

SPRACHMODELL-BERECHNUNG BEIM ÜBERGANG AUF EINE NEUE ANWENDUNG

Johann Adelhardt, Elmar Nöth, Georg Stemmer, Heinrich Niemann

Universität Erlangen-Nürnberg, Lehrstuhl für Mustererkennung,

Martensstr. 3,

D-91058 Erlangen, Germany

adelhardt@informatik.uni-erlangen.de

Kurzfassung: In diesem Beitrag wird die Integration eines n-Gramm-Sprachmodells einer Anwendung in eine neue Anwendung vorgestellt, die Dialoge von anderen Domänen verarbeitet. In beiden Anwendungen wird das Sprachmodell zur Klassifikation von syntaktisch-prosodischen M-Grenzen bei der Verarbeitung von kontinuierlicher, spontaner Sprache eingesetzt. Wegen der Verarbeitung von spontaner Sprache wird zur Klassifikation ein kategorienbasiertes Sprachmodell verwendet, für das ein Kategoriensystem auf Basis eines schon bestehenden der ersten Anwendung entwickelt wurde. Wegen des mit der Entwicklung verbundenen Aufwands, der für weitere neu zu verarbeitende Domänen teilweise wiederholt werden muss, wird für die Suche nach einem Kategoriensystem für das Sprachmodell ein iterativer Algorithmus vorgeschlagen, der ein für diese Anwendung geeignetes Kategoriensystem finden soll. Motiviert wird die Automatisierung der Kategorienauswahl für das Sprachmodell durch die fortlaufende Vergrößerung der Menge der Trainingsdialoge in weiteren Domänen im Laufe des Projekts, die ansonsten mit einer steten manuellen Weiterentwicklung des Kategoriensystems verbunden ist. Aus dem durch den Algorithmus gewonnenen Sprachmodell und dem vorhandenen wird dann per Interpolation in einem weiteren Schritt ein neues Sprachmodell gewonnen, um die Daten der vorhandenen Anwendung ebenfalls zu nutzen. Der Klassifikator, der mit dem neuen Korpus alleine trainiert und gefunden wurde, erreichte für das Zwei-Klassen-Problem "Grenze/Nicht-Grenze" 89% Erkennungsrate. Die Interpolation der beiden Sprachmodelle lieferte 91% Erkennungsrate.

1 Einleitung

In der Sprachverarbeitung stehen meist zu wenig Trainingsdaten für das Training der Klassifikatoren zur Verfügung. Im Verbundprojekt SmartKom¹ (SK) [7], einem Nachfolgeprojekt des Verbundprojekts Verbmobil (VM) [6], wurden deswegen insbesondere am Anfang des Projekts Daten und Ergebnisse aus VM verwendet, um das Datenproblem zu verringern. In VM wurden Dialoge zur Terminplanung mit dem Ziel der Übersetzung von spontaner Sprache verarbeitet. Das Projekt VM wird in [6] umfassend vorgestellt. In SK wird spontane Sprache in einem multimodalen Dialogsystem verarbeitet, das mehrere Domänen beherrschen soll wie z.B. Kinoauskunft und -reservierung, Fernsehprogrammauskunft, Fernseh- und Videorecorderansteuerung,

¹Die diesem Beitrag zugrunde liegenden Arbeiten wurden mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) unter dem Förderkennzeichen 01 IL 905K7 im Rahmen des Verbundprojekts SmartKom gefördert. Die Verantwortung für den Inhalt dieses Beitrags liegt bei den Autoren. Wir danken allen SmartKom-Teilnehmern, die die prosodische Information etikettiert bzw. in ihre Analysemodule integriert haben, für die ausserordentlich gute Zusammenarbeit.

- d007_pkd_004_ANB: da ist nichts für mich dabei M3 dann möchte' ich doch lieber ins Kino gehen (M3)
- w004_pkd_030_AAB: wie kann ich dort fahren M3 also mit der \$U-Bahn oder mit der Straßenbahn (M3)
- w004_pkd_032_AAB: aha M3 um wieviel Uhr beginnt die den Film M3 um siebzehn Uhr nein neunzehn Uhr (M3)

Abbildung 1 - Beispiele für Benutzeräußerungen in SK-Dialogen mit äußerungsinternen M3-Grenzen

Routenplanung und Touristeninformation. Das Dialogsystem soll dem Benutzer natürliche Interaktion mit dem Rechner über seine Mimik, Gestik und Sprache erlauben, auf die das System mit Information via Sprach- und graphischer Display-Ausgabe reagiert und mit dem Benutzer kommuniziert, also auch Fragen stellt. Weitere Eigenschaften von SK werden in [7] erläutert.

In diesem Beitrag wird ein Aspekt der Sprachverarbeitung behandelt, die Segmentierung mit syntaktisch-prosodischen M-Grenzen [2]. Die Verarbeitung der prosodischen Merkmale der Sprache erfolgt im Prosodiemodul [1], das für VM entwickelt wurde und in SK weiterentwickelt wird. Das Prosodiemodul berechnet verschiedene prosodische Merkmale und wird zur Segmentierung von Benutzeräußerungen, aber auch zur Bestimmung des Benutzerzustandes eingesetzt. Syntaktisch-prosodische M-Grenzen werden im Prosodiemodul mit Hilfe von Sprachmodellen bestimmt. Aus Darstellungsgründen erfolgt die Berechnung der M-Grenzen hier auf der besten Wortkette. In VM und SK selbst erfolgt die Berechnung im Worthypothesengraphen [4], die Übertragung der Berechnung der Grenzwahrscheinlichkeit ist straight-forward.

In Kapitel 2 werden M-Grenzen, Sprachmodelle und Kategoriensysteme vorgestellt. In Kapitel 3 wird die Bestimmung eines Kategoriensystems für eine neue Anwendung beschrieben, bevor dann in Kapitel 4 ein Verfahren zur automatischen Auswahl von sog. Handkategorien vorgestellt wird. Auf die Experimente und Ergebnisse wird in Kapitel 5 eingegangen, daran schließt sich noch ein Ausblick und die Zusammenfassung an.

2 M-Grenzen, Sprachmodelle und Kategoriensysteme

Ein Ziel der Verarbeitung von spontaner Sprache ist die Zerlegung des Gesprochenen in syntaktische Einheiten, um davon ausgehend den Inhalt zu erfassen. In VM wurde hierfür ein speziell auf spontane Sprache zugeschnittenes syntaktisch-prosodisches Label-System entwickelt, die M-Grenzen [2]. Hiermit wurde der Beobachtung Rechnung getragen, dass syntaktische Grenzen oft prosodisch markiert sind.

Bei der Annotation des Trainingsmaterials wird jede Wortgrenze mit einer von 25 Grenzklassen beschriftet. Für die Segmentierung mit M-Grenzen werden die 25 Grenzklassen des Label-Systems auf die beiden Oberklassen *starke syntaktisch-prosodische Grenze* (M3) vs. *keine syntaktisch-prosodische Grenze* (M0) abgebildet, für Ergebnisse einer Abbildung auf fünf Oberklassen siehe [1, 4]. Starke Grenzen treten z.B. zwischen Hauptsätzen, zwischen Haupt- und Nebensätzen, zwischen freien Phrasen und bei Links- und Rechtsversetzung auf.

Zur Darstellung von syntaktisch-prosodischen Grenzen werden beispielhaft in Abbildung 1 einige Dialogausschnitte aus SK-Domänen gezeigt, die im Rahmen von Wizard-of-Oz-Experimenten aufgenommen wurden. Die Dialogausschnitte sind am Anfang alle mit einer Beschriftung versehen, die einzelne Dialoge charakterisiert und den Experimenten zuordenbar macht. Die Äußerungen wurden manuell mit M-Grenzen gelabelt, die in Abbildung 1 bereits auf die Oberklasse M3 und M0 abgebildet wurden. Jede Wortgrenze, die nicht als M3-Grenze annotiert ist, wird implizit als M0-Grenze betrachtet. Durch die Labelung wird eine Segmentierung

	Types	Tokens	Kat 20	Handkategorien	Ø Wortzahl je HK
VM	ca. 15000	ca. 403000	ca. 1000	158	8
SK Man.	ca. 2500	ca. 19500	85	173	7
SK Autom.	ca. 2500	ca. 19500	85	5 + 76	10

Tabelle 1 - Lexikon und Handkategorien in VM und SK

der Äußerung erreicht, um Sinneinheiten zu erhalten, die in der geschriebenen Sprache vielfach durch Zeichensetzung ausgedrückt würden. Der triviale Fall, die M3-Grenze am Ende der Äußerung, wie sie oben in Klammern angegeben ist, wird (auch bei den Experimenten in Kapitel 5) nicht weiter betrachtet.

Bei der Klassifikation von M-Grenzen werden in unserem Modul Wortketten der Form

$$w_{i-1}w_iM^jw_{i+1}w_{i+2}, \quad j = 0, 3 \quad (1)$$

betrachtet und abgeschätzt, ob zwischen w_i und w_{i+1} die Grenze M0 oder die Grenze M3 auftritt. Die M-Grenzen werden mit Hilfe von Sprachmodellen bestimmt. Sprachmodelle sind ein statistisches Verfahren zur Modellierung von Wortfolgen in einer Benutzeräußerung und werden auch als n-Gramme bezeichnet.

Wie bei der Verarbeitung von spontaner Sprache üblich, werden kategorienbasierte Sprachmodelle verwendet, die zusätzliche Information aus einem Lexikon, dem Kategoriensystem, benutzen. Das Kategoriensystem enthält die Information, in welche Kategorie jedes Wort des Lexikons eingeordnet wird. Jedem Wort wird nur eine Kategorie zugeordnet, so dass man ein System disjunkter Kategorien ohne überlappende Kategorien erhält (vgl. [5]). Können einem Wort mehrere Kategorien zugeordnet werden, dann wird die Kategorie gewählt, die für SK (bzw. VM) typisch ist. Für die Klassifikation mit kategorienbasierten Sprachmodellen muß die jeweils betrachtete Wortfolge noch in die zugehörige Kategorienfolge umgesetzt und die Wahrscheinlichkeit für das Wort beim Auftreten der Kategorie berücksichtigt werden [8].

Die zu der Wortkette in (1) gehörige 5-Gramm-Wahrscheinlichkeit wird in (2) durch die Trigramm-Wahrscheinlichkeit für das Auftreten der Wortketten und der Wahrscheinlichkeit für die Kategoriezugehörigkeit der Wörter approximiert. Der Klassifikator berechnet die Wahrscheinlichkeit für beide M-Grenzen und entscheidet sich für die Klasse mit größerer Wahrscheinlichkeit.

$$P(w_{i-1}w_iM^jw_{i+1}w_{i+2}) = P(M^j|c_{i-1}, c_i) \cdot P(c_{i+1}|c_i, M^j) \cdot P(c_{i+2}|M^j, c_{i+1}) \cdot P(w_{i-1}|c_{i-1}) \cdot P(w_i|c_i) \cdot P(w_{i+1}|c_{i+1}) \cdot P(w_{i+2}|c_{i+2}), \quad j = 0, 3 \quad (2)$$

Die Segmentierung wird mit einem kategorienbasierten Sprachmodell gemacht, das trainiert werden muß. Für das kategorienbasierte Sprachmodell ist ein noch zu bestimmendes Kategoriensystem erforderlich. Darauf wird im nächsten Abschnitt eingegangen.

3 Ein Kategoriensystem für eine neue Anwendung

Im Folgenden wird kurz auf die beiden Sprachkorpora als Basis für die Experimente eingegangen. Die Entwicklung eines anwendungsspezifischen Sprachkorpus für die Dialogverarbeitung und auch für die Verarbeitung der Prosodie ist für das Projekt VM abgeschlossen. Das Korpus für SK, das mehrere neue Domänen abdeckt und durch die neuen Interaktionsmöglichkeiten mit dem System eine neue Dialogführung ermöglicht, wird momentan noch erweitert. Die beiden Korpora werden in Tabelle 1 einander gegenüber gestellt. Es fällt auf, dass das VM-Korpus

sowohl hinsichtlich Umfangs des Lexikons als auch der Beispieldialoge wesentlich größer ist. Derzeit stehen für SK etwa 5% des Datenumfangs von VM zur Verfügung. Das SK-Korpus wird jedoch nicht so umfangreich werden wie jenes von VM. Damit wird motiviert, das VM-Korpus weiter zu verwenden, zumal es zu einer Verbesserung der Erkennungsraten des SK-Systems beiträgt, und die neuen SK-Daten schrittweise in das Training der Klassifikatoren für das System zu integrieren. Die Angabe Man. bzw. Autom. in der Tabelle 1 bezieht sich wie die Angabe der sog. Handkategorien auf die unten erläuterte Verwendung eines sog. Handkategoriensystems. Mit \emptyset *Wortzahl je HK* ist die durchschnittliche Wortzahl je Handkategorie gemeint. Die Anzahl in Spalte *Kat 20* der Tabelle 1 bezieht sich auf die Anzahl der Wörter, die aufgrund ihrer Häufigkeit, hier 20 Beobachtungen, eine eigene Kategorie zugewiesen erhalten.

3.1 Entwurf eines Kategoriensystems

Beim Entwurf eines Kategoriensystems kann man im einfachsten Fall jedem Wort eine eigene Kategorie zuweisen, was natürlich keinen Informationsgewinn bringt. Bei der Einordnung eines Wortes werden vier Kriterien verwendet, die Häufigkeit des Wortes im Korpus, semantische und syntaktische Gründe und anwendungsabhängige Gründe, wie sie auch [8] motiviert (vgl. nachfolgendes Bsp. mit der Kategorie Z AHL12).

Einen Teil der Generierung eines Kategoriensystems kann man vollständig automatisieren, das sind die über die Häufigkeit des Wortvorkommens bestimmten Kategorien. Der restliche Teil muß überwiegend manuell erledigt werden, wobei man hier sich z.B. durch automatisierte Betrachtung des Wortendes via der letzten n Buchstaben aller Wörter und einer darauf beruhenden Vorauswahl helfen kann. Damit findet man beispielsweise den Teil der Straßennamen, die auf *weg*, *straße* oder *gasse* enden.

Im nicht automatisierten Teil werden zur Gruppierung der Wörter die Handkategorien bestimmt, dabei handelt es sich um die manuelle Zuordnung von Wörtern zu einer Kategorie im Lexikon, die Kategorie ist also zusätzliche Information über das Wort. Hier hat man Wörter, die wichtig erscheinen, aber nicht so häufig vorkommen, dass ihnen durch die Vorgabe automatisch eine eigene Kategorie zugewiesen würde, die auf diesem Weg manuell eine eigene Kategorie im Handkategoriensystem erhalten. Wörter der selben Bedeutung, z.B. Straßennamen, werden in eine Kategorie eingeordnet, weil sie einzeln nicht häufig genug für eine eigene Kategorie vorkommen, aber insgesamt im Text in der gleichen Art auftreten und deshalb ähnlich behandelt werden. Straßennamen unterscheiden sich jedoch teilweise im Geschlecht. Dem wird Rechnung getragen durch Aufteilung der Straßennamen in mehrere Kategorien, z.B. Namen, die auf *straße*, die auf *gasse* und die auf *weg* enden. Eine weitere Gruppierung erfolgte in VM beispielsweise für Zahlen in Z AHL12 (1,2, ..., 12), Z AHL24 (13, 14, ..., 24), Z AHL59 (25, ..., 59) entsprechend dem Wortvorkommen für Uhrzeiten. In die Handkategorien werden auch viele Verben, Adjektive oder Substantive eingeordnet, die öfters auftreten und Gemeinsamkeiten aufweisen. Verschiedene, gebeugte Formen des selben Wortes werden aber durchaus unterschiedlichen Kategorien zugeordnet, weil der Kontext eines Wortes dessen Auftreten bestimmt [3].

Eine besondere Kategorie ist RAREBIRDS, der die selten auftretenden Wörter zugewiesen werden, für welche eben nicht durch die Häufigkeit oder durch manuelle Zuordnung eine Kategorie im Lexikon festgelegt wurde.

²Die Kategorie MONAT z.B. enthält 14 Elemente, weil für die Monate Juni und Juli als alternative Sprechweise auch *Juno* und *Julei* angegeben ist.

VM	Kategorie	VM	Kategorie	SK	Kategorie	SK	Kategorie
355	NACHNAME	29	NAME	359	NACHNAME	29	TERMIN
128	ORTSNAME	27	FEIERTAG	131	ORTSNAME	28	ZAHL
53	VORNAME	26	ZAHL	58	VORNAME	26	FEIERTAG
36	ZAHL31STEN	19	MANN	44	NAME	23	FILMTITEL
35	ZAHL31STE	15	HOTEL	36	ZAHL31STEN	19	MANN
34	ZAHL31STER	14	ZAHL24	35	ZAHL31STE	16	HOTEL
33	ORT	14	TREFFEN	35	ORT	14	ZAHL24
30	LETTER	14 ²	MONAT	34	ZAHL31STER	14	TREFFEN
29	TERMIN	12	ZAHL12	30	LETTER	14 ²	MONAT

Tabelle 2 - Die häufigsten Handkategorien in VM und SK, sowie die Anzahl der darin enthaltenen Wörter

3.2 Übergang auf eine neue Domäne

Bei der Entwicklung eines Kategoriensystems wurden drei Gruppen von Kategorien unterschieden. Die erste Gruppe sind die Wörter, die aufgrund ihrer Häufigkeit in eine Kategorie eingeordnet werden. Das sind die im vorherigen Abschnitt erläuterten Kategorien *Kat 20*. Die zweite Gruppe sind die durch die Handkategorien bestimmten Wörter und die restlichen Wörter, denen die Kategorie RAREBIRDS zugewiesen wird, fallen in die dritte Gruppe.

In Tabelle 1 wird neben der Anzahl der VM-Kategorien auch die Anzahl der SK-Handkategorien gezeigt, wobei es sich bei SK Man. um ein VM-Kategoriensystem handelt, das durch den SK-Anteil der Wörter ergänzt wurde. Tabelle 2 zeigt die häufigsten Handkategorien der beiden Kategoriensysteme.

Beim Übergang auf eine neue Anwendung ist wegen des neuen Wortschatzes die Inspektion der bisherigen Handkategorien erforderlich. Handkategorien mit wenigen Beobachtungen im Korpus müssen manuell untersucht werden. Teilweise sind durch die neue Anwendung 'alte' Handkategorien überflüssig oder hinderlich in dem Sinn, dass sie sogar zu einer Verschlechterung der Erkennungsrate führen. Die Frage ist zu klären, wieviele Kategorien wieviele Wörter im neuen Korpus abdecken. Bei neu eingeführten Handkategorien muß vor allem geprüft werden, ob sie sich hinsichtlich der Klassifizierung so verhalten, wie das erwartet wird. Man prüft also, wie sich durch Hinzunahme von neuen Handkategorien die Erkennungsrate im Klassifikationsprozeß verhält.

Hier werden für das neue Korpus ins Lexikon eingefügten Wörtern Handkategorien zugewiesen und die Überprüfung hinsichtlich günstiger oder ungünstiger Einordnung in Handkategorien durch den nachfolgend vorgestellten Algorithmus übernommen. Ob eine Handkategorie günstig oder ungünstig ist, wird durch das Kriterium Erkennungsrate beurteilt.

4 Automatische Aussuche von Handkategorien

Wie gut wird man durch die automatische Aussuche von Handkategorien, wenn man diese aus einem Pool von möglichen Handkategorien auswählen kann? Der Algorithmus in Abbildung 2 zeigt das Verfahren zur Bestimmung des Handkategoriensystems für das Sprachmodell auf. Begonnen wird mit dem Lesen der Initialwerte für die beiden Mengen aus Konfigurationsdateien, den sog. Basis-Handkategorien, die auf alle Fälle als Handkategorien in den Sprachmodellen verwendet werden, und der Menge der möglichen Handkategorien³.

³Neue Wörter erhalten im Lexikon eine semantisch motivierte Handkategorie, die für die Verarbeitung im SK-System verwendet wird.

READ Basis-Handkategorien	
READ Pool der möglichen Handkategorien	
verbesserung := true	
WHILE verbesserung = true	
Bestimme Basiserkennungsrate für Verbesserung	
verbesserung := false	
FORALL $k \in$ Pool der möglichen Handkategorien	
Bestimme ein Sprachmodell mit k als zusätzlicher Kategorie in den Handkategorien	
Bestimme Erkennungsrate mit Validierungsmenge	
IF	Erkennungsrate > Basiserkennungsrate
THEN	Markiere k
IF	Existieren markierte Handkategorien
THEN	verbesserung := true
	Entferne markierte Handkategorien aus Pool und füge sie zur Menge der Basis-Handkategorien hinzu

Abbildung 2 - Algorithmus zur Bestimmung des besten Kategoriensystems

In einer Schleife wird dann überprüft, ob durch Hinzunahme jeweils einzelner Kategorien zu den Handkategorien die Erkennungsrate erhöht werden kann. Dazu wird im ersten Schritt der Schleife festgelegt, welche Erkennungsrate im aktuellen Schritt das Maß für den Vergleich ist. Dies erfolgt mit einem Sprachmodell und einer “Dummy-Handkategorie”, die ein Wort enthält, das sicher nicht im Vokabular des Lexikons und dadurch impliziert auch nicht in der Trainings-, Validierungs- oder Testmenge vorkommt. Danach wird für alle Elemente der möglichen Handkategorien in einem eigenen Durchlauf probiert, durch Hinzunahme dieses Elements eine Erhöhung der Erkennungsrate zu erreichen. Wird für ein Element eine Verbesserung erreicht, so wird dieses Element markiert.

Am Ende einer while-Iteration werden die markierten Elemente aus der Menge der möglichen Handkategorien entfernt und in die Menge der Basis-Handkategorien eingefügt, sofern in dem Durchlauf Elemente markiert wurden. Das Abbruchkriterium der Iteration ist erreicht, wenn in einem Schritt keine Elemente markiert wurden. Die Iteration terminiert sicher, weil die endliche Menge der möglichen Handkategorien streng monoton verkleinert wird.

Nach dem Durchlauf steht ein geeignetes Handkategoriensystem und auch das Sprachmodell mit der besten Erkennungsrate fest. Mit dem Sprachmodell wird anschließend in einem folgenden Schritt mit einer Testmenge, die weder in der Trainings- noch in der Validierungsmenge enthalten ist, die Erkennungsrate des Experiments bestimmt. In den verschiedenen Durchläufen, die in den Experimenten für diesen Beitrag durchgeführt wurden, waren durchschnittlich vier bis fünf Iterationen nötig, bis sich die Erkennungsrate nicht mehr verbesserte. In Ausnahmefällen waren es sieben oder acht Iterationen.

5 Experimente und Ergebnisse

Das Vorgehen ist in die folgenden Schritte gegliedert: manuelle Bestimmung einer Basismenge von Handkategorien und einer Menge von möglichen Handkategorien, Training der Sprachmodelle unter Berücksichtigung der Handkategorien und Suche nach der Menge von Handkategorien mit maximaler Gesamterkennungsrate und Interpolation des besten SK-Sprachmodells mit dem VM-Sprachmodell.

Die Ergebnisse beruhen auf Klassifikation mit einem Trigramm-Sprachmodell, siehe (2). Das

	VM	SK Man.	SK Autom.	VM u. SK Autom. mit rat. Interpol.
Erkennungsrate in %	95	89	89	91

Tabelle 3 - Erkennungsraten für die Sprachmodelle

Kategoriensystem wurde jeweils mit den Handkategorien gebildet, den Wörtern, die nicht durch die Handkategorien einer Kategorie zugeordnet waren, wurde eine eigene Kategorie zugeordnet, wenn sie mindestens 20 mal im Trainingskorpus vorkamen, sonst wurde das entsprechende Wort der Kategorie RAREBIRDS zugeordnet. Das resultierende Sprachmodell wurde per rationaler Interpolation [5] mit dem VM-Sprachmodell verknüpft.

5.1 Experimente mit manuell ausgesuchten Handkategorien

Zwei prinzipiell verschiedene Experimente wurden durchgeführt. Für das erste Experiment wurde ein Handkategoriensystem manuell für dieses Korpus entwickelt. Die Entwicklung bzw. der Aufwand besteht darin, das Vokabular und die Häufigkeit der einzelnen Wörter zu analysieren und nach Wörtern zu suchen, die gleichen Kategorien zugeordnet werden sollen. Dann werden manuell einzelne Tests, sprich Trainingsläufe, durchgeführt und damit ausprobiert, ob verschiedene Kategorien zu einer Verbesserung führen. Mit einem auf das Korpus basierenden und so entworfenen Kategoriensystem wurden 89% Erkennungsrate erreicht.

5.2 Experimente mit automatischer Suche nach Handkategorien

Für das zweite Experiment wurde eine Vorschlagsliste für Handkategorien entwickelt. Basierend auf dieser Vorschlagsliste wurde dann die Iteration gestartet und in der Iteration nach einem Handkategoriensystem gesucht, welches für dieses Korpus geeignet ist. Das Ergebnis der Iteration ist ein Handkategoriensystem, mit dem ebenfalls 89% Erkennungsrate erreicht wurden. Für die Erzeugung dieses Kategoriensystems war allerdings wesentlich weniger Zeitaufwand erforderlich, da die Auswahlsschritte automatisch durchgeführt wurden.

Mit dem Basissystem von 5 Handkategorien und der restlichen Einordnung der Wörter gemäß der Häufigkeit in Einzelkategorien und RAREBIRDS wurden Erkennungsraten von über 84% erreicht. Hier wurden die Wörter, die mindestens 20 mal im Trainingskorpus gezählt wurden, in Kategorien mit nur einem Wort und die restlichen in die eine Kategorie RAREBIRDS eingeordnet.

Das Ergebnis der Iterationen des Algorithmus ist ein Sprachmodell, das mit dem als global betrachteten Verbmobil-Sprachmodell interpoliert wird. Das Verfahren führt zur Verbesserung der Erkennungsrate für das Zweiklassenproblem mit M0/M3-Grenzen. Die rationale Interpolation mit dem VM-Sprachmodell als globalem Sprachmodell führt zu einer Verbesserung von ca. 2% auf 91%. In VM wurden Erkennungsraten von 95% erreicht, VM hatte aber auch keine so breite Streuung der Domänen und wesentlich mehr Trainingsmaterial. In Tabelle 3 werden die erzielten Erkennungsraten einander nochmals gegenüber gestellt.

6 Ausblick

Für weitere Arbeiten kann bei der automatischen Kategorienauswahl beispielsweise für die Auswahl der Kategorien in den einzelnen Iterationsschritten statt gleichzeitiger Hinzunahme aller zur Verbesserung führenden Handkategorien in einem Schritt auch nur eine einzelne oder ein

Teil der Kategorien hinzu genommen werden. Die hinzuzunehmenden Handkategorien könnte man anhand der jeweils durch die Kategorie erreichten Verbesserung der Erkennungsrate auswählen. Aufgefallen war, dass bei den Kategorien, die zu einer Verbesserung führten, manche Kategorien eine stärkere Verbesserung ergaben als andere. Wechselwirkungen zwischen den einzelnen Kategorien sind zu untersuchen. Bei der Bildung des Kategoriensystems sind in den einzelnen Iterationsschritten des Algorithmus andere Strategien denkbar. Die Iteration kann mit dem kompletten Pool von Vorschlägen für die Handkategorien gestartet werden. Durch selektives Reduzieren der Menge der Handkategorien entsprechend der Verbesserung bzw. Verschlechterung der Erkennungsrate kann die Auswahl der Kategorien durchgeführt und dadurch Rückschlüsse auf Wechselwirkungen zwischen den Kategorien gezogen werden. Untersuchungen zur Stabilität der Erkennungsrate der beiden Sprachmodelle wurden noch nicht durchgeführt.

7 Zusammenfassung

Ziel der hier vorgestellten Arbeiten war die Erstellung eines Sprachmodells mit unzureichenden Daten von einer neuen Anwendung unter Verwendung von Trainingsdaten eines bestehenden Sprachmodells. Hierfür ist eine Zuordnung der beobachteten Wörter zu Kategorien notwendig. Es wurde ein Verfahren vorgestellt, das die verwendeten Kategorien automatisch aus einer Menge von Vorschlägen auswählt und Erkennungsraten liefert, wie sie auch mit einem System, das durch aufwendige manuelle Kategorienauswahl bestimmt wurde, erzielt werden. Damit wurde eine Erkennungsrate von 89% erreicht. Eine weitere Verbesserung der Erkennungsrate auf 91% ergab sich durch Interpolation der beiden Sprachmodelle.

Literatur

- [1] A. Batliner, A. Buckow, H. Niemann, E. Nöth, and V. Warnke. The Prosody Module. In Wahlster [6], pages 106–121.
- [2] A. Batliner, R. Kompe, A. Kießling, M. Mast, H. Niemann, and E. Nöth. M = Syntax + Prosody: A syntactic–prosodic labelling scheme for large spontaneous speech databases. *Speech Communication*, 25(4):193–222, 1998.
- [3] James H. Martin Daniel Jurafsky. *Speech and Language Processing*. Prentice Hall, New Jersey, 2000.
- [4] R. Kompe. *Prosody in Speech Understanding Systems*. Lecture Notes for Artificial Intelligence. Springer–Verlag, Berlin, 1997.
- [5] E. G. Schukat-Talamazzini. *Automatische Spracherkennung – Grundlagen, statistische Modelle und effiziente Algorithmen*. Vieweg, Braunschweig, 1995.
- [6] W. Wahlster, editor. *Verbmobil: Foundations of Speech-to-Speech Translations*. Springer, Berlin, 2000.
- [7] W. Wahlster, N. Reithinger, and A. Blocher. SmartKom: Multimodal Communication with a Life-like Character. In *Proc. European Conf. on Speech Communication and Technology*, volume 3, pages 1547–1550, Aalborg, 2001.
- [8] Volker Warnke. *Integrierte Segmentierung und Klassifikation von Äußerungen und Dialogakten mit heterogenen Wissensquellen*. Studien zur Mustererkennung. Logos Verlag, Berlin, erscheint voraussichtl. noch 2002.