DETEKTION UNBEKANNTER WÖRTER IN DER AUTOMATISCHEN ERKENNUNG SPONTANER SPRACHE

André Berton und Olaf Schreiner

DaimlerChrysler Research and Technology
Postfach 2360, 89013 Ulm, Deutschland
andre.berton@daimlerchrysler.com

Abstract: Aufgrund der Vielfalt von Ausdrucksmöglichkeiten in Spontansprache verwenden Benutzer von Spracherkennungssystemen Wörter, die nicht im begrenzten Erkennervokabular vorhanden sind, sogenannte unbekannte Wörter. Solche Wörter führen zu Erkennungsfehlern und müssen zur Verbesserung des Sprachverstehens detektiert werden. Dieser Beitrag stellt zwei Methoden zur Detektion gegenüber. Die explizite Modellierung verwendet spezielle HMM für unbekannte Wörter, die jeweils auf allen Wörtern einer bestimmten Länge trainiert wurden. Die implizite Methode basiert auf erkenner-internem Wissen (Konfidenz-merkmalen), das unbekannte Wörter von Wörtern des Vokabulars unterscheidet. Beide Methoden eignen sich zur Detektion unbekannter Wörter, wobei gezeigt werden kann, daß die vorgestellten impliziten Modelle signifikant bessere Leistungen als die expliziten Modelle erreichen. Sie sind in der Lage, etwa jedes dritte unbekannte Wort zu detektieren, wobei jede zweite OOV-Hypothese korrekt ist. Erstmals werden hybride Ansätze vorgestellt, die beide Methoden kombinieren. Sie erhöhen die Präzision der Detektion geringfügig, jedoch kompensiert diese Verbesserung den wesentlich höheren Rechenaufwand nicht. Eine weitere Präzi-sierung der expliziten Modelle wird zur Verbesserung der Detektion vorgeschlagen.

1 Einführung

Automatische Spracherkennungssysteme verfügen stets nur über einen begrenzten Wortschatz, der sich von wenigen Wörtern (Ziffernerkennung) bis hin zu mehreren hunderttausend Wörtern (Diktiersystem) erstrecken kann. Im Gegensatz zu diesem definierten Wortschatz steht die deutsche Sprache, die aufgrund ihrer vielfältigen Wortbildungsformen (Flexion, Derivation, Kompositabildung) unendlich ist. Das führt dazu, daß ein Benutzer Wörter verwenden kann, die dem Spracherkennungssystem unbekannt sind. Diese Wörter nennt man Out-Of-Vocabulary-Wörter (OOV). Als unbekannte Wörter werden laut Definition neben gültigen deutschen Wörtern auch sogenannte irreguläre OOV-Wörter berücksichtigt, wie z.B. Wortfragmente, Neustarts, Neologismen und Wortabbrüche.

Die Äußerung solcher OOV-Wörter führt zur Substitution des unbekannten Wortes durch ein bekanntes Wort, welches phonetisch und grammatikalisch möglichst gut passt. Aus einem OOV-Wort entstehen im Mittel zwei Erkennungsfehler [1], die in den meisten Fällen wiederum Fehlinterpretationen in der Anwendung zur Folge haben. Unbekannte Wörter treten besonders häufig in Spontansprache auf, da sich der Benutzer während des spontanen Sprechens noch in der Entschlussfindungsphase befindet [2]. Geht man im Mittel von 2% unbekannten Wörtern bei Spontansprache aus (etwa jeder fünfte Satz enthält ein OOV-Wort), so zeigt sich die Signifikanz des Problems besonders deutlich.

Doch wie kann man das Problem der unbekannten Wörter lösen? Die einfachste Lösung des OOV-Problems besteht in einer Vokabularvergrößerung. Der Anteil unbekannter Wörter sinkt damit deutlich, allerdings zulasten einer steigenden Verwechselbarkeit von Wörtern und einer sinkenden Erkennungsrate: eine inakzeptable Lösung [3]. Eine andere Methode basierte

darauf, geeignetere Basiseinheiten für die Spracherkennung als Wörter zu finden [4], um die Vielfalt der Wortbildungsmöglichkeiten der deutschen Sprache besser abzudecken. Eine Zerlegung der Wörter in Komposita-Konstituenten verringerte die Anzahl unbekannter Wörter und verbesserte die Erkennung, wobei sich die nachträgliche Zusammenfügung der Konstituenten zu Komposita als sehr aufwändig herausstellte.

Die geeignete Lösung wäre eine dynamische Erweiterung des Vokabulars durch automatische Transkription der OOV-Wörter, wofür deren Detektion eine notwendige Voraussetzung ist. Aus der Literatur sind zwei verschiedene Detektionsmethoden für OOV-Wörter bekannt. Bei der expliziten Modellierung werden spezielle akustische Modelle zur Abdeckung phonetischer Eigenschaften unbekannter Wörter und Sprachmodelle zur Beschreibung des unmittelbaren Kontextes unbekannter Wörter trainiert [1,5]. Die implizite Modellierung hingegen basiert auf erkenner-internen Merkmalen, die in Form von Konfidenzschätzungen zur Verfügung stehen, um unbekannte Wörter zu detektieren [6]. Auf beide Methoden wird in diesem Beitrag näher eingegangen. Aus ihnen werden hybride Methoden entwickelt, welche die Vorteile beider Methoden zur Detektion unbekannter Wörter berücksichtigen.

In diesem Beitrag werden Lösungen zur Detektion unbekannter Wörter und somit zur Fehlervermeidung für Spontansprache präsentiert. Die Untersuchungen sind für sprachbediente Anwendungen im Automobil relevant, da sich derzeit der Übergang von kommandobasierten Applikationen hin zu freierer Eingabe ankündigt.

Dieser Beitrag ist folgendermaßen organisiert: Abschnitt 2 beschreibt die Methoden zur Detektion unbekannter Wörter, die gegenübergestellt werden. Die Evaluierungskriterien werden im Abschnitt 3 dargelegt. In Abschnitt 4 wird die Stichprobe und der zugrundeliegende Spracherkenner beschrieben, bevor die experimentellen Ergebnisse diskutiert werden. Die Schlußfolgerungen und der Ausblick schließen das Dokument im Abschnitt 5 ab.

2 Methode

2.1 Explizite Modellierung durch OOV-HMM

Automatische Spracherkenner suchen die Wortfolge $\hat{W}$ mit der höchsten Rückschlusswahrscheinlichkeit $P(W \mid O)$ für die beobachtete Folge akustischer Merkmalsvektoren O des Sprachsignals. Die Suche kann mathematisch auf die folgende Bayes-Formel zurückgeführt werden:

$$P(W \mid O) = \frac{P(O \mid W)P(W)}{P(O)}$$

Für die Suche der besten Wortfolge muss die a-priori-Wahrscheinlichkeit der Beobachtungsfolge akustischer Merkmalsvektoren $P(O)$ nicht geschätzt werden, da das Sprachsignal und somit auch der Schätzwert für alle konkurrierenden Wortfolgen identisch ist. Somit beruht die Dekodierung der besten Wortfolge auf der Schätzung der akustischen Wahrscheinlichkeiten $P(O \mid W)$ und der Sprachmodellwahrscheinlichkeiten $P(W)$.

Dies gilt auch für die explizite Modellierung unbekannter Wörter. Um möglichst hohe Wahrscheinlichkeiten $P(O \mid W)$ bei ungesehenen Wörtern zu erzielen, werden spezielle, akustische Modelle trainiert, die im allgemeinen besser auf ungesehene Wörter passen als trainierte HMM für Wörter des Vokabulars. Diese Modellierung wird durch den Ansatz generischer, längenabhängiger OOV-HMM aus [1] realisiert, der nur geringfügig modifiziert wurde, da dieser Ansatz bei einem geringen Rechenzeit- und Trainingsaufwand gute Detektionsleistungen erreichte. Die Topologie der 15 längenabhängigen Markov-Modelle ist

in Tabelle 1 dargestellt. Sie wurde derart gewählt, dass für jedes Modell ausreichend Trainingsdaten zur Verfügung stehen. Die Parameter für die OOV-HMM wurden durch den bekannten Baum-Welch-Algorithmus bestimmt.

.

Modell	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Zustandsanzahl	3	4	5	6	7	8	9	10	11	12	13	14	15	15	15
Phonemanzahl	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17+

Tabelle 1 – Topologie der 15 längenabhängigen OOV-Modelle

Die Wahrscheinlichkeit des Wortüberganges für unbekannte Wörter P(W) wird für spontansprachliche Anwendungen durch statistische Sprachmodelle approximiert. Da das Vokabular nicht alle Wörter des Korpus für das Sprachmodell-Training enthält, können die Übergangswahrscheinlichkeiten für OOV-Wörter auf dieselbe Art und Weise geschätzt werden, wie die Übergangswahrscheinlichkeiten für Wörter aus dem Vokabular. Der Anteil unbekannter Wörter in der Trainingsstichprobe und in der realen Anwendung (Teststichprobe) sollte ungefähr übereinstimmen, um bestmögliche Detektionsergebnisse zu erreichen. Zur Korrektur möglicher Abweichungen wurde die Methode der Iterativen Substitution aus [7] verändert, indem der Korpus nicht geteilt, sondern vervielfältigt wurde, wobei für jede Kopie des Korpus andere (seltene) Wörter aus dem Vokabular entfernt wurden. Schließlich wurde ein zusätzlicher Straffaktor für jeden Übergang zu einem unbekannten Wort in der Dekodierung eingeführt, um die Zahl der Hypothesen unbekannter Wörter steuern zu können.

2.1.1 Berücksichtigng der Silbenstruktur für explizite OOV-Modelle

Eine spezielle Kategorie stellen die unbekannten Wörter dar, die im Sprachmodell wohldefiniert sind, aber aus strategischen Gründen nicht im Lexikon sein können. Diesen Fall haben wir betrachtet für Ortsnamen. Für eine Anwendung sollten 60.000 Ortsnamen in einem spontansprachlichen Satz gesprochen werden können. Dazu wurde ein statistisches Sprachmodell auf Sätzen mit regulären Ortsnamen trainiert, die Sprachmodellklasse der Ortsnamen aber später mit OOV-Modellen gefüllt. Der Ansatz der längenabhängigen expliziten OOV-Modelle wurde dazu insofern verfeinert, dass zusätzlich zur Wortlänge auch die Wortstruktur berücksichtigt wurde. Dazu wurden die speziellen phonetischen Beschreibungen jeder Silbe der Ortsnamen sukzessive verallgemeinert, indem jeweils ähnliche Phoneme zu Phonemklassen zusammengefasst wurden, wobei als kleinste phonetische Einheit jeweils das Startkonsonanten-Cluster, der Vokalkern und das Endkonsonanten-Cluster betrachtet wurden. Die Lexikoneinträge für die OOV-Modelle wurden dann aus einem bis drei HMM je Silbe aufgebaut, so dass die Anzahl der Silben der Ursprungswörter im wesentlichen die Länge der OOV-Lexikoneinträge bestimmt. Schließlich wurden 5 Gruppen verschiedener Granularität für die Experimente ausgewählt. Tabelle 2 gibt zu jeder der Gruppen die Anzahl der verschiedenen Cluster, HMM und Lexikoneinträge an. Gruppe 1 entspricht dabei grob den längenabhängigen OOV-Modellen aus [1]. Die resultierenden akustischen Modelle wurden dann jeweils mit den ihnen zugeordneten Ortsnamen trainiert.

Gruppe	**Startkons.-Cluster**	**Vokal-Cluster**	**Endkons.-Cluster**	**HMMs**	**Lexikon-einträge**
1	1	1	1	3	33
2	2	1	2	5	399
3	2	2	2	6	567
4	2	2	3	7	4509
5	3	2	3	8	4936

Tabelle 2 – Definition der fünf OOV-Modell-Gruppen

2.2 Implizite Modellierung durch Konfidenzen

Die implizite Modellierung detektiert unbekannte Wörter durch Verwendung erkennerinterner Parameter. Sie beruht auf Konfidenzmerkmalen, die mit der Korrektheit der Erkennung eines gesprochenen Wortes korrelieren [8]. Ein zweistufiges Schätzverfahren generiert eine Reihe geeigneter Konfidenzmerkmale (s. Abbildung 1) und schätzt aus dem Merkmalsvektor die Rückschlusswahrscheinlichkeiten für die drei Klassen "richtig", "falsch" und "OOV" nach der folgenden Formel:

$$P(k \mid v) = \frac{P(v \mid k)P(k)}{\sum_{k} P(v \mid k)P(k)}$$

wobei v den Merkmalsvektor und k die Klasse repräsentiert, der der Vektor zugeordnet werden soll. Die Klassenzugehörigkeit $\hat{k}$ für das hypothetisierte Wort wird entsprechend der maximalen Rückschlusswahrscheinlichkeit $P(k \mid v)$ aller Klassen bestimmt. Eine Vielzahl potentieller Konfidenzmerkmale aus allen Stufen eines Spracherkenners wurde in [8] untersucht. Zu den erfolgreichsten Kandidaten gehören:

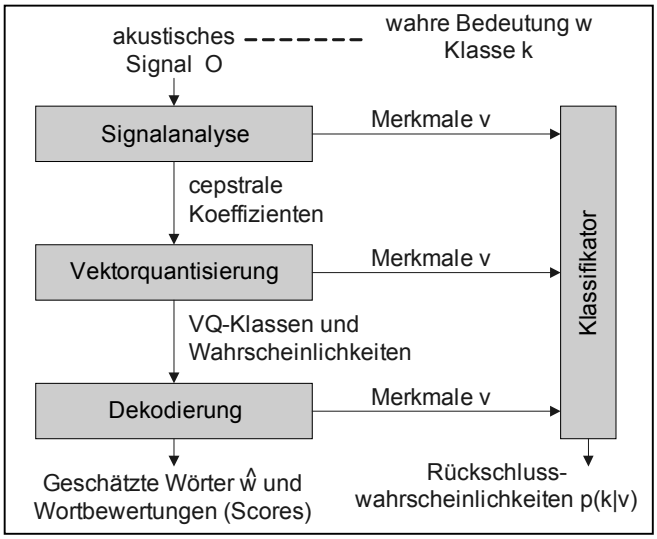

Abbildung 1 – Zweistufiges Verfahren zur Schätzung von Konfidenzen

- SNR: Signal-Rausch-Abstand
- VQMaxProb: Maximale VQ-Wahrscheinlichkeit aller Codebuchklassen
- GraphHomogen: Homogenitätsmaß für Kantendurchgangswahrscheinlichkeit des aktuellen Wortes
- SpeakRate: Sprechgeschwindigkeit (Frames pro Zustand)
- LRGarbBestFrm: Likelihoodverhältnis zwischen Wortscore und Score des besten Müllwortes (frameweise berechnet)
- LMProbAvg: Mittlere Sprachmodell-Übergangswahrscheinlichkeit zum aktuellen Wort
- WortLen: Länge des Wortes in Frames
- NormScore: Längennormierter akustischer Score

Wie gut sich die einzelnen Konfidenzmerkmale zur Detektion unbekannter Wörter eignen, soll experimentell untersucht werden. Die beste Kombination der relevanten Merkmale wird durch iterative Hinzunahme des jeweils besten Kandidaten gesucht.

Bevor die Rückschlusswahrscheinlichkeiten $P(k \mid v)$ jedoch geschätzt werden können, müssen die Erkennungsergebnisse für Sprachstichproben mit den entsprechenden Klassen k annotiert werden. Auf den annotierten Stichproben werden Polynomklassifikatoren zur Schätzung der entsprechenden Verteilungen trainiert, da sie bereits bei geringem Trainingsaufwand gute Ergebnisse erzielen.

2.3 Hybride Techniken

Zwei Methoden, die explizite und implizite Modelle kombinieren, sollen daraufhin untersucht werden, ob sie die Detektion unbekannter Wörter verbessern können.

ExplicitDetect+ConfidenceVerify: Dieser Ansatz versucht, die Detektionsleistung der expliziten Modelle mit der Verifikationsleistung der Konfidenzmaße zu kombinieren. Die expliziten Modelle hypothetisieren während der Dekodierung unbekannte Wörter, die in einer nachfolgenden Verifikationsstufe durch Konfidenzmaße bewertet und aufgrund einer Schwellwert-Operation akzeptiert oder zurückgewiesen werden. Dieser hybride Ansatz erlaubt es, eine größere Anzahl von OOV-Hypothesen zu generieren, unter der Voraussetzung, daß die Verifikationsstufe die meisten falschen Hypothesen zurückweist und die Wortfehlerrate des Gesamtsystems nicht verringert wird (s.Abbildung 2).

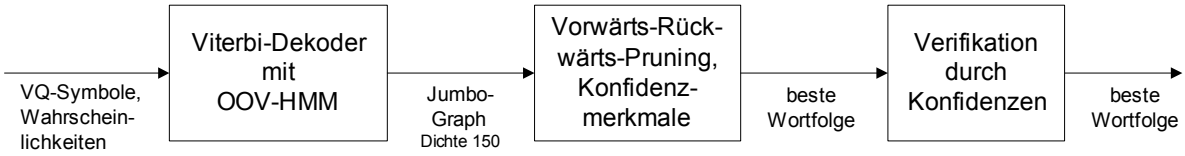

Abbildung 2 – Schematische Darstellung der hybriden Methode ExplicitDetect+ConfidenceVerify

FullDetect+ConfidenceVerify: Dieser Ansatz erweitert die vorherige Methode derart, dass nach der Generierung von OOV-Hypothesen durch explizite Modelle noch eine Klassifikationsstufe zur OOV-Detektion durch implizite Modelle (Konfidenzschätzer) folgt, bevor die Verifikationsstufe die unwahrscheinlichen Hypothesen verwirft (s.Abbildung 3). Die beiden verschiedenen Detektionsansätze für unbekannte Wörter erlauben es, aufgrund ihrer komplementären Eigenschaften eine größere Anzahl von OOV-Hypothesen zu generieren. Die zwei Konfidenzschätzer dienen der Detektion unbekannter Wörter bzw. der Rückweisung falscher Hypothesen. Sie werden auf unterschiedlichen Konfidenzmerkmalen oder deren Kombinationen beruhen, da anzunehmen ist, dass sich einige Merkmale besser zur Detektion, andere hingegen besser zur Verifikation eignen.

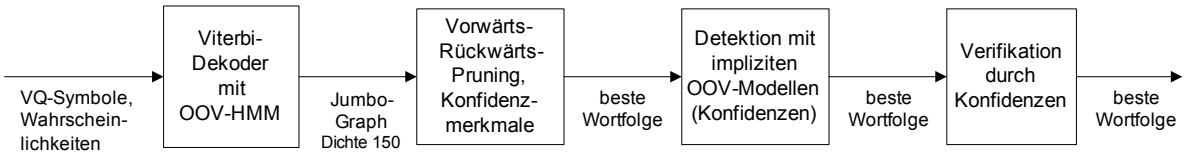

Abbildung 3 – Schematische Darstellung der hybriden Methode FullDetect+ConfidenceVerify

3 Evaluierungskriterien

Die Güte der Detektion unbekannter Wörter wird üblicherweise durch *Detektionsrate* und *Präzision* gemessen. Die Detektionsrate *DR* gibt an, welcher Anteil der unbekannten Wörter der Referenztransliteration korrekt detektiert wurde. Die Präzision PRZ misst den Anteil korrekter OOV-Hypothesen. Sie bewertet dadurch, wie genau die Modellierung unbekannter Wörter funktioniert. Die Klassifikationsfehlerrate KFR beschreibt den Anteil fehlerhafter Klassifikatorentscheidungen. Sie dient der Klassifikatoroptimierung.

$$DR = \frac{\#OOV_{COR}}{\#OOV_{REF}} \qquad PRZ = \frac{\#OOV_{COR}}{\#OOV_{HYP}} \qquad KFR = 1 - \frac{\#\hat{k}_{COR}}{\#k_{REF}}$$

4 Experimente und Ergebnisse

4.1 Der Spracherkenner

Für die Experimente wurde ein deutscher 16-kHz-Spracherkenner [9] mit semikontinuierlichen HMM auf etwa 50 Stunden Spontansprache aus der Verbmobil-Stichprobe [10] trainiert. Bei der Dekodierung kommt der Viterbi-Algorithmus mit akustischen Modellen

(HMM), verbaumtem Lexikon und einem statistischen Bigramm-Sprachmodell zum Einsatz. Dessen Ergebnis wird in einer weiteren Verarbeitungsstufe durch Dynamische Programmierung mit einem Trigramm-Sprachmodell ausgedünnt wird. Die statistischen Sprachmodelle wurden mit einem domänenspezifischen Korpus (etwa 600.000 Wörter) trainiert.

4.2 Die Stichprobe

Als Teststichprobe soll die letzte Evaluierungsstichprobe des Projektes Verbmobil-II dienen (Eval'2000). Da die Anzahl unbekannter Wörter dieser Stichprobe zu klein ist, um Klassifikatoren robust trainierten zu können, wurde sie in der Stichprobe durch Verkleinerung des Vokabulars um 15% etwa vervierfacht. Tabelle 3 beschreibt die Charakteristika der modifizierten Stichprobe.

Stichproben	Training	Test
Anzahl Referenzwörter	35.943	10.916
Vokabulargröße	8.590	8.590
Anzahl unbekannter Wörter	4.062	1.251
OOV-Rate	11,3%	11,5%

Tabelle 3 – Charakteristik der Stichprobe

4.3 Baseline-Experiment

Welche Ergebnisse erreicht der Spracherkenner ohne OOV-Detektion mit verkleinertem Vokabular? Tabelle 4 stellt die Evaluierungsergebnisse mit zeitlicher Zuordnung der Worthypothesen zur Referenztransliteration zusammen. Durch die Verringerung der Vokabulargröße stieg die Wortfehlerrate erwartungsgemäß um etwa 10% absolut, was sich zum Großteil auf den gestiegenen Anteil unbekannter Wörter zurückführen lässt.

Stichproben	Training	Test
Anzahl Worthypothesen	36.551	11.227
Wortfehlerrate	43,9 %	48,5%

Tabelle 4 – Baseline-Ergebnisse

4.4 Ergebnisse expliziter Methoden

Die 15 längenabhängigen OOV-HMM wurden auf der gesamten Trainingsstichprobe (50 Stunden Sprache) trainiert, wobei alle Wörter mit einer bestimmten Anzahl von Phonemen zum Training eines der 15 OOV-HMM dienten. Für das Training des Sprachmodells wurde der Anteil unbekannter Wörter künstlich durch iterative Vokabularverkleinerung erhöht, um eine realistische Größenordnung für die gegebene Anwendung zu erzielen. Nach Optimierung des Straffaktors für den Übergang zu einem unbekanntes Wort, erreichten die Modelle eine Detektionsrate von 25,4% bei einer Präzision von 41,9%. Diese Ergebnisse liegen in einer vergleichbaren Größenordnung wie bei [1].

Für den Test der akustischen Silbenstruktur-Modelle wurden aus der Teststichprobe die 1000 Sätze ausgewählt, die Ortsnamen enthalten. Dadurch wurde eine künstliche OOV-Rate von 23,2% hergestellt. Der Test verlief in zwei Stufen: In einer ersten Stufe wurde ein Erkennungslauf mit dem OOV-Setup durchgeführt. In einer zweiten Stufe wurden die Signalstellen, an denen der Erkenner in der ersten Stufe OOV-Wörter gefunden hatte, mit einem Einzelwort-Erkenner gegen die echten Referenz-Ortsnamen getestet, um zu überprüfen,

	Stufe 1			**Stufe 2**
Gruppe	**WFR**	**DR**	**PRZ**	**WFR_{EW}**
1	33.9 %	86.1 %	78.2 %	83.6 %
2	31.0 %	89.7 %	76.3 %	77.3 %
3	28.9 %	92.7 %	76.3 %	69.7 %
4	18.8 %	97.1 %	80.4 %	40.0 %
5	18.7 %	97.0 %	80.9 %	39.2 %

Tabelle 5 – Ergebnisse der 2 Stufen

wie präzise die OOV-Modelle die fehlenden Ortsnamen in der ersten Stufe auffüllen. Tabelle 5 zeigt, dass durch eine geeignete Erhöhung der Anzahl der akustischen OOV-Modelle um zwei Größenordnungen (von Gruppe 1 nach Gruppe 4) Wortfehlerrate (WFR), Detektionsrate (DR) und Präzision (PRZ) erheblich verbessert werden konnten. Die Ergebnisse stehen für den weiteren Vergleich jedoch außer Konkurrenz, da sie auf den Spezialfall der Ortsnamen zugeschnitten waren.

4.5 Ergebnisse impliziter Methoden

Zur Klassifikation der Teststichprobe in die 3 Zielklassen wurde ein Polynomklassifikator zweiten Grades trainiert. Er erreicht eine Detektionsrate von 33,7% bei einer Präzision von 49,8%. Die impliziten Modelle detektieren somit deutlich besser als die expliziten Modelle.

Welche Konfidenzmerkmale tragen besonders zur Unterscheidung der Klassen bei? Aus Abbildung 4 geht hervor, daß *GraphHomogen* den größten Beitrag leistet: Es trennt die zwei häufigsten Klassen „richtig“ und „falsch“. Weitere Konfidenzmerkmale, die diese zwei Klassen gut trennen, sind *NormScore* und *SpeakRate*. Die Merkmale *LMProbAvg*, *LRGarbBestFrm* und *WordLen* dienen der Trennung der Klassen „falsch“ und „OOV“.

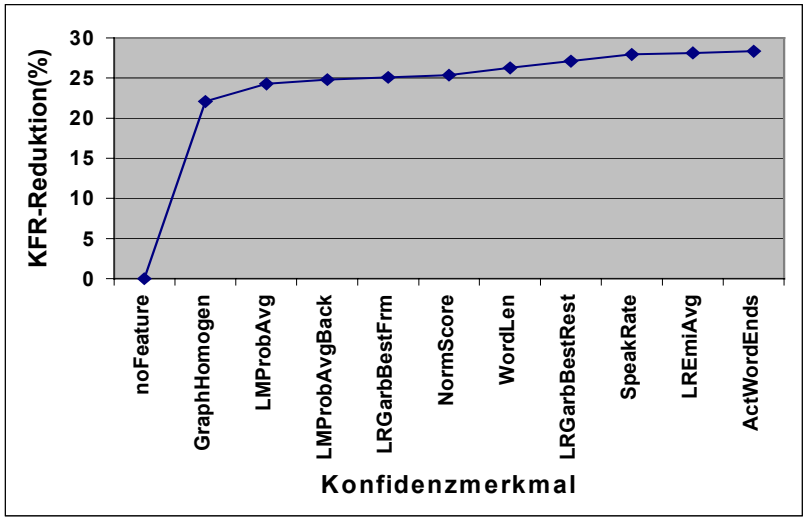

Abbildung 4 – Iterative Auswahl der Merkmale zur Klassentrennung

4.6 Ergebnisse hybrider Methoden

In einer letzten Experimentenreihe wird versucht, explizite und implizite Detektionstechniken für unbekannte Wörter zu kombinieren. Abbildung 5 zeigt die Ergebnisse aller Modellierungstechniken vergleichend. Die Methode ExplicitDetect+ConfidenceVerify arbeitet zwar aufgrund der Verifikation präziser als die expliziten Modelle, erreicht aber mit 22,2% nur eine ungenügende Detektionsrate. Die beste Leistung wird durch den Ansatz FullDetect +ConfidenceVerify mit 34,1% Detektionsrate und 56,4% Präzision erreicht. Da dieser Ansatz aber nur eine 5% höhere Präzision als die viel einfacheren impliziten Modelle aufweist, ist der Mehraufwand für die hybride Modellierung nicht gerechtfertigt.

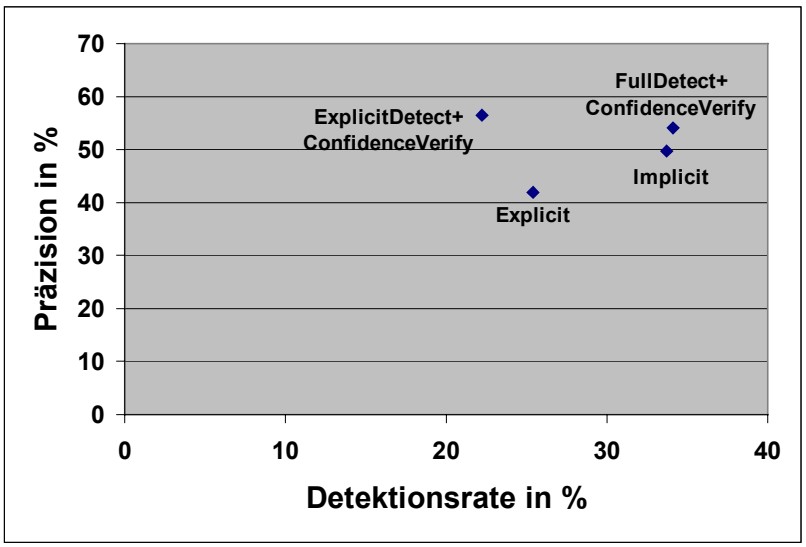

Abbildung 5 – Vergleich der Detektionsmethoden

5 Schlussfolgerungen und Ausblick

Unbekannte Wörter stellen ein signifikantes Problem für die Erkennung von Spontansprache dar, da ihr Anteil an der Sprache etwa 2-3% beträgt und jedes unbekannte Wort im Durchschnitt zu zwei Erkennungsfehlern führt. Zur Verbesserung des Sprachverstehens müssten die unbekannten Wörter zuerst detektiert werden. Hierfür wurden zwei Methoden gegenübergestellt. Die explizite Modellierung verwendete spezielle HMM für unbekannte Wörter, die jeweils auf allen Wörtern einer bestimmten Länge trainiert wurden. Die Leistungsfähigkeit des Ansatzes hängt dabei wesentlich von der Anzahl der verwendeten Modelle und ihrer phonetischen Ähnlichkeit zu den unbekannten Wörtern ab. Die implizite Methode basiert auf erkenner-internem Wissen, Konfidenzmerkmalen, die unbekannte Wörter von Wörtern des Vokabulars unterscheiden können. Die vorgestellten impliziten Modelle erreichen eine signifikant bessere Detektionsrate und Präzision als die expliziten Modelle. Sie sind in der Lage, etwa jedes dritte unbekannte Wort zu detektieren, wobei jede zweite OOV-Hypothese korrekt ist. Durch hybride Ansätze, die beide Methoden kombinierten, konnte zwar die Präzision um etwa 5% erhöht werden, allerdings wurde damit keine signifikante Steigerung der Detektion erreicht. Die größten Verbesserungen der Detektionsrate werden durch eine Verfeinerung der expliziten Modelle erwartet, indem beispielsweise die Wortanfänge genauer modelliert werden. Desweiteren wird daran gearbeitet, präzisere explizite OOV-Modelle für bestimmte Klassen von Wörtern zu trainieren, z.B. Städte- oder Straßennamen.

Literatur

[1] Fetter, P.: Detection and Transcription of Out-Of-Vocabulary Words in Continuous Speech Recognition, Dissertation, TU Berlin, 1998.

[2] Shriberg, E.E.: Preliminaries to a Theory of Speech Disfluencies, Dissertation, University of California, Berkeley, USA, 1994.

[3] Hetherington, I.L.: A Characterization of the Problem of New, Out-Of-Vocabulary Words in Continuous-Speech Recognition and Understanding, Dissertation, MIT, Cambridge, USA, 1995.

[4] Berton, A. und Fetter, P. und Regel-Brietzmann, P.: Compound Words in Large-Vocabulary German Speech Recognition Systems, In Proc. ICSLP'96, Vol. 3, Philadelphia, USA, 1996.

[5] Asadi, A. und Schwartz, R. und Makhoul, J Automatic Detection of New Words in a Large Vocabulary Continuous Speech Recognition System, In Proc. ICASSP'90, Albuquerque, USA, 1990.

[6] Chase, L.L.: Error-Responsive Feedback Mechanisms for Speech Recognizers, Dissertation, Carnegie Mellon University, Pittsburgh, USA, 1997.

[7] Fetter, P. und Class, F. und Haiber, U. und Regel-Brietzmann, P.: Detection of Unknown Words in Spontaneous Speech, In Proc. EuroSpeech'95, Vol. 3, Madrid, Spanien, 1995.

[8] Berton, A.: Konfidenzmaße und deren Anwendungen in der automatischen Sprachverarbeitung, Dissertation, BTU Cottbus, 2004.

[9] Class, F. und Kaltenmeier, A. und Regel-Brietzmann, P.: Optimization of an HMM-Based Continuous Speech Recognizer. In Proc. EuroSpeech’93, Berlin, 1993.

[10] Wahlster, W.: Verbmobil -- Translation of Face-to-Face Dialogs, In Proc. EuroSpeech'93, Berlin, 1993.