber auf der a-posteriori Wahrscheinlichkeit der Zustände basiert. Der Viterbialgorithmus bestimmt die optimale zeitliche Zustandssequenz unabhängig für jedes Modell und berechnet so eine Wahrscheinlichkeit für die gesamte HMM-Folge. Problematisch ist allerdings die Anwendbarkeit der Prädiktionsgleichung. Die Zustandswahrscheinlichkeiten $p(s_t \mid m)$ hängen ausschließlich von der Startverteilung $\Pi = \{\pi_s^m\}$ ab und können einen stationären Endwert $p(s \mid m)$ erreichen, der unabhängig vom Zeitpunkt *t* ist. Im nächsten Abschnitt wird hierfür zunächst eine Lösung angeboten, die auf der bekannten Kalmanfilter-Technik beruht.

4.1 HMM's und verallgemeinerte Kalmangleichungen

Ein Nachteil des oben beschriebenen Algorithmus besteht darin, dass die Prädiktion nicht von der Berechnung der a-posteriori WK-Verteilung profitieren kann. In der modifizierten Realisierung der Prädiktion des Modells λ^m wird nun eine Vorhersage des Zustands s_t im nächsten Zeitpunkt aufgrund des a priori Wissens der Zustandsübergänge $p(s_t \mid s_{t-1})$ und der vergangenen Merkmalsvektoren $\mathbf{x}_{0:t-1}$ getroffen. Dieses a-priori Wissen stellt zeitliches Kontextwissen über Sprachsignale dar, dagegen spiegelt die Likelihood des aktuellen Merkmalvektors nur den aktuellen Messwert und damit lokales Wissen wider. Somit kann zunächst eine modifizierte Prädiktionsgleichung angegeben werden.

$$p(s_t \mid \mathbf{x}_{0:t-1}, m) = \sum_{s_{t-1}} p(s_t \mid s_{t-1}, \mathbf{x}_{0:t-1}, m) \cdot p(s_{t-1} \mid \mathbf{x}_{0:t-1}, m) \tag{4.4}$$

Da die Übergänge durch ein Modell basierend auf a priori Wissen gegeben sind, ergibt sich die Vereinfachung $p(s_t \mid s_{t-1}, \mathbf{x}_{0:t-1}, m) = p(s_t \mid s_{t-1}, m)$.

$$p(s_t \mid \mathbf{x}_{0:t-1}, m) = \sum_{s_{t-1}} p(s_t \mid s_{t-1}, m) \cdot p(s_{t-1} \mid \mathbf{x}_{0:t-1}, m) \tag{4.5}$$

Mit dieser Vorhersage kann nun zu jedem Zeitschritt ein Update erfolgen, d.h. die a-posteriori Schätzung des vorangegangenen Zeitschritts wird zur a-priori Vorhersage des aktuellen Zeitschritts.

$$p(s_t \mid \mathbf{x}_{0:t}, m) = \frac{p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}, s_t, m) \cdot p(s_t \mid \mathbf{x}_{0:t-1}, m)}{p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}, m)} \tag{4.6}$$

Weiterhin kann die Annahme getroffen werden, dass die Likelihood unabhängig von den vorausgegangenen Merkmalsvektoren ist, da die Likelihood vorrangig die Übereinstimmung zwischen einem Messwert und einem statistischen Modell und weniger die zeitliche Historie angibt. Dies führt mit $p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}, s_t, m) = p(\mathbf{x}_t \mid s_t, m)$ und $p(\mathbf{x}_t \mid \mathbf{x}_{0:t-1}, m) = p(\mathbf{x}_t \mid m)$ auf Gleichung (4.7), welche auf den Komponenten Prädiktion, Innovation und Update basiert und somit eine verallgemeinerte Form der Kalmangleichungen realisiert.

$$p(s_t \mid \mathbf{x}_{0:t}, m) = \frac{p(\mathbf{x}_t \mid s_t, m) \cdot p(s_t \mid \mathbf{x}_{0:t-1}, m)}{p(\mathbf{x}_t \mid m)} \tag{4.7}$$

Setzt man nun Gleichung (4.5) in Gleichung (4.7) ein so erhält man Gleichung (4.8).

$$p(s_t \mid \mathbf{x}_{0:t}, m) = \frac{p(\mathbf{x}_t \mid s_t, m) \cdot \sum_{s_{t-1}} p(s_t \mid s_{t-1}, m) \cdot p(s_{t-1} \mid \mathbf{x}_{0:t-1}, m)}{p(\mathbf{x}_t \mid m)} \tag{4.8}$$

Diese Gleichung entspricht der bekannten zeitlichen Forward-Notation für die a-posteriori WK-Verteilung der Zustände eines Modells λ^m :

$$a_j(t) = \left[\sum_{i=1}^{N} \alpha_i(t-1) \cdot a_{i,j} \right] \cdot b_j(\mathbf{x}_t) \tag{4.9}$$

Allerdings liegt hier nicht die Situation vor, dass die prädiktive Verteilung ausschließlich von der Modellebene getrieben wird, sondern auch von den vergangenen Messwerten auf der diagnostischen Ebene. Mit dem Übergang zu mehrschichtigen bidirektionalen HMM's kann ein echter corticaler Algorithmus konzipiert werden.

5 Mehrschichtiger corticaler Algorithmus

Die Überführung des einschichtigen HMM in ein mehrschichtiges bidirektionales HMM (*engl.* ***C**ascaded **B**idirectional **HMM***, CBHMM) basiert auf der Anwendung der verallgemeinerten Kalmangleichungen auf ein räumlich-zeitliches HMM. Weiterhin müssen die a-posteriori-WK-Verteilungen der räumlichen und zeitlichen Komponenten im Bayes'schen Sinne optimal fusioniert werden. Zusätzlich wird hier ein Index $0 \le d \le D-1$ eingeführt, mit dem die einzelnen Schichten adressiert werden. Die höchste Schicht wird mit *d=0* und die niedrigste Schicht mit *d=D-1* bezeichnet. In Analogie zu Gleichung (4.7) werden nun die verallgemeinerten Kalmangleichungen für ein CBHMM in Bottom-Up Richtung hergeleitet, und in eine verallgemeinerte Forward-Notation überführt. Anschließend werden aus Platzgründen die verallgemeinerten Kalmangleichungen des CBHMM in Top-Down Richtung angegeben und in eine verallgemeinerte Backward-Notation überführt. Abschließend werden die Updategleichungen formuliert.

Zunächst wird eine Hilfsgröße **e** eingeführt, mit der ein verallgemeinerter Emissionsvektor bezeichnet wird. Ähnlich wie in Gleichung (4.6) kann man zunächst die a-posteriori WK-Verteilung der Zustände in der Schicht *d* angeben, dabei wird der Modellindex *m* weggelassen.

$$p(s_t^d \mid \mathbf{e}_{0:t}^d) = \frac{\overbrace{p(\mathbf{e}_t^d \mid \mathbf{e}_{0:t-1}^d, s_t^d)}^{Likelihood} \cdot \overbrace{p(s_t^d \mid \mathbf{e}_{0:t-1}^d)}^{zeitl.\,\mathrm{Pr}\ddot{a}diktion}}{p(\mathbf{e}_t^d \mid \mathbf{e}_{0:t-1}^d)} \tag{5.1}$$

Ausgehend von den statischen Eigenschaften der Likelihood kann wie in Gleichung (4.7) folgende Vereinfachung vorgenommen werden:

$$p(s_t^d \mid \mathbf{e}_{0:t}^d) = \frac{p(\mathbf{e}_t^d \mid s_t^d) \cdot p(s_t^d \mid \mathbf{e}_{0:t-1}^d)}{p(\mathbf{e}_t^d)} \tag{5.2}$$

Zunächst kann die Likelihood unter Verwendung der Bayes-Regel noch einmal umgeformt werden:

$$p(\mathbf{e}_t^d \mid s_t^d) = \frac{\overbrace{p(s_t^d \mid \mathbf{e}_t^d)}^{Innovation} \cdot p(\mathbf{e}_t^d)}{p(s_t^d)} \tag{5.3}$$

Setzt man nun (5.3) in (5.2) ein, so erhält man schließlich Gleichung (5.4) mit einem Innovationsterm.

$$p(s_t^d \mid \mathbf{e}_{0:t}^d) = \frac{p(s_t^d \mid \mathbf{e}_t^d) \cdot p(s_t^d \mid \mathbf{e}_{0:t-1}^d)}{p(s_t^d)} \tag{5.4}$$

Die beiden Terme des Zählers entsprechen zum einen einer zeitlichen Prädiktion der Zustandsverteilung basierend auf dem Kontext der Schicht *d* und zum anderen einer Zustandsverteilung basierend auf einer Innovation, welche von der darunter liegenden Schicht getrieben wird.

Prädiktion:

$$p(s_t^d \mid \mathbf{e}_{0:t-1}^d) = \sum_{s_{t-1}^d} p(s_t^d \mid s_{t-1}^d, \mathbf{e}_{0:t-1}^d) \cdot p(s_{t-1}^d \mid \mathbf{e}_{0:t-1}^d) \tag{5.5}$$

Innovation:

$$p(s_t^d \mid \mathbf{e}_t^d) = \sum_{s_t^{d+1}} p(s_t^{d+1} \mid \mathbf{e}_{0:t}^{d+1}) \cdot p(s_t^d \mid s_t^{d+1}, \mathbf{e}_{0:t}^{d+1}) \tag{5.6}$$

Nun kann mit der Innovation und der zeitlichen Prädiktion die a-posteriori WK-Verteilung der Zustände in der Schicht *d* berechnet werden.

$$p(s_t^d \mid \mathbf{e}_{0:t}^d) = \frac{\sum_{s_t^{d+1}} p(s_t^{d+1} \mid \mathbf{e}_{0:t}^{d+1}) \cdot p(s_t^d \mid s_t^{d+1}, \mathbf{e}_{0:t}^{d+1}) \cdot \sum_{s_{t-1}^d} p(s_t^d \mid s_{t-1}^d, \mathbf{e}_{0:t-1}^d) \cdot p(s_{t-1}^d \mid \mathbf{e}_{0:t-1}^d)}{p(s_t^d)} \tag{5.7}$$

Die folgenden beiden Ausdrücke aus (5.7) können als Markovketten 1.Ordnung in räumlicher bzw. zeitlicher Richtung aufgefasst werden

$$p(s_t^d \mid s_{t-1}^d, \mathbf{e}_{0:t-1}^d) = p(s_t^d = j \mid s_{t-1}^d = i) = a_{i,j}^d \tag{5.9}$$

$$p(s_t^d \mid s_t^{d+1}, \mathbf{e}_{0:t}^{d+1}) = p(s_t^d = j \mid s_t^{d+1} = l) = b_{j,l}^d \tag{5.10}$$

Mit dieser Vereinbarung kann (5.7) auch in verallgemeinerter Forward-Notation notiert werden [6].

$$\alpha_j^d(t) = \underbrace{\left[\sum_i \alpha_i^d(t-1) \cdot a_{i,j}^d\right]}_{zeitl.\ Markovkette} \cdot \underbrace{\left[\sum_l \alpha_l^{d+1}(t) \cdot b_{j,l}^d\right]}_{räumliche\ Emission} \tag{5.11}$$

Analog zu Gleichung (5.7) kann ebenfalls eine verallgemeinerte Kalmangleichung in Top-Down Richtung notiert werden.

$$p(s_d^t \mid \mathbf{e}_{0:d}^t) = \frac{\sum_{s_d^{t-1}} p(s_d^t \mid \mathbf{e}_{0:d}^{t-1}) \cdot p(s_d^t \mid s_d^{t-1}, \mathbf{e}_{0:d}^{t-1}) \cdot \sum_{s_{d-1}^t} p(s_d^t \mid s_{d-1}^t, \mathbf{e}_{0:d-1}^t) \cdot p(s_{t-1}^d \mid \mathbf{e}_{0:d-1}^t)}{p(s_d^t)} \tag{5.12}$$

Mit der Annahme einer Markovkette 1. Ordnung in räumlicher und zeitlicher Richtung kann Gleichung (5.12) in Backward-Notation angegeben werden.

$$\beta_j^t(d) = \underbrace{\left[\sum_l \beta_l^t(d-1)\cdot b_{j,l}^d\right]}_{räuml.\ Markovkette} \cdot \underbrace{\left[\sum_i \beta_i^{t-1}(d)\cdot a_{i,j}^d\right]}_{zeitliche\ Emmision} \tag{5.13}$$

Schließlich müssen die Zustandsverteilungen in jeder Schicht und zu jedem Zeitpunkt im Bayes'schen Sinne fusioniert werden. Ohne Herleitung wird hier nur die Formulierung der Updategleichung und deren Interpretation vorgestellt.

$$p(s_t^d \mid \mathbf{e}_{0:t}^d, e_{0:d}^t) = \frac{p(s_t^d \mid \mathbf{e}_{0:t}^d)\cdot p(s_d^t \mid \mathbf{e}_{0:d}^t)}{p(s_d^t)} \tag{5.14}$$

Der linke Teil des Zählers entspricht einer diagnostischen Zustandsverteilung in Bottom-Up Richtung, der rechte Teil dagegen beschreibt die prädiktive Zustandsverteilung in Top-Down Richtung. Die resultierende Zustandsverteilung entspricht nun einer a-posteriori Schätzung.

6 Corticaler Algorithmus in Forward-Backward Notation

Ein CBHMM der Schicht *d* kann als 5-Tupel definiert werden: $H_d = (\sigma_d, \kappa_d, \pi_d, A_d, B_d)$. Dabei gelten folgende Vereinbarungen: σ_d -Menge der Zustandssymbole, κ_d -Menge der Ausgabesymbole, π_d - Startverteilung, A_d - zeitliche Zustandsübergänge, B_d - räumliche Zustandsübergänge. Im Folgenden wird nun der Ablauf des corticalen Algorithmus mit einem CBHMM beschrieben. In Ebene 0 wird zunächst eine Prädiktions-Verteilung für den nächsten Zeitschritt berechnet. Ausgehend von dieser Prädiktions-Verteilung wird durch Abwärtspropagation die entsprechende Prädiktions-Verteilung für alle darunterliegenden Schichten berechnet. Dieser Schritt kann als Vorbahnung der aktuellen Erkennung - basierend auf dem aktuellen Kontext - interpretiert werden. Auf der unteren Ebene *D-1* wird anschließend der Merkmalvektor ***x*** erfasst und die diagnostische Verteilung WK-Verteilung berechnet. Ausgehend von dieser Verteilung werden durch Aufwärtspropagation die dia-gnostischen Verteilungen für alle darüberliegenden Ebenen berechnet. Die Kopplung der Schichten erfolgt dabei durch die räumliche Markovkette. Darüber hinaus erfolgt bei der Aufwärtspropagation eine Update bzw. die Berechnung der Belief-Verteilung für jede einzelne Schicht. Auf dieser Verteilung setzt dann die Berechnung in den nächsten Zeitschritten auf.

Initialisierung:

$$\beta^{-1}(d) = \pi^d ; \alpha^d(-1) = \pi^d \tag{6.1}$$

Prädiktion(t); d=0:D-1

$$\beta_j^t(d) = \underbrace{\left[\sum_{k\in\sigma_{d-1}} \beta_k^t(d-1)\cdot b_{j,k}^d\right]}_{Markovkett\,e-Raum\,,} \cdot \underbrace{\left[\sum_{i\in\sigma_d} \beta_i^{t-1}(d)\cdot a_{i,j}^d\right]}_{zeitl.\ Emmision} \tag{6.2}$$

Diagnostik(t); d=D-1:0

$$\alpha_j^d(t) = \underbrace{\left[\sum_{i\in\sigma_d} \alpha_i^d(t-1)\cdot a_{i,j}^d\right]}_{Markovkette-Zeit} \cdot \underbrace{\left[\sum_{k\in\sigma_{d+1}} \alpha_k^{d+1}(t)\cdot b_{j,k}^d\right]}_{räuml.\ Emission} \tag{6.3}$$

Belief(t), Update (t); d=D-1:0

$$\gamma_t^d(t) = \frac{\beta^t(d)\cdot\alpha^d(t)}{NORM} \quad \xrightarrow{\underline{Update}} \quad \beta^t(d), \alpha^d(t) \tag{6.4}$$

Abschließend muss noch auf eine Besonderheit aufmerksam gemacht werden. Der rechte Term der Diagnostik hat in der untersten Schicht *d=D-1* die Bedeutung einer konventionellen Emission:

$$\sum_{k\in\sigma_{d+1}} \alpha_k^D(t)\cdot b_{j,k}^{D-1} = \sum_k p(\omega_k \mid \mathbf{x}_t)\cdot c_{j,k}^m = b_s^m(\mathbf{x}_t) \tag{6.5}$$

Die im rechten Term von (6.3) adressierte noch tiefer liegende Ebene D entspricht demnach der Ebene der Vektorquantisierung. Die Berechnung des rechten Terms in der untersten HMM-Ebene als Skalarprodukt bleibt strukturell aber identisch zur Berechnung in den höheren Ebenen.

7 Schlussfolgerung

Die Formulierung eines einschichtigen corticalen Algorithmus in seiner einfachsten Form hat wesentliche Schwächen in der Prädiktionsqualität aufgezeigt. Mit der Einführung der verallgemeinerten Kalmangleichungen konnte dieses Problem überwunden werden. Die Analyse der Kalmangleichungen zeigt, dass mit einem Kalmanfilter und dessen Eigenschaft als Zustandsschätzer die Dynamik eines HMM's vollständig erfasst werden kann. Es wurde gezeigt, dass sich die verallgemeinerten Kalmangleichungen in die bekannte Forward-Notation überführen lassen. Das Ziel dieser Arbeit bestand nun darin, einen corticalen Algorithmus zu entwickeln, in dem für jede Schicht die Diagnostik und die Prädiktion als zwei voneinander unabhängige Zustandsschätzungen im Bayesschen Sinne optimal fusioniert werden. Dies gelingt mit einem mehrschichtigen bidirektionalen HMM. Dabei kommt nun zusätzlich zur zeitlichen Dimension eine räumliche Dimension hinzu. In beiden Fällen wird die Dynamik durch verallgemeinerte Kalmangleichungen beschrieben. Fusioniert man die räumliche und zeitliche Zustandsverteilung in jeder Schicht, liegt schließlich ein corticaler Algorithmus vor. Die Verwendung der verallgemeinerten Kalmangleichungen hat zur Folge, dass der corticale Algorithmus äußerst kompakt durch eine verallgemeinerte Forward-Backward Notation formuliert werden kann. In dieser Notation zeigt sich, dass konventionelle HMM's lediglich den unidirektionalen Sonderfall von mehrschichtigen bidirektionalen HMM's darstellen.

8 Ausblick

Nach dem gezeigt worden ist, dass die Dynamik eines mehrschichtigen bidirektionalen HMM's durch verallgemeinerte Kalmangleichungen formuliert werden kann, folgt nun auch die Notwendigkeit Algorithmen für die Dekodierung und das Training mehrschichtiger HMM's anzugeben. Darüber hinaus kann der corticale Algorithmus auch in umgekehrter zeitlicher Richtung formuliert werden, damit sind dann alle Informationsquellen in Raum und Zeit erschlossen. Eine wesentliche Beschränkung des aktuellen Algorithmus besteht allerdings darin, dass jede Schicht ihre Zustandsübergänge zeitsynchron ausführt. Ähnlich wie für einen baumartig strukturierten BPA sollte es möglich sein, auch hier ein baumartig strukturiertes bidirektionales HMM zu entwickeln, in dem die Zustandsübergänge auch asynchron erfolgen können. Dabei wird der corticale Algorithmus eine weitere Verallgemeinerung erfahren.

9 Literatur

[1] Römer, R.: Robuste Spracherkennung auf der Basis recheneffizienter auditiver Modelle, Dissertation TU-München, 2009.

[2] Mountcastle, V.: An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System, The Mindful Brain (Gerald M. Edelman and Vernon B. Mountcastle, eds.) Cambridge, MA: MIT Press, 1978.

[3] George, D.; Hawkins, J.: A Hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex, IEEE International Joint Conference on Neural Networks 2005.

[4] Pearl, J.: Probabilistic Reasoning in Intelligent Systems, Morgan Kaufman 1988.

[5] Lee, T.S.; Mumford, D.: Hierarchical Bayesian Inference in the Visual Cortex, JOSA Vol.20, No.7/July 2003.

[6] Blaylock, N.; Allen, J.: Fast Hierarchical Goal Schema Recognition, Proceedings of AAAI-06, Boston, July 2006.

[7] Fink, G.A.: Mustererkennung mit Markov-Modellen, Theorie-Praxis-Anwendungsgebiete, B.G. Teubner Verlag, Oktober 2003.