

Inkrementelle Erstellung von Worthypothesengraphen

T. Pfau und G. Ruske

Lehrstuhl für Mensch-Maschine-Kommunikation
Technische Universität München
Arcisstr. 21, 80290 München
ruske@e-technik.tu-muenchen.de

KURZFASSUNG

Das hier vorgestellte inkrementell arbeitende Verfahren zur Generierung von Worthypothesengraphen (WHG) ermöglicht es, Worthypothesengraphen schritthaltend während des Suchvorgangs zu erzeugen und direkt an höhere Verarbeitungsstufen weiterzuleiten. Eine Pufferung der von der Suche emittierten Wortfolgehypothesen macht es möglich, sinnvolle lokale Entscheidungen über die Aufnahme von Worthypothesen in den WHG zu treffen. Weitere Maßnahmen - wie ein zusätzliches Pruning auf Wortebene - vermindern die Redundanz der im Worthypothesengraph enthaltenen Information und wirken sich günstig auf Speicher- und Rechenzeitbedarf des Systems aus. Darüberhinaus kann mithilfe des entwickelten Mechanismus die Größe des WHG je nach Anwendungsgebiet unabhängig von der Suchraumgröße eingestellt werden.

1. EINFÜHRUNG

Linguistische Information sollte im Hinblick auf eine schritthaltende Verarbeitung und ein prinzipiell mögliches Echtzeitverhalten eines Spracherkennungssystems idealerweise direkt in den datengetriebenen Beam-Search-Algorithmus eingebunden werden, so daß viele Wissensquellen gleichzeitig ausgewertet werden können. Für dieses Vorgehen gibt es zur Zeit noch keinen befriedigenden Lösungsansatz. Alternativ bietet sich der WHG als Schnittstelle zwischen akustischer Satzerkennung und linguistischer Verarbeitung an. Worthypothesengraphen stellen Wortnetze dar, in denen jede Kante eine Worthypothese verkörpert und die alternative Erkennungsergebnisse für Wörter mit alternativen Segmentierungen enthalten. Im Gegensatz zu vielen anderen Systemen zur Generierung von WHGen, die eine mehrmalige Bearbeitung der Hypothesen erfordern [Kuh96, Mur93], wurde bei der Entwicklung des hier vorgestellten Vorgehens auf ein inkrementelles Vorgehen Wert gelegt.

2. ANFORDERUNGEN AN DEN WORTHYPOTHESENGRAPHEN

Bei der Generierung eines WHG besteht grundsätzlich das Problem, den Suchraum, der auf Wortebene im

wesentlichen die Struktur eines Baumes besitzt, in einen geschlossenen, nicht zyklischen Graphen zu transformieren, der die N besten Hypothesen des Suchraumes enthalten soll. Bei der Umsetzung des Suchraumes in den WHG ist im einzelnen auf folgendes zu achten [Pfa96]:

- Die N besten Hypothesen sollen im WHG enthalten sein.
- Durch die in der Suche angewendeten Pruning-Mechanismen entstehen im Suchraum Äste, die nicht bis zum Ende der Äußerung weiterverfolgt werden (siehe Abbildung 1). Im WHG stellen grundsätzlich nur Äste, die vom Anfangsknoten bis zum Endknoten des WHG verfolgt werden können, sinnvolle Information dar.
- Die Bigramm-Bindungen des Suchraums sollen beim Aufbau des WHG Berücksichtigung finden.
- Die Redundanz der im Worthypothesengraph enthaltenen Information soll möglichst gering sein. Geringfügig unterschiedliche Segmentierungen derselben Wortfolgehypothese (Wort und zugehöriges Vorgängerwort) sollen im WHG nicht mehrfach enthalten sein.
- Eine Einstellung der Komplexität des WHG unabhängig von der Suchraumgröße ist wünschenswert. Dies ist im Hinblick auf einen flexiblen Einsatz des Erkennungssystems von großer Bedeutung. Im Offline-Betrieb kann es sinnvoll sein, komplexe Graphen zu erzeugen, so daß Aussagen über die Konkurrenten der einzelnen Wörter getroffen werden können. Im Online-Betrieb dagegen erfordern nach einer strengen Grammatik als falsch bewertete, spontansprachlich aber akzeptierte Äußerungen den Einsatz einer wenig restriktiven Grammatik, die das akustische Scoring aber auch nur wenig verändern kann. Damit erscheinen sehr tiefe Graphen in diesem Fall wenig sinnvoll.
- Wortrekombinationen zur Reduktion des Baumes in einen geschlossenen Graphen sind notwendig.

Im folgenden wird ein Vorgehen vorgestellt, das die erläuterten Probleme bei der inkrementellen Erstellung von Worthypothesengraphen berücksichtigt.

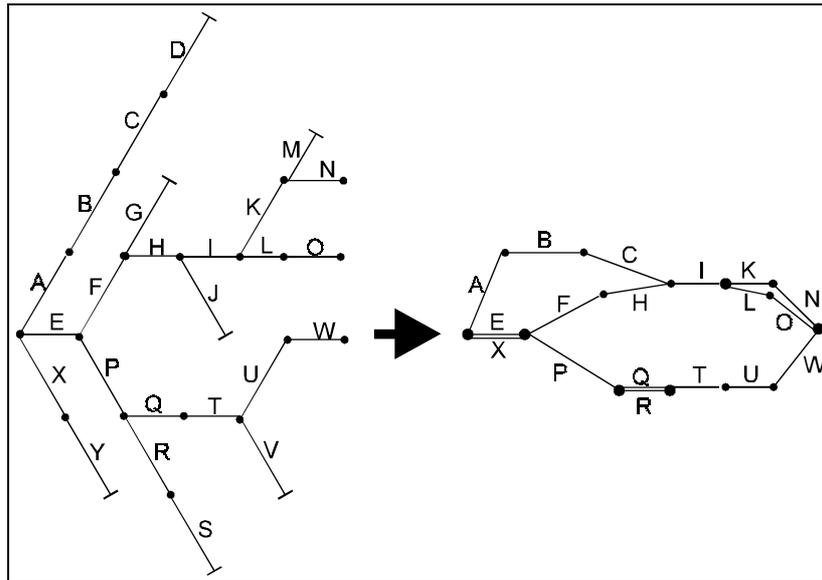


Abbildung 1: Suchraum => geschlossener Graph (symbolische Darstellung: Wörter A, B, ...W)

3. INKREMENTELLES VORGEHEN

Das Verfahren zur inkrementellen Erstellung von Worthypothesengraphen basiert auf den von der Suche pro Frame (10ms) emittierten, mit den akkumulierten Scores bewerteten Wortfolgehypothesen. Diese dienen als Basis für die in jedem Zeitschritt (inkrementell) zu treffenden lokalen Entscheidungen über die Aufnahme von Hypothesen in den WHG. Die inkrementelle Vorgehensweise kann im Gegensatz zu einem nichtinkrementellen Vorgehen prinzipiell nicht sicherstellen, daß der generierte WHG die N besten Hypothesen enthält. Die global N besten Hypothesen sind i.a. nicht identisch mit den pro Frame gefundenen N besten Hypothesen.

Viele Äste des Suchraumes enden aufgrund der in der Suche angewendeten Pruning-Mechanismen, bevor das Ende der Sprachäußerung erreicht ist. Um zu vermeiden, daß die Aufnahme von Hypothesen, die solchen Ästen angehören, auch im WHG zu offenen Enden führt, wird es notwendig, die Wahrscheinlichkeit des weiteren „Überlebens“ einer Hypothese abzuschätzen. Für diesen „Blick in die Zukunft“ ist es erforderlich, die von der Suche emittierten Hypothesen in einem begrenzten zeitlichen Umfang zu puffern. Entscheidungen über die Aufnahme in den Graphen können auf diese Art für Worthypothesen früherer Zeitschritte getroffen werden. Durch die Pufferung entsteht allerdings ein gewisser zeitlicher Versatz zwischen Hypothesenemission und Aufnahme in den WHG.

Der Wahl der Größe des Puffers kommt eine besondere Bedeutung zu. Grundsätzlich wird als sicher angesehen, daß eine Hypothese bis zum Satzende weiterverfolgt werden kann, die zumindest K weitere Folgewörter besitzt. Der Puffer muß daher Wortfolgehypothesen der Länge K+1 aufnehmen können (aktuelles Wort und K Folgewörter). Die Größe K muß nun so gewählt werden, daß zum einen anhand der Folgewörter die „Überlebenswahrscheinlichkeit“ gut abgeschätzt werden kann, zum anderen aber auch im Hinblick auf eine Echtzeitverarbeitung der Versatz zwischen Hypothesenemission und Aufnahme in den WHG nicht zu groß wird. Grundsätzlich kann man aufgrund dieses Vorgehens die Generierung der WHG in zwei Phasen aufteilen:

Phase 1: Aufnahme der Hypothesen in den Puffer:

Um die Größe des Worthypothesengraphen unabhängig von den Suchraumparametern einstellen zu können, wurde ein zusätzlicher Pruning-Mechanismus auf Wortebene eingeführt, der den Suchvorgang und damit verbunden den am besten erkannten Satz (1-best) unbeeinflusst läßt. Das Pruning erfolgt analog zum Pruning des Suchraumes anhand eines Pruningfaktors, der relativ zum besten endenden Wort angewendet wird. Die Redundanz der im WHG enthaltenen Information wird schon bei der Aufnahme der Hypothesen in den Puffer durch die Suche nach identischen Wortfolgehypothesen mit ähnlichen Segmentierungen verringert. Segmentierungen, deren Anfangs- und Endzeitpunkte

sich um weniger als den festgelegten Frameradius R unterscheiden, werden als identisch angesehen.

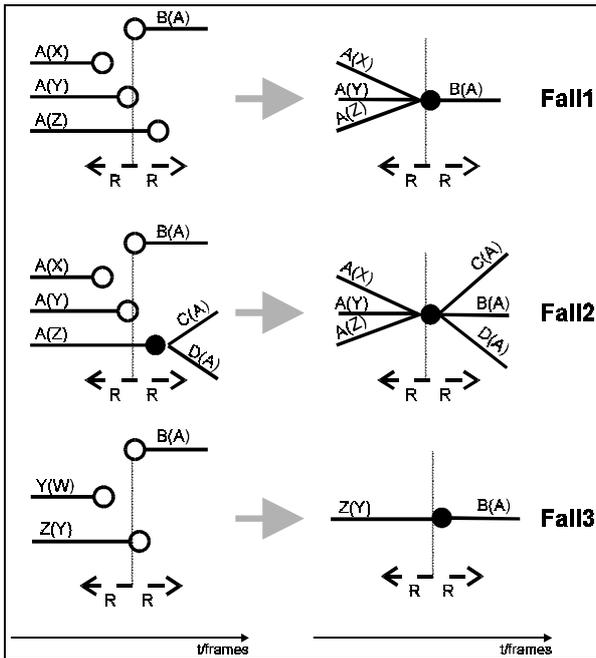


Abbildung 2: Aufnahme einer Hypothese in den Puffer ($B(A)$ = Wort B mit Vorgänger A)

Für aufzunehmende Hypothesen werden solche Vorgänger im Puffer gesucht, die der Bigramm-Bindung entsprechen:

- Es können ein oder mehrere der Bigramm-Bindung entsprechende Wörter im Puffer enthalten sein. Die Endknoten der passenden Vorgänger werden zu einem Knoten rekombiniert, wobei als gemeinsamer Knoten derjenige Endknoten des Vorgängers Verwendung findet, der schon mindestens einen

Nachfolger besitzt. (Abbildung 2: Fall 2). Besitzen mehrere Vorgänger schon Nachfolger, so entscheidet der Score. Besitzt keiner der passenden Vorgänger Nachfolger, so wird derjenige Knoten mit dem höchsten akkumulierten Score als Rekombinationsknoten ausgewählt. (Abbildung 2: Fall 1).

- Aufgrund des zusätzlich eingeführten Prunings auf Wortebene kann es aber auch vorkommen, daß für den aufzunehmenden Knoten kein Vorgänger gefunden wird, da dieser dem Wortpruning zum Opfer gefallen ist, obwohl der entsprechende Ast im Suchraum weitergeführt wurde. Um die Hypothese nicht verwerfen zu müssen, wird die Hypothese unter Verletzung der Bigramm-Bindung an das am besten bewertete im Frameradius R zu findende Wort angehängt. (Abbildung 2: Fall 3).

Phase 2: Übernahme von Hypothesen aus dem Puffer in den Worthypothesengraphen:

Der Puffer ist als Ringpuffer angelegt, daher werden die am längsten im Puffer befindlichen Hypothesen aus diesem entfernt, wenn der Puffer voll ist. Anhand der Länge der im Puffer befindlichen Folgewortketten wird entschieden, ob die Hypothese in den WHG aufgenommen wird.

Besitzt die Hypothese eine Nachfolgerkette der Länge K, die als ausreichend lang für ein sicheres „Überleben“ angesehen wird, wird die Hypothese aus dem Puffer entfernt und in den WHG aufgenommen.

Ist die Nachfolgerkette kürzer als K-1 Wörter, obwohl ein ausreichend großer zeitlicher Ausschnitt im Puffer enthalten ist, so wird angenommen, daß der Pfad dem Pruning zum Opfer gefallen ist. Die Hypothese wird aus dem Puffer entfernt, aber nicht in den Graphen aufgenommen.

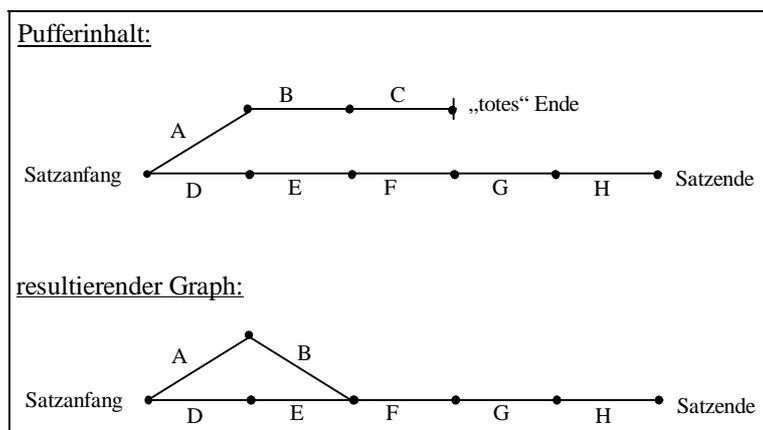


Abbildung 3: Präventives Einhängen ($K=2$)

Hypothesen mit Nachfolgerketten der Länge $K-1$ erfordern eine Sonderbehandlung. Der Vorgänger einer solchen Hypothese hat zum Zeitpunkt der Entscheidung über seine Aufnahme in den Puffer K Nachfolger besessen. Ein Verwerfen dieser Hypothese würde folglich zu einem „toten Ende“ im Graphen führen. Um dies zu vermeiden, werden Hypothesen mit $K-1$ Nachfolgewörtern unter Verletzung der Bigramm-Bindung an einen Nachfolger angehängt, der K Nachfolger besitzt. Dieser Vorgang wird „präventives Einhängen“ genannt.

Durch Anwendung des beschriebenen Verfahrens wird bei richtiger Dimensionierung des Puffers vermieden, daß „tote Enden“ und „tote Anfänge“ im Graphen entstehen.

4. ERGEBNISSE

Um das Verfahren zu testen, wurden Untersuchungen mithilfe des im folgenden beschriebenen Spracherkennungssystems durchgeführt.

- Vorverarbeitung mit 20 spektralen Kanälen entsprechend dem menschlichen Gehör. Erste und zweite Ableitung der spektralen Kanäle, sowie Energie und Nulldurchgangsrate ergänzen den Parametersatz auf 66 Dimensionen.
- Akustische Modellierung: kontinuierliche HMM-Modelle für Phoneme und größere - auf Silbenteilen basierenden - Einheiten. (ca. 1.5 Mio Parameter)
- Suche: Beam-Search-Algorithmus mit Histogramm-Pruning.
- Trainingsmaterial: Verbmobil CD1-3 (ca. 10h Spontansprache aus dem Gebiet Terminabsprache)
- Testset: Verbmobil CD4.

Aufgrund der Vielzahl der zu optimierenden Parameter konnten bisher nur Testläufe auf einer Untermenge des Evaluierungs-Sets '95 durchgeführt werden. Erste Ergebnisse zeigen, daß die im 1-best Modus gefundenen Sätze (Worterkennungsrate 67,1%) praktisch immer auch in den generierten WHGen enthalten sind. Eine Erhöhung der Tiefe der WHGen kann also die Worterkennungsrate nur weiter erhöhen.

Das Verfahren zur inkrementellen Erstellung der WHGen führt dazu, daß die Knoten der WHGen nicht durchgehend nummeriert sind. Zur Evaluierung wird deshalb ein zusätzliches Programm benötigt, das die Knoten der generierten WHGen mit durchgehenden Nummern versieht.

5. ZUSAMMENFASSUNG

Mithilfe des hier vorgestellten Verfahrens können bei geeigneter Wahl der Parameter inkrementell während des Suchvorgangs Worthypothesengraphen erstellt

werden. Der Umfang der Graphen kann dabei je nach Anwendungsgebiet unabhängig von der Suchraumgröße variiert werden. Es konnte experimentell gezeigt werden, daß trotz der angewendeten Heuristiken, die zwangsläufig aus der Forderung nach inkrementellem Vorgehen resultieren, der im 1-best Modus als am besten erkannte Satz auch im generierten Worthypothesengraphen enthalten ist. Weitere Evaluierungen sollen dazu genutzt werden, einen genaueren Zusammenhang zwischen den verschiedenen Parametern der Suche, der Worthypothesengraphengenerierung und der Worterkennungsrate sowie der Wortfehlerrate aufzuzeigen.

Die vorliegende Arbeit wurde im Rahmen des Verbundvorhabens Verbmobil vom Bundesministerium für Forschung und Technologie (BMFT) unter dem Förderkennzeichen 01 IV 102 C/6 gefördert. Die Verantwortung für den Inhalt dieser Arbeit liegt bei den Autoren.

6. LITERATUR

- [Kuh96] T.Kuhn, P.Fetter, A.Kaltenmeier, P.Regel-Brietzmann, DP-Based Wordgraph Pruning. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Seite 861-864, Atlanta, 1996.
- [Mur93] H.Murveit, J.Butzberger, V.Digalakis und M.Weintraub, Large-Vocabulary Dictation Using SRI's DECIPHER Speech Recognition System: Progressive Search Techniques. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Seite 319-322, Minneapolis, 1993.
- [Pfa96] T.Pfau, G.Ruske, Inkrementelle Erstellung von Worthypothesengraphen, Verbmobil Report 106, TU München, erschienen am DFKI Saarbrücken, 1996.