

Informatik-Fachberichte 141

Herausgegeben von W. Brauer
im Auftrag der Gesellschaft für Informatik (GI)

Ernst Günter Schukat-Talamazzini

Generierung von Worthypothesen in kontinuierlicher Sprache



Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo

Autor

Ernst Günter Schukat-Talamazzini
Lehrstuhl für Informatik 5 (Mustererkennung)
Universität Erlangen-Nürnberg
Martensstr. 3, 8520 Erlangen

CR Subject Classifications (1987): I.5, I.2.7, I.2.8, I.2.6

ISBN-13:978-3-540-17998-6 e-ISBN-13:978-3-642-72795-5

DOI: 10.1007/978-3-642-72795-5

CIP-Kurztitelaufnahme der Deutschen Bibliothek. Schukat-Talamazzini, Ernst Günter: Generierung von Worthypothesen in kontinuierlicher Sprache / Ernst Günter Schukat-Talamazzini. – Berlin; Heidelberg; New York; London; Paris; Tokyo: Springer, 1987. (Informatik-Fachberichte; 141)
ISBN-13:978-3-540-17998-6

NE: GT

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der Fassung vom 24. Juni 1985 zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

© by Springer-Verlag Berlin Heidelberg 1987

Geleitwort

Die automatische Erkennung kontinuierlich gesprochener Sprache bei großem Wortschatz und vielen Sprechern ist einer der zu Zeit im Brennpunkt stehenden Forschungsschwerpunkte der Mustererkennung, speziell der Spracherkennung und des Sprachverstehens. Wie in anderen Bereichen der Mustererkennung auch haben Methoden der Wissensverarbeitung hier ihren festen Platz gefunden. Gesprochene Sprache zu verstehen ist für Menschen keine Schwierigkeit, aber für eine Maschine immer noch nur unter drastischen Einschränkungen möglich. Dieses macht das Problem aus rein wissenschaftlichen Gründen zu einer Herausforderung. Gesprochene Sprache ist auch das bevorzugte Medium menschlicher Kommunikation, so daß die automatische Spracherkennung eine Fülle von Anwendungen in künftigen Informationssystemen eröffnet. Eine Grundvoraussetzung dafür ist offenbar, daß man in der Lage ist, in normaler fließender Rede gesprochene Wörter automatisch zu erkennen. Diesem Problem ist das vorliegende Buch gewidmet.

Zunächst werden die Grundlagen des Vergleichs von Zeitmustern, bekannte Verfahren der Wortpositionierung und Lautähnlichkeiten diskutiert. Diese Zusammenstellung wichtiger Ergebnisse aus der Literatur vermittelt einerseits einen breiten Überblick über die Problematik und die relevanten Lösungsansätze, liefert aber andererseits auch zahlreiche mathematische Einzelheiten in exakter Form, zum Beispiel den Beweis des Satzes von Baum für den *Baum-Welch*-Algorithmus. Aus dieser Diskussion wird dann ein vollständiger Modul zur Generierung von Worthypothesen in kontinuierlicher Sprache entwickelt. Da beim gegenwärtigen Stand der Kunst an jeder Stelle des Sprachsignals mehrere Alternativen für Wörter gefunden werden, werden diese als Worthypothesen bezeichnet. Für die Leistung eines Verfahrens spielt die Bewertung der Worthypothesen eine entscheidende Rolle, wie in der Arbeit ausführlich gezeigt wird. Eine sehr sorgfältige experimentelle Untersuchung der Zuverlässigkeit des Moduls unter verschiedenen Randbedingungen bildet den Abschluß.

Das Buch bietet mit dem allgemeinen Überblick und dem in allen Einzelheiten ausgearbeiteten, in seiner Leistungsfähigkeit im internationalen Vergleich hervorragenden Verfahren zur Worthypothesengenerierung einen wichtigen Beitrag zur Spracherkennung.

H. Niemann

Vorwort

Das vorliegende Buch entstand als Dissertation im Rahmen eines längerfristigen Forschungsprojekts, das seit dem Jahr 1979 am Lehrstuhl für Informatik 5 (Mustererkennung) der Universität Erlangen-Nürnberg verfolgt wird und die Grundlagenforschung zum automatischen Erkennen und Verstehen gesprochener Sprache zum Inhalt hat. Die Arbeiten wurden seitdem vom Bundesministerium für Forschung und Technologie (BMFT) und der Deutschen Forschungsgemeinschaft (DFG) sowohl unmittelbar als auch innerhalb von Verbund- bzw. Schwerpunktprogrammen gefördert.

Gegenstand der nachfolgenden Untersuchungen ist ein wesentlicher Teilaspekt automatischen Sprachverstehens: das Problem der Worterkennung in fließender Rede. Neben einer einführenden, vereinheitlichenden Darstellung unterschiedlicher Ansätze zum Vergleich eindimensionaler (insbesondere *zeitabhängiger*) Muster, der Konzeption eines Worterkennungssystems, das auch einen großen Wortschatz dank einer kompakten Organisation des Aussprachelexikons effizient verarbeiten kann, und den Resultaten zahlreicher vergleichender Untersuchungen zur Erkennungssicherheit bietet der Band zwei neue Verfahren zur Lokalisierung eindimensionaler Muster in einem kontinuierlichen Datenstrom sowie eine (auch experimentelle) Gegenüberstellung statistischer und struktureller Modellierung des Verwechslungsverhaltens automatischer Lautklassifikation, die den betreffenden Originalarbeiten des Autors entstammen.

Da über zahlreiche eigene Untersuchungen zu berichten und eine Fülle einschlägiger Literatur zu verwerthen war, blieb die Diskussion verwandter Fragestellungen wie etwa der Wortkettenerkennung ebenso wie eine ausführliche Darstellung phonetisch oder phonologisch orientierter Gesichtspunkte ausgeklammert, zumal ausgezeichnete Literatur zu diesem Themenkreis bereits vorhanden ist. Unberücksichtigt aus Erwägungen im Sinne inhaltlicher Homogenität heraus blieb auch die Behandlung von Worthypothetisierungsmethoden unter Verwendung assoziativer Lexikonrepräsentationen, die im Verlaufe des Projekts ebenfalls vorgeschlagen und realisiert wurden.

An dieser Stelle möchte ich all jenen herzlich danken, die an der Entstehung dieses Buches durch Unterstützung in vielerlei Formen Anteil hatten, sei es, indem sie Anregungen gaben, kritisch zuhören oder Geduld bewiesen:

— dem Bundesministerium für Forschung und Technologie (BMFT) für die Finanzierung des Forschungsvorhabens im Rahmen des Projekts "*Semantisch-pragmatische Programmierbausteine für ein Frage-Antwort-System mit kontinuierlicher Spracheingabe und Sprachausgabe*"

- dem Leiter des Projekts, Herrn Prof. Dr. H. Niemann. Ihm und Herrn Prof. Dr. E. Paulus gilt überdies mein Dank für die Übernahme der Dissertationsgutachten.
- den Mitarbeitern/innen am Lehrstuhl für Informatik 5 (Mustererkennung) der Universität Erlangen-Nürnberg, insbesondere Frau Dr. A. Brietzmann, Herrn P. Regel und Herrn Dr. R. Mühlfeld für die kollegiale Zusammenarbeit im Rahmen des Spracherkennungsprojekts sowie Herrn Dr. D. Jäpel für seine ständige Diskussionsbereitschaft,
- den Herren A. Cieslik, W. Obermayer und F. Popp für das Erstellen der Abbildungen,
- und schließlich allen Studenten, die im Rahmen einer Diplomarbeit, Studienarbeit oder Hilfskraftbeschäftigung zum Erfolg des Projekts beitrugen, insbesondere den Herren H. Fischer, S. Heunisch, H. Schneider und H.M. Sperber.

E.G. Schukat-Talamazzini

Inhalt

1. Einführung	1
1.1 Generierung von Worthypothesen	1
1.2 Übersicht	3
2. Vergleich von Zeitmustern	6
2.1 Abstand zwischen Zeichenketten	8
2.2 Dynamische Zeitverzerrung	11
2.3 Markoffmodelle	13
2.3.1 Markoffketten und -modelle	14
2.3.2 Schätzung der Modellparameter	17
2.3.3 Viterbi-Training	18
2.3.4 Baum-Welch-Algorithmus	19
2.3.5 Problemspezifische Modifikationen	24
3. Positionierung von Wörtern im Sprachsignal	27
3.1 Detektion von Schlüsselwörtern	28
3.2 Einige Verfahren zur Positionierung	29
3.3 Vertikale Summation	33
3.4 Invertiertes Markoffmodell (IMM)	38
4. Lautähnlichkeit und -verwechslung	40
4.1 Ähnlichkeitsrelationen zwischen Lauten	40
4.2 Lautverwechslungswahrscheinlichkeiten	42
4.2.1 Kanalmodell für die akustisch-phonetische Erkennung	42
4.2.2 Das Markoffmodell für den Kanal	43
4.2.3 Der Schätzalgorithmus	46
4.2.4 Experimentelle Ergebnisse	47
4.3 Lautoberklassen	50
4.3.1 Oberklassen mit maximaler Transinformation	50
4.3.2 Oberklassen und Kohorten	55
4.4 Lautverwechslungsmatrix nach Regeln	57
5. Ein Modul zur Generierung von Worthypothesen	59
5.1 Die Repräsentation des Lexikons als Baum	59
5.1.1 Linguistische Datenbasis und Lautumschrift	59
5.1.2 Lexikonbaum, Profil und Suche	61
5.1.3 Ein Präprozessor für das Aussprachelexikon	65
5.1.4 Beispiele	66

5.2	Eine parametrisierte Familie von Worterkennungsalgorithmen	68
5.2.1	Segmentweise Verwechslung	68
5.2.2	Rekursionsformel für den Profilvektor	70
5.2.3	Baumentwicklung	72
5.2.4	Hypothesengenerierung	72
5.2.4.1	Positionswahl	72
5.2.4.2	Anfangspunkt	73
5.2.4.3	Kohortenhypthesen	74
5.2.4.4	Worthypothesen	74
5.2.4.5	Hypothesenbewertung	74
5.2.4.6	Elimination dominierter Hypothesen	74
5.2.4.7	Reduktion	75
5.3	Performanzkriterien für die Worthypothesengenerierung	75
5.3.1	Motivation	75
5.3.2	Definition von Rangmaßen	76
5.3.2.1	Worthypothesen und -vorkommen	76
5.3.2.2	Sätze	78
5.3.2.3	Stichproben	79
5.3.2.4	Separierte Auswertung	79
5.3.3	Einige Anmerkungen	80
6.	Experimente und Resultate zur Worterkennung	82
6.1	Übersicht und experimentelle Voraussetzungen	82
6.2	Das erweiterte Markoffmodell (EMM)	83
6.3	Einige Modifikationen zum EMM	84
6.3.1	Lokale Ähnlichkeit	84
6.3.2	Lokale Transitionen	85
6.3.3	Verknüpfungsoperationen	85
6.3.4	Positionswahl	86
6.4	Einige Vergleichsalgorithmen	87
6.4.1	HMM und DP	87
6.4.2	Zwei Fortentwicklungen	88
6.4.3	Knotenorientierte Modelle	88
6.4.4	Mustervergleich ohne Zeitverzerrung	90
6.5	Lautverwechslungsmatrizen	91
6.5.1	Geschätzte Verwechslungshäufigkeiten	92
6.5.2	Binäre Verwechslungsmatrizen nach Lautoberklassen	93
6.5.3	Verwechslungsmatrix nach Regelsystem	97
6.6	Wortposition	97
6.6.1	Positionsalternativen	97
6.6.2	Anfangs- und Endpunktdetektion	98
6.7	Separierte Auswertungen	99
6.7.1	Wortart	99
6.7.2	Wortlänge	100
6.7.3	Wortposition	101
6.8	Lexikon	101
6.8.1	Standardaussprache und Aussprachevarianten	101
6.8.2	Lexikonumfang	102

6.8.3	Teillexika längerer Wörter	105
6.9	Performanzvergleich	105
6.9.1	Einzelworterkennung	105
6.9.2	Worthypothesen	106
6.9.3	Verifikation	106
7.	Hypothesenbewertung	108
7.1	Problemstellung	108
7.1.1	Qualität und Priorität	108
7.1.2	Hypothesenbewertung als komparatives Urteil	109
7.1.3	Drei kritische Punkte	109
7.2	Verletzung des Optimalitätsprinzips	111
7.3	Längennormalisierung	113
7.4	Das Zweiklassenproblem für Worthypothesen	115
7.5	Die Bewertungsfunktion für das IMM	117
7.6	Wortfolgen und ihre Bewertung	118
Zusammenfassung	120
Bibliographie	123
Anhang	137