

Artificial Intelligence

Künstliche Intelligenz

herausgegeben von Wolfgang Bibel und Walther von Hahn

Künstliche Intelligenz steht hier für das Bemühen um ein Verständnis und um die technische Realisierung intelligenten Verhaltens.

Die Bücher dieser Reihe sollen Wissen aus den Gebieten der Wissensverarbeitung, Wissensrepräsentation, Expertensysteme, Wissenskommunikation (Sprache, Bild, Klang, etc.), Spezialmaschinen und -sprachen sowie Modelle biologischer Systeme und kognitive Modellierung vermitteln.

Auswahl der bisher erschienenen Titel:

Automated Theorem Proving

von Wolfgang Bibel

Prolog

von Ralf Cordes, Rudolf Kruse, Horst Langendörfer, Heinrich Rust

Wissensbasierte Systeme

von Doris Altenkrüger and Wilfried Büttner

Logische Grundlagen der Künstlichen Intelligenz

von Michael R. Genesereth and Nils J. Nilsson

Logische und Funktionale Programmierung

von Ulrich Furbach

Parallelism in Logic

von Franz Kurfeß

Wissensrepräsentation und Inferenz

von Wolfgang Bibel (with St. Hölldobler and T. Schaub)

Deduktive Datenbanken

von Armin B. Cremers, Ulrike Griefahn and Ralf Hinze

Neuronale Netze und Fuzzy-Systeme

von Detlef Nauck, Frank Klawonn and Rudolf Kruse

Fuzzy Systems in Computer Science

von Rudolf Kruse, Jörg Gebhardt and Rainer Palm

Fuzzy Sets and Fuzzy Logic

Foundations of Application – from a Mathematical Point of View
von Siegfried Gottwald

Automatische Spracherkennung

von Ernst Günter Schukat-Talamazzini

Ernst Günter Schukat-Talamazzini

Automatische Spracherkennung

Grundlagen, statistische Modelle
und effiziente Algorithmen



Das in diesem Buch enthaltene Programm-Material ist mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Der Autor, die Reihenherausgeber und der Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Art aus der Benutzung dieses Programm-Materials oder Teilen davon entsteht.

Alle Rechte vorbehalten

© Springer Fachmedien Wiesbaden 1995

Ursprünglich Erschienen bei Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig/Wiesbaden, 1995



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Gedruckt auf säurefreiem Papier

ISSN 090-0699

ISBN 978-3-528-05492-2

ISBN 978-3-322-96180-8 (eBook)

DOI 10.1007/978-3-322-96180-8

Vorwort

Das Thema der vorliegenden Arbeit ist die Problemstellung maschineller Erkennung gesprochener Sprache. Ein wichtiger Forschungszweig auf diesem Gebiet ist gegenwärtig die Entwicklung telefonischer Auskunftsdialogsysteme, die es einem Benutzer beispielsweise gestatten, Zugverbindungsinformationen abzurufen, Platz- oder Hotelreservierungen vorzunehmen oder Bankgeschäfte zu tätigen. Besondere Beachtung finden daher Verfahren zur sprecherunabhängigen Verarbeitung kontinuierlicher Äußerungen, die über einen großen Erkennungswortschatz verfügen, aber den Lösungsraum gleichzeitig durch anwendungsabhängige grammatische Restriktionen wirksam einschränken.

Zur Lösung der gestellten Aufgabe gilt es, der akustischen Variabilität der zu analysierenden Sprachdaten zu begegnen, die durch sprecherspezifische Artikulationseigenschaften und phonetische Verschleifungsphänomene hervorgerufen wird, und ein komplexes Suchproblem zu bewältigen, dessen Ursache im Umfang des Wortschatzes und in der Kombinatorik unbekannter Wortgrenzen liegt.

Nach einigen anfänglichen Versuchen, explizite Funktionsmodelle für den überaus komplexen Erzeugungs- und Wahrnehmungsprozeß gesprochener Sprache zu entwickeln und deren Arbeitsweise auf einem Digitalrechner zu simulieren, dominieren im Bereich der automatischen Spracherkennung heute — nunmehr seit Beginn der 80er Jahre — die wahrscheinlichkeits- oder informationstheoretisch ausgerichteten Lösungsparadigmen. Den statistischen Zugang zum Spracherkennungsproblem kennzeichnet der weitgehende Verzicht auf „handgefertigte“ Aussprache- und Grammatikmodelle zugunsten effizienter Strategien des maschinellen Lernens aus akustischen und textuellen Sprachdaten. Dieser statistische Ansatz hat sich als außerordentlich flexibel und tragfähig erwiesen; seiner ausführlichen Darstellung ist folglich auch der größte Teil dieses Buches gewidmet.

Der mathematische Formalismus statistischer Modelle, ihre parametrische Optimierung und schnelle Suchalgorithmen zur modellgestützten Analyse bilden die inhaltlichen Schwerpunkte dieser Arbeit; vereinigt werden diese Verfahrensbausteine schließlich in der Architektur des ISADORA-Musteranalysesystems, das in einem der umfangreicheren Kapitel dieses Buches detailliert behandelt wird und auch die Grundlage für die experimentellen Resultate des Schlußteils bildet. Die Darstellung des Instrumentariums statistischer Musteranalyse wird ergänzt durch sprachverarbeitungsspezifische Abschnitte über Merkmalgewinnung, phonetische Modellierung und grammatische Modellierung, in deren Verlauf auch zahlreiche innovative Verarbeitungskonzepte aus den Forschungsarbeiten des Autors erstmals in breiterer Form präsentiert werden.

Das Buch wendet sich an Fachleute, die innerhalb des weitgespannten Gebietes *Spracherkennung/Sprachverstehen* tätig sind, aber auch an Studenten, die vertiefte Kenntnisse auf diesen Gebieten erwerben wollen. Vorausgesetzt werden Grundkenntnisse der Wahrscheinlichkeitsrechnung, der linearen Algebra und der Analysis, die in etwa dem Vordiplom eines technisch-naturwissenschaftlichen Studienganges entsprechen. Angesichts der breiten, anwendungsübergreifenden Darstellung statistischer Methoden der Musteranalyse sollten aber auch Systementwickler im Bereich der wissensbasierten Musteranalyse und der Künstlichen Intelligenz von der Lektüre dieses Bandes profitieren können.

Das Buch entstand als Habilitationsschrift im Rahmen eines längerfristigen Forschungsprojekts, das am Lehrstuhl für Mustererkennung (Informatik 5) der Friedrich-Alexander-Universität Erlangen-Nürnberg verfolgt wird und die Grundlagenforschung zum Erkennen und Verstehen gesprochener Sprache zum Inhalt hat. Die Arbeiten wurden vom Bundesministerium für Forschung und Technologie, der Deutschen Forschungsgemeinschaft und der Kommission der Europäischen Gemeinschaft (ESPRIT-Programm) gefördert.

Ich möchte mich an dieser Stelle ganz herzlich bei all jenen bedanken, die das Entstehen dieses Buches und seine Ausarbeitung kritisch und kooperativ begleitet haben. Dazu zählen insbesondere meine Kollegen aus der Sprachverarbeitungsgruppe des Lehrstuhls für Mustererkennung: Dr. Th. Kuhn, Dr. S. Rieck und Herr W. Eckert, die im Rahmen ihrer Tätigkeit für das SUNDIAL-Sprachdialogprojekt entscheidend zum Erfolg der Untersuchungen beitrugen; Dr. E. Nöth, Dr. M. Mast, Herr R. Kompe und Herr A. Kießling, die durch unermüdlige Diskussionsbereitschaft und ideenreiche Zusammenarbeit die fachlich wie persönlich anregende Atmosphäre des Instituts prägten. Mein Dank gilt auch den Herren Prof. Dr. G. Sagerer, G. Fink (Univ. Bielefeld), Prof. Dr. H. Bunke (Univ. Bern), Dr. F. Mihelič, I. Ipšić (Univ. Ljubljana) und G. Schmid (FORWISS, Erlangen) für die fruchtbare Zusammenarbeit bei der Anwendung des ISADORA-Systems in Sprach-, Schrift- Bild- und Meßwertverarbeitung sowie Dr. G. Micca (CSELT, Turin) und G. Niedermair (Siemens AG, München) für die langjährige und ertragreiche Kooperation während des SUNDIAL-Projekts.

Ganz besonders herzlich danke ich Herrn Prof. Dr. H. Niemann, der mein Interesse für das faszinierende Forschungsgebiet der Mustererkennung geweckt und gefördert hat; ihm und Herrn Priv. Doz. Dr. G. Ruske (TU München) danke ich darüberhinaus für die Übernahme der Habilitationsgutachten. Außerordentlich verbunden bin ich auch meinen Kollegen Dr. V. Fischer und Dr. E. Nöth für ihre kritischen Einwände, sachgerechten Korrekturen und süffisanten Anmerkungen anlässlich der Durchsicht des Buchmanuskripts. Nicht zuletzt habe ich den Herausgebern dieser Reihe sowie den Mitarbeitern des Lektorats *Computerfachbuch/Informatik* im Verlag Vieweg für ihre freundliche Unterstützung zu danken.

Ernst Günter Schukat-Talamazzini

Inhaltsverzeichnis

1	Einführung	1
1.1	Mensch-Maschine-Kommunikation	1
1.2	Maschinelle Verarbeitung gesprochener Sprache	3
1.3	Entwurfsparameter spracherkennender Systeme	6
1.4	Warum ist automatische Spracherkennung schwierig?	8
1.5	Geschichte und Forschungsparadigmen	11
1.6	Wahrscheinlichkeitsorientierte Spracherkennung	15
1.7	Übersicht	18
2	Gesprochene Sprache	21
2.1	Artikulation und symbolphonetische Beschreibung	22
2.1.1	Artikulation von Sprachlauten	22
2.1.2	Phonologische Kategorisierung	25
2.1.3	Silbe und Intonation	28
2.1.4	Ausspracheverschleifung	30
2.2	Akustische Theorie der Spracherzeugung	32
2.3	Wahrnehmung	37
2.3.1	Gehörorgane und Schallwanderung	37
2.3.2	Lautheitswahrnehmung	40
2.3.3	Frequenzgruppen- und Tonhöhenwahrnehmung	41
2.4	Zusammenfassung	42
3	Merkmalgewinnung	45
3.1	Diskretisierung	46
3.1.1	Abtastung	46
3.1.2	Quantisierung	47
3.2	Kurzzeitanalyse	48
3.2.1	Zeitbereichsmerkmale	51
3.2.2	Spektrumanalyse	53
3.2.3	Cepstrumkoeffizienten	58
3.2.4	Lineare Vorhersage	61

3.2.5	Sonstige Ansätze	67
3.3	Zeitliche Veränderung des Sprachsignals	68
3.4	Zusammenfassung	73
4	Klassifikation	75
4.1	Numerische Klassifikatoren	76
4.1.1	Der optimale Klassifikator	76
4.1.2	Parametrische Verteilungsdichtefunktionen	79
4.1.3	Verteilungsfreie Klassifikatoren	81
4.1.4	Nichtparametrische Klassifikatoren	84
4.1.5	Die Schätzung der Fehlerrate	86
4.2	Überwachtes Lernen	87
4.2.1	Schätzung der Verteilungsdichteparameter	87
4.2.2	Informationstheoretische Optimierung	90
4.2.3	Verteilungsfreie Verfahren	93
4.3	Unüberwachtes Lernen	96
4.3.1	Vektorquantisierung	97
4.3.2	Identifikation von Mischverteilungen (Teil A)	100
4.3.3	Der EM-Algorithmus	102
4.3.4	Identifikation von Mischverteilungen (Teil B)	104
4.4	Suchverfahren	106
4.4.1	Metrische Nächster-Nachbar-Suche	108
4.4.2	Nichtmetrische Nächster-Nachbar-Suche	110
4.4.3	Sequentielle Klassifikation	112
4.5	Merkmaltransformationen	113
4.5.1	Karhunen-Loève-Transformation	114
4.5.2	Klassenbezogene Transformationen	116
4.6	Zusammenfassung	118
5	Markovmodelle	121
5.1	Das Einzelworterkennungproblem	122
5.2	Markovmodelle	125
5.2.1	Definitionen	127
5.2.2	Produktionswahrscheinlichkeiten	129
5.2.3	Die verborgene Zustandsfolge	131
5.2.4	Skalierung und Logarithmierung	134
5.3	ML-Schätzung der Modellparameter	135
5.3.1	Baum-Welch-Algorithmus	136
5.3.2	Viterbi-Training	139
5.4	Kontinuierliche Zustandsausgabefunktionen	140

5.4.1	Normalverteilungsdichten	141
5.4.2	Gaußsche Mischverteilungsdichten	142
5.4.3	Semikontinuierliche Markovmodelle	144
5.5	Lernstichprobe und Parameterraum	145
5.5.1	Mehrfache Modelle — mehrfache Trainingsbeispiele	146
5.5.2	Parameterverklebung	147
5.5.3	Interpolation	149
5.5.4	Glättung	151
5.6	Verallgemeinerte Modellkonzepte	153
5.6.1	Dauermodellierung	153
5.6.2	Korrelierte Ausgabedichten	156
5.6.3	Kartesische Produktmodelle	158
5.6.4	Weitere Lernverfahren	159
5.6.5	Integrierte Merkmaltransformation.	161
5.7	Zusammenfassung	163
6	Akustisch-phonetische Wortmodellierung	165
6.1	Wortmodelle	167
6.1.1	Die Schätzung von Wortmodellen	167
6.1.2	Ganzwortmodelle und Wortuntereinheiten	170
6.2	Kontextunabhängige Wortuntereinheiten	173
6.2.1	Phonologisch orientierte Wortuntereinheiten	174
6.2.2	Hierarchische Wortrepräsentationen	176
6.2.3	Akustisch orientierte Wortuntereinheiten	178
6.3	Kontextabhängige Wortuntereinheiten	179
6.3.1	Phoneme in Kontext	180
6.3.2	Modelle kontextabhängiger Phoneme	181
6.3.3	Verallgemeinerte Triphone	183
6.3.4	Polyphone	186
6.3.5	Subphonemische Modellierung	187
6.4	Wortgrenzenmodellierung	190
6.5	Nichtwörter, unbekannte Wörter und neue Wörter	192
6.6	Zusammenfassung	196
7	Grammatische Sprachmodelle	199
7.1	Linguistisch gesteuerte Spracherkennung	200
7.2	Stochastische Grammatiken	204
7.2.1	Diskrete stochastische Prozesse	204
7.2.2	Äquivalenzklassenbildung	207
7.2.3	Faktorisierung und Adaption des Sprachmodells	210

7.2.4	Konzeptuelle Grammatiken	214
7.3	Schätzung der Sprachmodellparameter	215
7.3.1	Glättung relativer Häufigkeiten	216
7.3.2	Rückfall auf vergrößerte Statistiken	219
7.3.3	Interpolation	220
7.3.4	Kategorien	225
7.4	Zusammenfassung	229
8	Dekodierung kontinuierlicher Sprache	231
8.1	Synchrone Suche	233
8.1.1	Netzwerke von Markovmodellen	233
8.1.2	Die beste Wortsegmentierung	236
8.1.3	Strahlsuche	240
8.1.4	Vorwärts-Rückwärts-Suche	243
8.2	Asynchrone Suche	244
8.2.1	Der A*-Algorithmus	245
8.2.2	Die Kellersuche	247
8.3	Wortschatzorganisation	253
8.4	Mehrphasendekodierung	259
8.4.1	Das Prinzip der schrittweisen Verfeinerung	259
8.4.2	Die n besten Wortketten	261
8.5	Zur Interaktion zwischen akustischem und linguistischem Modell	265
8.6	Zusammenfassung	268
9	Das ISADORA-System	271
9.1	Rekursive Markovmodelle	273
9.1.1	Definition	273
9.1.2	Die Vorwärts- und Rückwärtswahrscheinlichkeiten	275
9.2	Die Systemarchitektur	278
9.2.1	Strukturierte Markovmodelle	280
9.2.2	Netzwerkformalismus	281
9.2.3	Akustische Modelle	283
9.2.4	Reduzible und irreduzible A-Knoten	284
9.3	Das Netzwerk zur maschinellen Spracherkennung	287
9.3.1	Spracheinheiten unterhalb der Wortebene	288
9.3.2	Spracheinheiten innerhalb und oberhalb der Wortebene	291
9.3.3	Auswahl expliziter Modelle	294
9.3.4	Baumförmige Wortschatzorganisation	295
9.4	Lernen	297
9.4.1	Initialisierung	297

9.4.2	Standardlernverfahren	298
9.4.3	Das A.P.I.S.-Lernverfahren für hierarchische Wortmodelle	300
9.5	Erkennen	306
9.5.1	Rekursiver Viterbi-Algorithmus	306
9.5.2	Rückverfolgung	308
9.5.3	Geschachtelte symbolische Beschreibungen	310
9.5.4	Opake Knoten	312
9.6	Zusammenfassung	314
10	Experimentelle Untersuchungen	317
10.1	Basiskonfiguration zur Spracherkennung	318
10.1.1	Sprachdaten und Erkennungsaufgabe	318
10.1.2	Konfiguration des Erkennungssystems	319
10.2	Detaillierte Auswertungen zur Spracherkennung	321
10.2.1	Sprecherabhängige Worterkennung	321
10.2.2	Kodebuchentwurf	322
10.2.3	Phonetische Modellierung	326
10.2.4	Schätzung linguistischer Sprachmodelle	330
10.2.5	Mehrphasendekodierung	331
10.3	Erkennung spontan produzierter Äußerungen	335
10.3.1	Das Erlanger Bahnauskunftssystem	336
10.3.2	Evaluierung spontansprachlichen Datenmaterials	338
10.4	Zusammenfassung	341
11	Zusammenfassung	343
	Literaturverzeichnis	349
	Sachregister	388
A	Spracherkennung mit ISADORA	395
A.1	Phonetische Umschriftsysteme	396
A.2	Phonemische Basiseinheiten	397
A.3	Die Kardinalzahlwörter von 1 bis 999 999	398
A.4	Kategoriepaargrammatik für Uhrzeitangaben	399
A.5	Evaluierung verschiedener Wortuntereinheiten	400
B	Rekursive Markovmodelle	401
B.1	Schätzung der Modellparameter	401
B.2	Überführung in ein HMM	402