

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 27. September 2018

Module und Studiengänge

Zur Vorlesung

Zur Modulprüfung

Zum Inhalt

Literatur zur Lehrveranstaltung

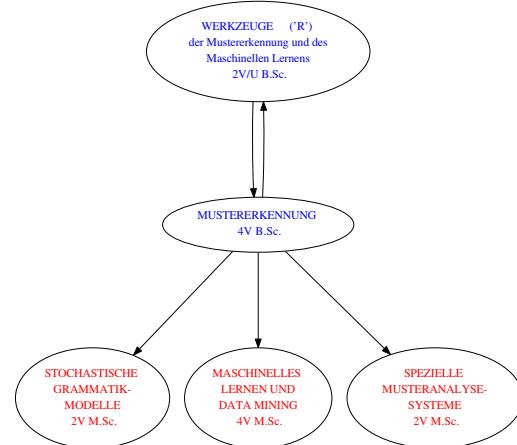
Professur für Musteranalyse

Lehrangebot Wintersemester & Sommersemester

Lehrbereich

Informatik
WP-Bereich Intelligente Systeme
Vertiefung Künstliche Intelligenz und Mustererkennung
Statistische Musteranalyse

Inhaltliche Abhängigkeiten



Meine Lehrveranstaltungen für ...

Informatiker & Bioinformatiker & Informatikerinnen & Bioinformatikerinnen

	SOMMERSEMESTER	WINTERSEMESTER
ASQ und EF Inf.	Intelligente Systeme 4V	Literaturarbeit und Präsentation ASQ 2S
Bachelor	Mustererkennung 4V Werkzeuge ME/ML 2Ü	Strukturiertes Programmieren 4V+2Ü
Master	Stochastische Grammatikmodelle 2V Biometriesysteme (Seminar) 2S	Maschinelles Lernen und Datamining 4V Spezielle Musteranalyse-systeme 2V

Wie studiere ich MUSTERANALYSE ?

Studiengänge: Informatik · Bioinformatik · Angewandte Informatik

Bachelor

Mustererkennung^{6LP}



WeRkzeuge ME/ML^{3LP}



Vertiefungsangebote

auf Antrag beim Prüfungsamt:

Maschinelles Lernen^{6LP}



Master

Maschinelles Lernen^{6LP}



Stochast. Grammatik^{3LP}



Musteranalysesysteme^{3LP}



Nivellierungsmodule

auf Antrag beim Prüfungsamt:

Mustererkennung^{6LP}



WeRkzeuge ME/ML^{3LP}



Mustererkennung

Vorlesung der Bachelor/Master-Studiengänge · 4V · 6 LP

Modultitel (deutsch)	Mustererkennung
Modultitel (englisch)	Pattern Recognition
Modulnummer	FMI-IN0036
Art des Moduls (Pflicht-, Wahlpflicht- oder Wahlmodul)	Wahlpflichtmodul (INT) für den B.Sc. Informatik Wahlpflichtmodul (INT) für den B.Sc. Angewandte Informatik Wahlpflichtmodul (Wahlpflichtbereich 2) für den B.Sc. Bioinformatik Wahlpflichtmodul (INT) für den M.Sc. Informatik (auf Antrag) Wahlpflichtmodul für den M.Sc. Bioinformatik (Bereich Informatik) Pflichtmodul für das Anwendungsfach Computational Neuroscience zum B.Sc. Angewandte Informatik Wahlpflichtmodul für das Lehramt Informatik
Modul-Verantwortlicher	Ernst Günter Schukat-Talamazzini
Leistungspunkte (ECTS credits)	6
Arbeitsaufwand (work load) in:	- Präsenzstunden: 180 Std. - Selbststudium (einschl. Prüfungsvorbereitung): 60 Std. - 120 Std.
Lehrform (SWS)	3 V + 1 Ü
Häufigkeit des Angebots (Modulturnus)	jährlich im Sommersemester
Dauer des Moduls	1 Semester
Voraussetzung für die Zulassung zum Modul	
Empfohlene Vorkenntnisse für das Modul	- FMI-IN0070 (Grundlagen der Modellierung und Programmierung) oder FMI-IN0040 (Grundlagen der Modellierung und Programmierung (Grundteile)) oder FMI-IN0025 (Strukturiertes Programmieren für Bioinformatiker) - FMI-IN0001 (Algorithmen und Datenstrukturen) - FMI-MA0007 (Einführung in die Wahrscheinlichkeitstheorie)
Voraussetzung für die Zulassung zur Modulprüfung	Bearbeitung der Übungsaufgaben Mindestens 50% der erzielbaren Punkte erreicht
Voraussetzung für die Vergabe von Leistungspunkten (Prüfungsform)	Klausur (120min) oder mündliche Prüfung (30min) zur Vorlesung Studienbegleitende Erfolgsmerkmale. Abgestufte (Prüfungs-)Anforderungen berücksichtigen das von Bachelor- und Masterstudierenden jeweils erreichbare Leistungsniveau.
Inhalte	Einführung in die Methoden der Mustererkennung zur maschinellen Modellierung und Simulation komplexer Informationsverarbeitungsprozesse, wobei sie insbesondere bei der Wahlerkennung und Auswertung visueller, akustischer oder taktiler Sinnesindrücke durch den Menschen auftreten. Diskretisierung/Filtrierung/Normierung; Merkmalauswahl und Merkmalstransformation; statistische, diskriminative und nichtparametrische Klassifikatoren; unüberwachtes Lernen; Zeitreihen

B.Sc. Informatik

WP-Bereich Int.Syst.

B.Sc. Bioinform.

WP-Bereich

B.Sc. Ang.Inform.

WP-Bereich Int.Syst.
Pflicht im Anw.fach CNS

M.Sc. Informatik

WP-Bereich Int.Syst.

M.Sc. Bioinform.

WP-Bereich Int.Syst.

LG Informatik

FS 6–9

Spezielle Musteranalysesysteme

Vorlesung der Master-Studiengänge · 2V · 3 LP

Modultitel (deutsch)	Spezielle Musteranalysesysteme
Modultitel (englisch)	Pattern Analysis Systems
Modulnummer	FMI-IN0054
Art des Moduls (Pflicht-, Wahlpflicht- oder Wahlmodul)	Wahlpflichtmodul (INT) für den M.Sc. Informatik Wahlpflichtmodul für den M.Sc. Bioinformatik (Bereich Informatik)
Modul-Verantwortlicher	Ernst Günter Schukat-Talamazzini
Leistungspunkte (ECTS credits)	3
Arbeitsaufwand (work load) in:	- Präsenzstunden: 90 Std. - Selbststudium (einschl. Prüfungsvorbereitung): 30 Std. - 60 Std.
Lehrform (SWS)	2V (mit Projektanteil)
Häufigkeit des Angebots (Modulturnus)	jährlich im Sommersemester
Dauer des Moduls	1 Semester
Voraussetzung für die Zulassung zum Modul	keine
Empfohlene Vorkenntnisse für das Modul	- FMI-IN0036 (Mustererkennung) - Vorkenntnisse aus den Bereichen Künstliche Intelligenz und Digitale Bildverarbeitung
Voraussetzung für die Vergabe von Leistungspunkten (Prüfungsform)	mündliche Prüfung (30min) zur Vorlesung oder Ausarbeitung/Präsentation zu einer Projektaufgabe
Inhalte	- Komplexe Musteranalyseaufgaben mit longitudinalen Daten (Sprach- und Sprecherkennung, (Hand)schrifterkennung, DNA-Motive, Musikretrieval) - Geeignete Lernverfahren (z.B. Hidden Markov Modelle; siehe Webseite zum Kurs für Detailinformationen), unterstützende Werkzeuge, Vorverarbeitung und Etikettierung der Lerndaten und syntaktische Modellierungsverfahren am Beispiel einer oder mehrerer ausgewählter Aufgabenstellungen
(Qualifikations-)Ziele	- Vertiefte Kenntnis der Methoden syntaktischer Musteranalyse - Kompetenzen der Analyse, des Designs und der Realisierung von Musteranalysesystemen realistischer Größenordnung - Fertigkeiten der Nutzung ausgewählter Softwarewerkzeuge der syntaktischen Musteranalyse

Werkzeuge Mustererkennung & Maschinelles Lernen

Vorlesung der Bachelor/Master-Studiengänge · 2V/P · 3 LP

Modultitel (deutsch)	Werkzeuge der Mustererkennung und des Maschinellen Lernens
Modultitel (englisch)	Tools for Pattern Recognition and Machine Learning
Modulnummer	FMI-IN0086
Art des Moduls (Pflicht-, Wahlpflicht- oder Wahlmodul)	Wahlpflichtmodul (INT) für B.Sc. Informatik Wahlpflichtmodul (INT) für B.Sc. Angewandte Informatik Wahlpflichtmodul (KIME.INT) für den M.Sc. Informatik (auf Antrag) Wahlpflichtmodul für den B.Sc. Bioinformatik (Bereich Informatik)
Modul-Verantwortlicher	Ernst Günter Schukat-Talamazzini
Leistungspunkte (ECTS credits)	3
Arbeitsaufwand (work load) in:	- Präsenzstunden: 90 Std. - Selbststudium (einschl. Prüfungsvorbereitung): 30 Std. - 60 Std.
Lehrform (SWS)	2V (mit Übung)
Häufigkeit des Angebots (Modulturnus)	jedes Sommersemester
Dauer des Moduls	1 Semester
Voraussetzung für die Zulassung zum Modul	Keine
Empfohlene Vorkenntnisse für das Modul	FMI-IN0036 (Mustererkennung) sollte gleichzeitig belegt werden
Voraussetzung für die Vergabe von Leistungspunkten (Prüfungsform)	50% der erreichbaren Punkte aus den Übungsaufgaben
Inhalte	Mündliche Prüfung oder Klausur Aufgabenstellungen aus den Bereichen Mustererkennung, Maschinelles Lernen, Datamining und ihre Bearbeitung mit geeigneten Softwarewerkzeugen: Klassifikation, Vorhersage, Clustering, Transformation, Visualisierung, Zeitreihen, Spektraldarstellung, Wahrscheinlichkeitsmodelle
(Qualifikations-)Ziele	- Fähigkeiten im praktischen Umgang mit Entwicklungswerkzeugen für maschinelles Lernen in Musteranalyse und Datamining - Grundlegende Kenntnisse über den Aufbau von Softwaresystemen und Programmierparadigmen für die maschinelle Datenanalyse - Kompetenzen in Datenanalyse, Versuchsplanung, Konfiguration von MU, Lösungen

B.Sc. Informatik

WP-Bereich Int.Syst.

B.Sc. Ang.Inform.

WP-Bereich Int.Syst.

B.Sc. Bioinform.

Zusatzmodul

M.Sc. Bioinform.

WP-Bereich Int.Syst.

M.Sc. Informatik

KI/ME & INT auf Antrag

Stochastische Grammatikmodelle

Vorlesung der Master-Studiengänge · 2V · 3 LP

Modul FMI-IN0146 Stochastische Grammatikmodelle	
Modulnummer/-code	FMI-IN0146
Modultitel (deutsch)	Stochastische Grammatikmodelle
Modultitel (englisch)	Stochastic Grammars
Modulverantwortlicher	Ernst Günter Schukat-Talamazzini
Voraussetzungen für Zulassung zum Modul	keine
Empfohlene bzw. erwartete Vorkenntnisse	keine
Art des Moduls (Pflicht-, Wahlpflicht- oder Wahlmodul)	Wahlpflichtmodul (KIME, INT) für den M.Sc. Informatik Wahlpflichtmodul für den M.Sc. Bioinformatik (Bereich bioinformatisch relevante Informatik) Wahlpflichtmodul für den M.Sc. Mathematik (Nebenfach Informatik) Wahlpflichtmodul (INF) für den M.Sc. Computational Science
Häufigkeit des Angebots (Zyklus)	jedes 2. Semester (ab Sommersemester)
Dauer des Moduls	1 Semester
Zusammensetzung des Moduls / Lehrformen (VL, Ü, S, Praktikum)	2V
Leistungspunkte (ECTS credits)	3LP
Arbeitsaufwand (work load)	90h
- Präsenzstunden	30h
- Selbststudium (einschl. Prüfungsvorbereitungen)	60h
Inhalte	Grammatische Modellierung von Zeichenfolgen natürlicher („Texte“) und künstlicher (z.B. Nukleotid- oder Aminosäure-Sequenzen) Sprachen. Vorlesungsthemen sind u.a.: <ul style="list-style-type: none">• Schwach kontextfreie Grammatiken (IG, TAG, HG, CG)• Information/Kompression• robuste Häufigkeitsschätzung (Bayes, Good-Turing, Zipf)• N-Gramme, Interpolation, Maximum-Entropiestochastische

Module und Studiengänge

Zur Vorlesung

Zur Modulprüfung

Zum Inhalt

Literatur zur Lehrveranstaltung

Maschinelles Lernen & Data Mining

Vorlesung der Master-Studiengänge · 4V · 6 LP

Modultitel (deutsch)	Maschinelles Lernen und Datamining
Modultitel (englisch)	Machine Learning and Datamining
Modulnummer	FMI-IN0034
Art des Moduls (Pflicht-, Wahlpflicht- oder Wahlmodul)	02.12.09 Wahlpflichtmodul (KIME, INT) für den M.Sc. Informatik Wahlpflichtmodul (INT) für den B.Sc. Informatik (zusätzliches Lehrangebot) Wahlpflichtmodul für den M.Sc. Bioinformatik (Bereich Informatik) Wahlpflichtmodul (INF) für den M.Sc. Computational Science Wahlpflichtmodul für das Lehramt Informatik
Modul-Verantwortlicher	Ernst Günter Schukat-Talamazzini
Leistungspunkte (ECTS credits)	6
Arbeitsaufwand (work load) in:	180 Std. - Präsenzstunden - Selbststudium (einschl. Prüfungsvorbereitung)
Lehrform (SWS)	4V (mit Projektanteil)
Häufigkeit des Angebots (Modulturnus)	jährlich im Wintersemester
Dauer des Moduls	1 Semester
Voraussetzung für die Zulassung zum Modul	Keine
Empfohlene Vorkenntnisse für das Modul	FMI-IN0036 (Mustererkennung)
Voraussetzung für die Zulassung zur Modulprüfung	Keine
Voraussetzung für die Vergabe von Leistungspunkten (Prüfungsform)	Klausur (120min) oder mündliche Prüfung (30min) zur Vorlesung
Inhalte	Strukturaufdeckung, Klassifizierung oder Entwicklungsvorhersage aus großen Datenflüssen (Finanzprozesse, Handel und Transport, med./biol. Datensätze, Klimamesswerte, elektronische Dokumente, Fertigungsautomatisierung) – Vorlesungsthemen sind u.a.: Skalentypen; Visualisierung hochdimensionaler Daten (PCA, MDS, ICA); überwachte Lernverfahren (Versionenraum, Entscheidungsbaum, lineare/logistische Modelle); unüberwachte Lernverfahren (hierarchisch, (fuzzy) K-means, spektral); Graphische Modelle (Bayesnetze, Markovnetze, Induktion und Inferenz)
(Qualifikations-)Ziele	Tiefgreifende Fachkenntnisse des Gebiets Maschinelles Lernen Fähigkeit zur Analyse, Design und Realisierung von ML-Systemen Flächendeckende Übersicht aktueller Techniken des Datamining

Vorlesung

Nutzung der Folienpräsentation

- Die Folien sollen vom Mitschreiben während der Vorlesung entlasten.
- Das Mitschreiben wird dadurch nicht überflüssig.
- Die Folien sind kein Lehrbuch.
- Die Folien sind daher im allgemeinen nur mit den Erläuterungen während der Vorlesung und entsprechenden eigenen Notizen verständlich.

Vorlesung

[Mathematische Sachverhalte](#)

- Wichtige mathematische Grundlagen werden in Steilkursen wiederholt.
- Die entsprechenden Fakten sind (oft) im letzten Abschnitt eines Vorlesungsteils dargestellt.
- Schwierige mathematische Zusammenhänge werden in der Anwendung verständlicher.
- Umfangreiche mathematische Formeln erscheinen viel harmloser, nachdem man/frau sie einmal programmtechnisch umgesetzt hat.

Vorlesung

[Elektronisches Folienskript](#)

Die PDF-Fassung des Folienskripts enthält einige Hyperlinks:

- **Verweise auf externe Webseiten**
Detaillierte Zusatzinformationen, Daten, Bilder
(funktioniert nicht während der Vorlesung ...)
- **Literaturangaben**
Verweis auf Quellenangaben am Ende des Dokuments
- **Programmcode**
'R'-Code zur Erstellung einer Grafik oder Tabelle
'dot'-Code zur Erzeugung eines (gerichteten) Graphen

[Module und Studiengänge](#)

[Zur Vorlesung](#)

[Zur Modulprüfung](#)

[Zum Inhalt](#)

[Literatur zur Lehrveranstaltung](#)

Prüfungsvorgang

Mündliches „Verhör“ · circa 30 Minuten

Prüfungstermine

[mehr Information](#)

Erstprüfung am Mi (13/20) 27 Februar 2019

Wiederholung am Fr 5 April 2019

Prüfungsstoff

Vorlesungsinhalte in Schrift und Wort

Anmeldung und Zulassung

zur Teilnahme **und** zur Prüfung

Deadline: **Montag, 24.12.2018**



Module und Studiengänge

Zur Vorlesung

Zur Modulprüfung

Zum Inhalt

Literatur zur Lehrveranstaltung

Form, Zweck & Ziel

Lehrveranstaltungsform

Vorlesung (2V)

Zulassungsvoraussetzungen

auf eigene Verantwortung: *keine*

Themengebiet

Syntaktische Analyse von 1D-Mustern mit HMMs
↳ 4V+2Ü Mustererkennung & Werkzeuge ME/ML

Zweck

Brücke zwischen mathematischer Theorie und F&E-Praxis

Lernziele

- | | |
|---|------------------------------|
| Vertiefte Theorie der HMMs | <i>RMM</i> |
| Lern- und Analysetechniken | <i>Vererbung/Dekodierung</i> |
| · umfangreiche Datenmengen | |
| · große Domänenmodelle | |
| · Suche in kombinatorischen Lösungsräumen | |
| Komplexe Anwendungsszenarien | <i>HSE/ASE</i> |
| Systemarchitektur von HMM-Werkzeugen | <i>Isadora</i> |

Module und Studiengänge

Zur Vorlesung

Zur Modulprüfung

Zum Inhalt

Literatur zur Lehrveranstaltung

Musteranalyse (allgemein)

Empfohlene Bücher zur Vorlesung

-  **Joachim Schenk and Gerhard Rigoll.**
Mensch-Maschine-Kommunikation.
Springer, 2010.
-  **King Sun Fu.**
Syntactic Pattern Recognition and Applications.
Advances in Computing Science and Technology. Prentice Hall, Englewood Cliffs, NJ, 1982.
-  **Keinosuke Fukunaga.**
Introduction to Statistical Pattern Recognition.
Academic Press, Boston, MA, 1994.
-  **H. Niemann.**
Pattern Analysis and Understanding, volume 4 of *Series in Information Sciences*.
Springer, Berlin Heidelberg, 1990.
-  **J.P. Marques de Sa.**
Pattern Recognition. Concepts, Methods and Applications.
Springer, 2001.

Hidden Markov Modelle

Empfohlene Bücher zur Vorlesung

 **Frederick Jelinek.**

Statistical Methods for Speech Recognition.

MIT Press, Cambridge, MA, 1997.

 **J.J. Nijtmans.**

Speech Recognition by Recursive Stochastic Modelling.

PhD thesis, Den Haag, 1992.

 **Gernot A. Fink.**

Mustererkennung mit Markov-Modellen.

Teubner, Wiesbaden, 2003.

 **Eugene Charniak.**

Statistical Language Learning.

MIT Press, Cambridge, Massachusetts, 1993.

Spracherkennung

Empfohlene Bücher zur Vorlesung

 **Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, and Raj Reddy, editors.**

Spoken Language Processing: A Guide to Theory, Algorithm and System Development.

Prentice Hall, 2001.

 **L.R. Rabiner.**

Fundamentals of Speech Recognition.

Signal Processing Series. Prentice Hall, Englewood Cliffs, NJ, 1993.

 **E.G. Schukat-Talamazzini.**

Automatische Spracherkennung.

Vieweg, Wiesbaden, 1995.

 **Stephan Euler.**

Grundkurs Spracherkennung.

Vieweg, Wiesbaden, 2006.

Sprachverarbeitung (allgemein)

Empfohlene Bücher zur Vorlesung

 **S. Furui.**

Digital Speech Processing, Synthesis, and Recognition.

Marcel Dekker, New York, 2001.

 **Ben Gold and Nelson Morgan.**

Speech and Audio Processing: Processing and Perception of Speech and Music.

John Wiley & Sons, 1999.

 **Daniel Jurafsky and James H. Martin.**

Speech and Language Processing.

Prentice Hall, 2008.

2nd edition.

 **Manfred R. Schroeder.**

Computer Speech: Recognition, Compression, Synthesis, volume 35 of Springer Series in Information Sciences.

Springer, 1999.

 **Erwin Paulus.**

Sprachsignalverarbeitung.

Spektrum Akademischer Verlag, 1998.

 **E. Zwicker and H. Fastl.**

Psychoacoustics, volume 22 of Information Sciences.

Springer, 2 edition, 1999.

Schrifterkennung

Empfohlene Bücher zur Vorlesung

 **Z.-Q. Liu, J.-H. Cau, and R. Buse.**

Handwriting Recognition, volume 133 of Studies in Fuzziness and Soft Computing.

Springer, 2003.

 **R. Plamondon and C.G. Leedham, editors.**

Computer Processing of Handwriting.

World Scientific, Singapore, 1990.

 **R. Plamondon, C.Y. Suen, and M.L. Simner, editors.**

Computer Recognition and Human Production of Handwriting.

World Scientific, Singapore, 1989.

 **Rolf-Dieter Bippus.**

Stochastische Modelle zur off-line Fließschrifterkennung.

Shaker, 2000.

 **Mario Pechwitz.**

Automatische Erkennung handgeschriebener arabischer Wörter.

Shaker, Aachen, 2005.

EUR 45.80.

 **S. Mori and H. Nishida.**

Optical Character Recognition.

John Wiley & Sons, 1999.

Anwendungen aus der Bioinformatik

Empfohlene Bücher zur Vorlesung

-  **Marina Axelson-Fisk.**
Comparative Gene Finding.
Computational Biology. Springer, 2010.
-  **Pierre Baldi and Søren Brunak.**
Bioinformatics. The Machine Learning Approach.
Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 1998.
-  **Warren J. Ewens and Gregory R. Grant.**
Statistical Methods in Bioinformatics: an Introduction.
Statistics for Biology and Health. Springer, 2001.
-  **J. Paetz.**
Soft Computing in der Bioinformatik.
Springer, 2006.
-  **Richard Durbin, Sean Eddy, Anders Krogh, and Graeme Mitchison.**
Biological Sequence Analysis.
Cambridge University Press, Cambridge, UK, 1998.

Information Retrieval in der Musik

Empfohlene Bücher zur Vorlesung

-  **Meinard Müller.**
Fundamentals of Music Processing.
Springer, 2015.
-  **Francesco Camastra and Alessandro Vinciarelli.**
Machine Learning for Audio, Image and Video Analysis.
Springer, 2010.
-  **Richard Kronland-Martinet, Solvi Ystad, and Kristoffer Jensen, editors.**
Computer Music Modeling and Retrieval. Sense of Sounds, volume 4969 of Lecture Notes in Computer Science.
Springer, 2008.
4th International Symposium, CMMR 2007, Copenhagen, Denmark, August 2007.
-  **Martin Neukom.**
Signale, Systeme und Klangsynthese.
Peter Lang Verlag ISBN 3-03910-125-0, 2003.

Mathematische Grundlagen

Empfohlene Bücher zur Vorlesung

-  **I.N. Bronstein and K.A. Semendjajew.**
Taschenbuch der Mathematik.
Verlag Harri Deutsch, Thun und Frankfurt/Main, 24 edition, 1989.
-  **Rüdiger Hoffmann and Matthias Wolff.**
Intelligente Signalverarbeitung, volume 1.
Springer, 2014.
-  **M. Drmota, B. Gittenberger, G. Karigl, and A. Panholzer.**
Mathematik für Informatik.
Heldermann Verlag, 2007.
438 Seiten fester Einband 36 EUR.
-  **L. Dümbgen.**
Stochastik für Informatiker.
Springer, 2003.
-  **Alfred Mertins.**
Signaltheorie.
Springer, 2013.
3. Auflage.
-  **Geoffrey Grimmett and David Stirzaker.**
Probability and Random Processes.
Oxford University Press, 2001.

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 19. Oktober 2018

Teil I

Syntaktische Musteranalyse

Numerische Klassifikation

Maschinelles Lernen

Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

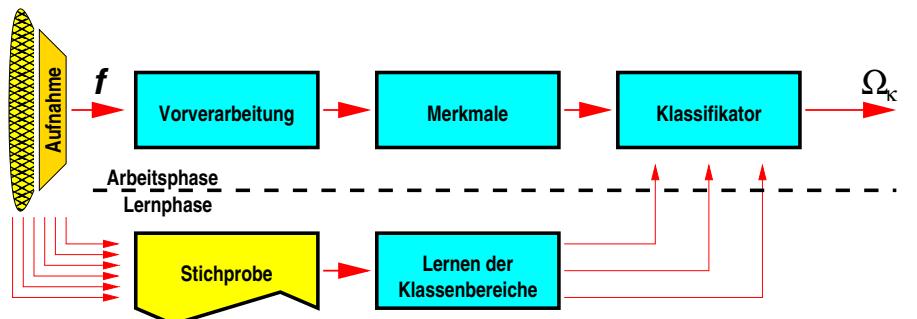
Einfaches Mustererkennungssystem

Diskretisierung Abtastung, Quantisierung Shannon-ATR, SNR $\approx 6B - 7.2$

Vorverarbeitung Filterung, Normierung TP/HP; Intensität, Größe

Merkmalextraktion Unvollst. Reihenentwicklung DFT, WHT, PCA, LDA

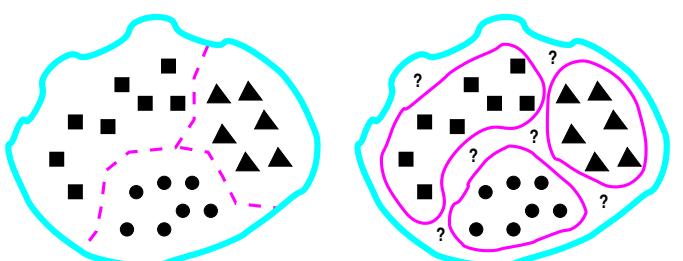
Klassifikation Raten der Musterklasse $\delta : \mathbb{R}^D \rightarrow \{1, \dots, K\}$



Klassifikationsaufgabe

Jedes **Muster** ist repräsentiert durch:

- einen Vektor numerischer **Merkmale** $x_1, \dots, x_D \in \mathbb{R}$
- seine wahre **Klassenzugehörigkeit** $\kappa \in \{1, \dots, K\}$



Geometrische Aufgabenstellung

- Muster $\hat{=}$ Punkt im euklidischen Raum \mathbb{R}^D
- Objektklasse $\hat{=}$ Punktwolke
- Klassifikator $\hat{=}$ K -Partition des Raumes

Die Bayesregel

Satz

Die Bayes-Entscheidungsregel (Maximum a posteriori Regel)

$$\delta^{MAP}(x) \stackrel{\text{def}}{=} \underset{\lambda}{\operatorname{argmax}} P(\lambda|x) = \underset{\lambda}{\operatorname{argmax}} \frac{P(\lambda) \cdot P(x|\lambda)}{P(x)}$$

ist optimal, d.h., sie garantiert den kleinstmöglichen erwarteten Klassifikationsfehler.

Definition

Die Fehlerrate

$$\varepsilon(\delta^{MAP}) \stackrel{\text{def}}{=} \underset{\delta}{\operatorname{argmin}} \mathcal{E}_{P_{K,X}}[\delta_K(X)]$$

der MAP-Regel heißt **Bayesfehlerrate** der Aufgabenstellung $P_{K,X}(\cdot, \cdot)$; sie kann von keinem Klassifikator unterboten werden.

Die optimale Entscheidungsregel

Definition

Die (randomisierte) **Entscheidungsregel**

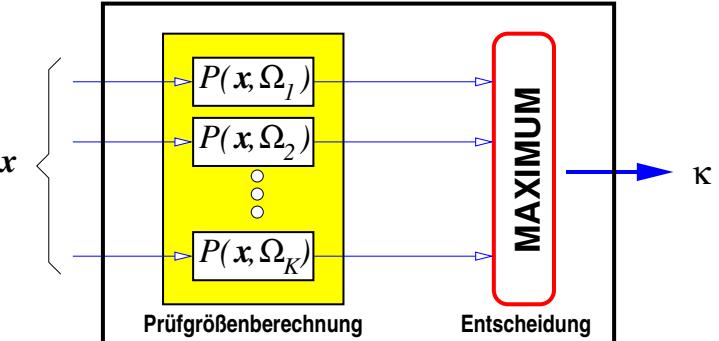
$$\delta : \mathbb{R}^D \rightarrow [0, 1]^K$$

heißt **optimal**, wenn sie den erwarteten Klassifikationsfehler

$$\begin{aligned} \varepsilon(\delta) &\stackrel{\text{def}}{=} 1 - \mathcal{E}_{P_{K,X}}[\delta_K(X)] \\ &= \mathcal{E}_{P_{K,X}} \left[\sum_{\lambda \neq K} \delta_\lambda(X) \right] \\ &= \sum_{\kappa=1}^K \int \left(P(\kappa, x) \cdot \sum_{\lambda \neq \kappa} \delta_\lambda(x) \right) dx \end{aligned}$$

minimiert.

Numerische Klassifikation — der Idealfall



Klassifikation $\hat{=}$ Maximierung der (idealen) Prüfgröße

$$u_\kappa(x) = P(x, \kappa) = P(\kappa) \cdot P(x|\kappa) \propto P(\kappa|x)$$

Numerische Klassifikation

Maschinelles Lernen

Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

Maschinelles Lernen

Problem

Die gemeinsame Verteilung $P_{\mathbb{K}, \mathbf{x}}$ und damit auch die a posteriori Klassenwahrscheinlichkeiten $P(\kappa|\mathbf{x})$ sind in realen Anwendungen nicht bekannt.

Lösung

Automatisches Lernen einer Entscheidungsregel (Parameter; Struktur) aus klassifizierten Beispieldaten

• **Statistisch.**

parametrische Verteilungsannahme $P(\mathbf{x}|\kappa) \hat{=} f(\mathbf{x}|\boldsymbol{\theta}_\kappa)$

NVK, BN, MRF

• **Diskriminativ.**

parametrische Trennfunktionen $P(\kappa|\mathbf{x}) \hat{=} h(\mathbf{x}|\boldsymbol{\theta}_\kappa)$

Poly, MLP, CME

• **Nichtparametrisch.**

Glättung der empirischen Datenverteilungsdichten (Potentialfunktion; Kernel)

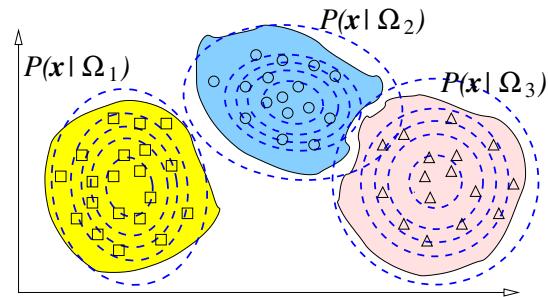
Parzen, UPS, kNN

• **Semiparametrisch.**

Struktur/Freiheitsgrade/Modellkapazität werden mitgelernt

SCT, SVM

Statistische Klassifikatoren



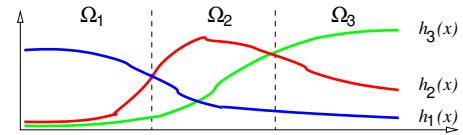
NVK

Normalverteilungsannahme:

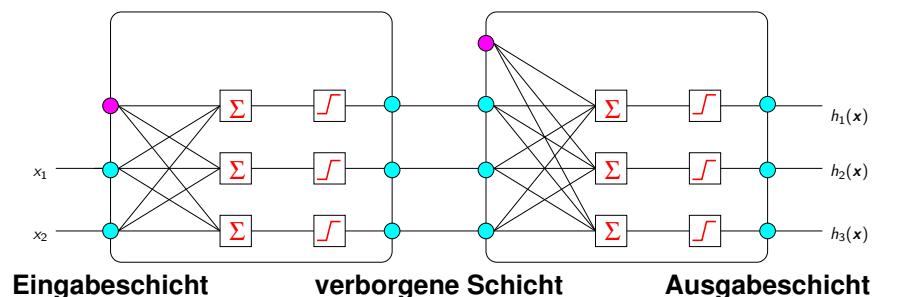
$$P(\mathbf{x}|\Omega_\kappa) \sim \mathcal{N}(\mathbf{x} | \boldsymbol{\mu}_\kappa, \mathbf{S}_\kappa)$$

$$\stackrel{\text{def}}{=} \frac{1}{\sqrt{\det(2\pi\mathbf{S}_\kappa)}} \cdot \exp \left\{ -\frac{1}{2} \cdot (\mathbf{x} - \boldsymbol{\mu}_\kappa)^\top \mathbf{S}_\kappa^{-1} (\mathbf{x} - \boldsymbol{\mu}_\kappa) \right\}$$

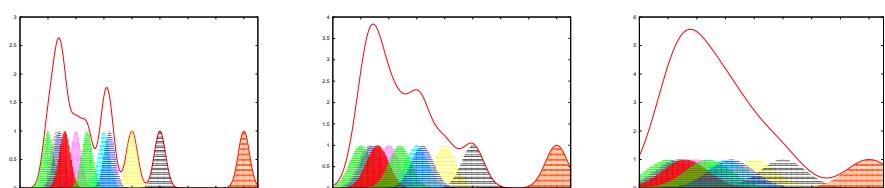
Diskriminative Klassifikatoren



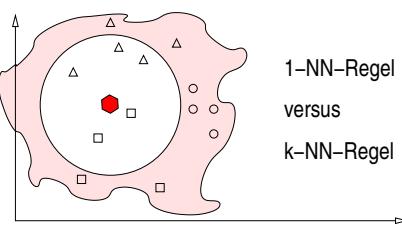
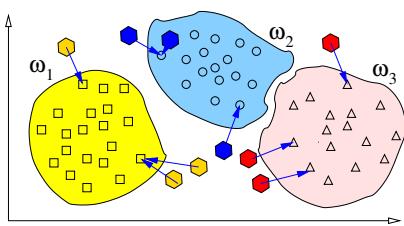
MLP



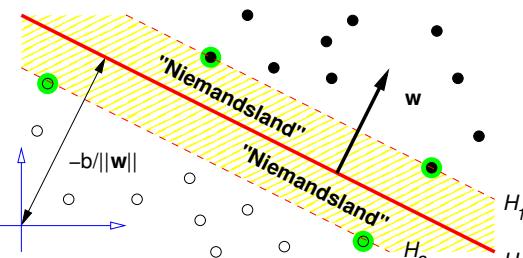
Nichtparametrische Klassifikatoren



Parzen-Dichteschätzung: mit Gaußkernen (↑) oder uniform (↓)

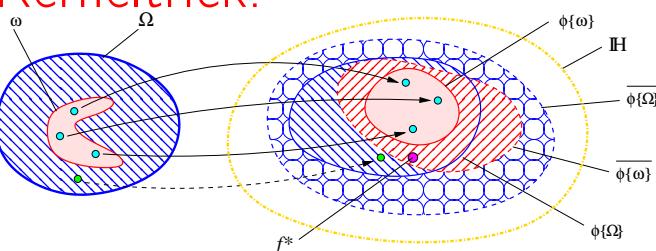


Semiparametrische Klassifikatoren 1

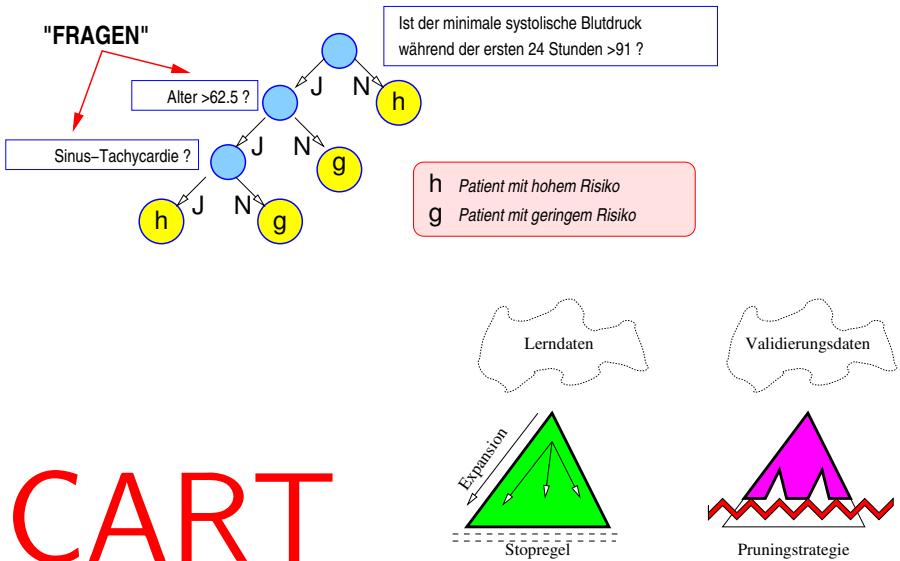


SVM

Kerneltrick:



Semiparametrische Klassifikatoren 2



Kapazität und induktiver Bias

Maschinelles Lernen $\hat{=}$ Optimierende Suche in einem Hypothesenraum

Definition

Unter der **Kapazität** eines Klassifikationsverfahrens verstehen wir die (Mächtigkeit der) Menge

$$\mathcal{H} \subseteq \{\delta | \delta : \mathbb{R}^D \rightarrow \{1, 2, \dots, K\}\}$$

aller damit prinzipiell lernbaren Entscheidungsfunktionen. Die kapazitätsbedingte Einschränkung des Lösungsraums heißt **induktiver Bias**.

Bemerkungen

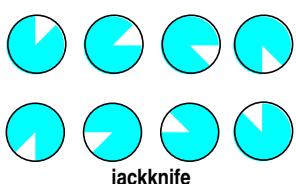
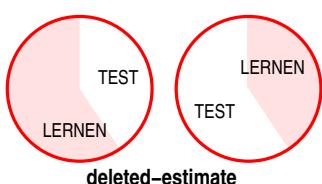
1. Der MAK (Minimumabstand-Klassifikator) generiert Klassengebiete in Form einer *Voronoi-Partition*
2. Ein NVK lässt ausschließlich Entscheidungsfunktionen δ mit *quadratischen Klassengrenzen* zu.
3. Ein (monothetischer) CART partitioniert den Merkmalraum in Parallelepipede
4. Hohe Kapazität $\hat{=}$ komplexes Modell, viele Parameter bzw. *Freiheitsgrade* („Gedächtnis“)
5. Parametrische Klassifikatoren besitzen Gedächtnis konstanter Größe
6. Nichtparametrische Klassifikatoren besitzen Gedächtnis vom Umfang der Lernstichprobe

Wahre und geschätzte Fehlerraten

Definition

Gegeben sei ein Klassifikationsverfahren, eine etikettierte Lerndatensammlung ω^ℓ und die gelernte Regel $\delta = \delta(\omega^\ell)$.

1. Der Erwartungswert $\varepsilon_\delta = \mathbb{E}[1 - \delta_{\mathbb{K}}(\mathbb{X})]$ heißt **wahre Klassifikationsfehlerrate**.
2. Für eine etikettierte Testprobe ω^t bezeichne $\varepsilon_\delta(\omega^t)$ die **empirische Fehlerrate** (bzgl. ω^t).
3. Die spezielle Fehlerrate $\varepsilon_\delta(\omega^\ell)$ heißt **Resubstitutionsfehlerrate**.

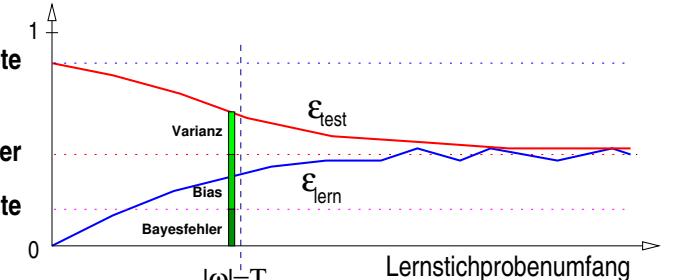


Überanpassung und Verallgemeinerung

Zufallsfehlerquote

Asymptotischer Fehler

Bayesfehlerrate



Bemerkungen

1. Für alle Entscheidungsregeln δ gilt $1 - 1/K \geq \varepsilon_\delta \geq \varepsilon_{\delta \text{MAP}}$.
2. $\varepsilon_\delta(\omega^t)$ ist ein **unverzerrter Schätzer** für ε_δ , falls nur ω^t unabhängig von ω^ℓ „gezogen“ wurde.
3. Für ω^ℓ selbst ist das nicht der Fall; $\varepsilon_\delta(\omega^\ell)$ wird den wahren Wert ε_δ deutlich überschätzen
4. **ÜBERANPASSUNG** \diamond kleine Lerndatensample \cdot geringer Bias
5. **VERALLGEMEINERUNG** \diamond große Lerndatensample \cdot paßförmiger Bias

$$\left\{ \begin{array}{c} \text{Fehlerrate} \\ \varepsilon_{\delta(\omega^\ell)} \end{array} \right\} = \left\{ \begin{array}{c} \text{Bayesfehler} \\ \varepsilon_{\text{MAP}} \end{array} \right\} + \left\{ \begin{array}{c} \text{Bias} \\ \varepsilon_{\delta_\infty} - \varepsilon_{\text{MAP}} \end{array} \right\} + \left\{ \begin{array}{c} \text{Varianz} \\ \varepsilon_{\delta_T} - \varepsilon_{\delta_\infty} \end{array} \right\}$$

Sollte das etwa schon alles gewesen sein ?!

- Nominal oder ordinal skalierte Attribute
- Objekte mit Attributen unterschiedlicher Skalen
- Klassifikation von Zeichenketten
- Analyse von Zeichenketten
- Klassifikation von 1D-Mustern
- **Analyse von 1D-Mustern**
- Klassifikation von 2D/3D-Objekten
- Analyse von 2D/3D-Objekten
- Vorhersage reeller statt nominaler Zielgrößen
- Nicht/kausale Strukturierung der Objektvariablen
- Hierarchische Gruppierung von Objekten

ML

(ML)

SGM

SGM

ME

SMAS

DBV

CV/RS

ML

ML

ML

Numerische Klassifikation

Maschinelles Lernen

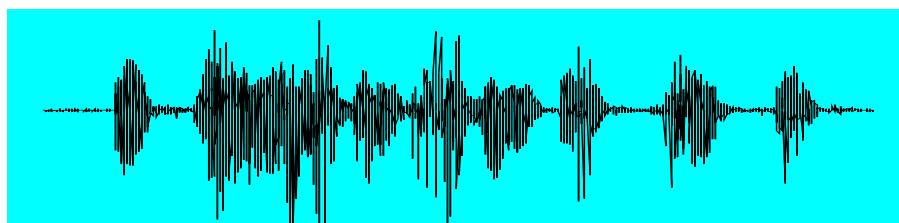
Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

Biometrische Sprecheridentifikation



AUSGABE:

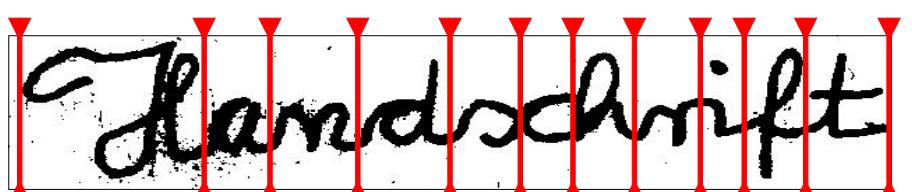
die Identität **eines** von $K \in \mathbb{N}$
möglichen Sprechern

$$\Omega_{\kappa} = \text{'Eva Hermann'}$$

EINGABE:

eine lineare Folge / Menge von
Spektralvektoren $x_1, \dots, x_T \in \mathbb{R}^D$

Segmentierung von Schriftzeichen — 1D



EINGABE:

eine lineare Folge von Bildspalten
 $x_1, \dots, x_T \in \mathbb{R}^D$

AUSGABE:

eine Folge von Objektgrenzen
 $(t_0, t_1, t_2, \dots, t_m)^\top \in \mathbb{N}^*$

Segmentierung von Schriftzeichen — 2D



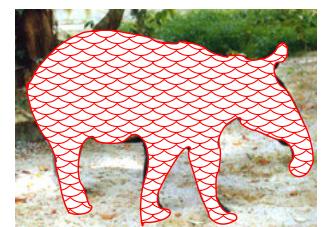
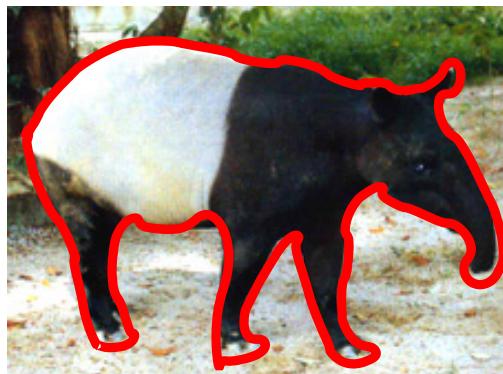
AUSGABE:

eine Menge von
Objektumschreibungen
 $(n_i, m_i, \nu_i, \mu_i)^\top \in \mathbb{N}^4, i = 1, 2, \dots$

EINGABE:

ein Grauwertbild $X \in [0, 255]^{N \times M}$

Segmentierung komplex geformter Objekte — 2D



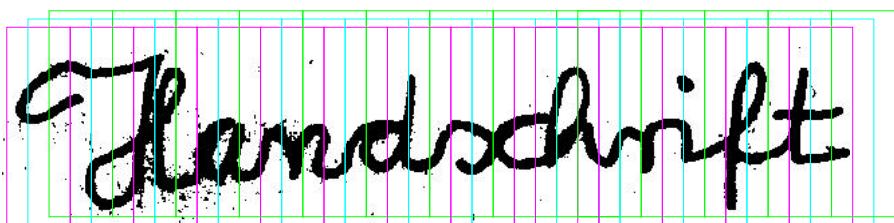
EINGABE:

ein Farbbild $X \in ([0, 255]^3)^{N \times M}$

AUSGABE:

eine geschlossene Konturlinie bzw.
eine zusammenhängende
Punktmenge
 $\mathcal{P}_{\text{tapir}} \subseteq [1 : N] \times [1 : M]$

Maschinelle Handschrifterkennung



AUSGABE:

eine Zeichenfolge $z \in A^*$ über dem

Alphabet

$\{A, B, \dots, Z, a, b, \dots, z, 0, 1, \dots, 9\}$

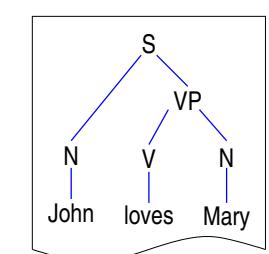
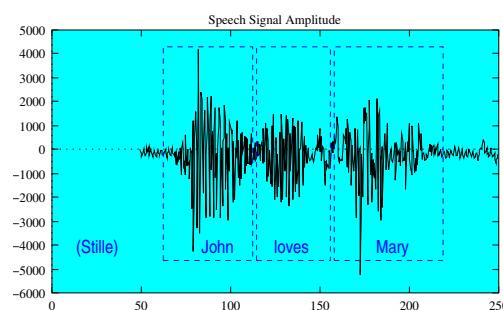
$$(H, a, n, d, s, c, h, r, i, f, t)$$

FINGARF

eine lineare Folge überlappender Bildfenster $\mathbf{X}_t \in \mathbb{R}^{N \times \mu}$, $t = 1, 2, \dots$

(, , , , , , , , ,)

Automatisches Verstehen gesprochener Sprache

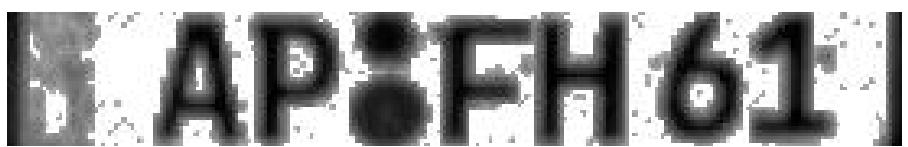


AUSGABE-

AUSGABE:
ein Ableitungsbaum aus Wortformen
und Syntaxkategorien:

S(N(John), VP(V(loves), N(Mary)))

Syntaxgesteuerte KFZ-Kennzeichenerkennung



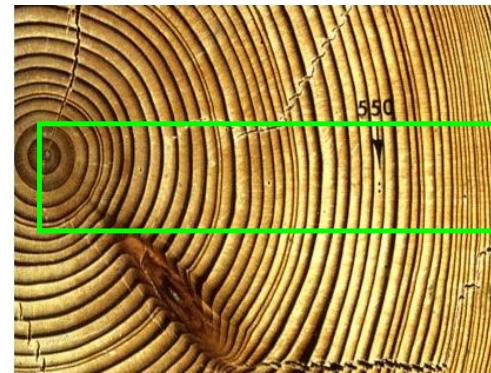
AUSGABE:

eine strukturierte Kennzeichenidentifikation aus dem aktuell gültigen Fahndungsblatt

EINGABE:
eine lineare Folge überlappender
Bildfenster $\mathbf{X}_t \in \mathbb{R}^{N \times \mu}$, $t = 1, 2, \dots$

$$(AP, FH, 61) \in \mathcal{C} \times \mathcal{L} \times [1 : 9999]$$

Altersbestimmung bei Nutzhölzern



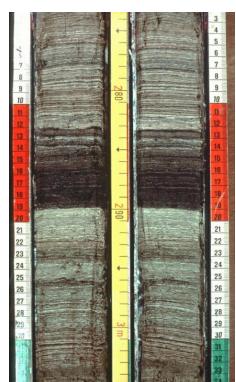
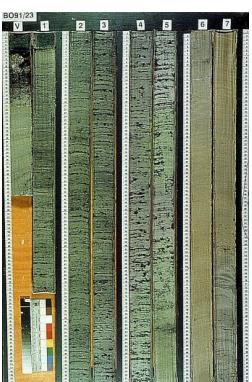
EINGABE:

eine lineare Folge überlappender
Bildfenster $\mathbf{X}_t \in \mathbb{R}^{N \times \mu}$, $t = 1, 2, \dots$
auf einem radialen
Querschnittstreifen

AUSGABE:

die Anzahl der Jahresringe, ermittelt aus der symbolischen Beschreibung $\mathfrak{B} \in \{R, Z\}^*$

Analyse von Bohrkernen — Klassifikation, Annotation, Lokalisierung



AUSGABE:

EINGABE:
eine lineare Folge überlappender
Bildfenster $X_t \in \mathbb{R}^{N \times \mu}$, $t = 1, 2, \dots$
in Bohrkernrichtung

1. der vorliegende Formationstyp
2. Sedimentsequenz (m/o Schichtdicke)
3. Siedlungsspuren (ja/nein & Position)

Numerische Klassifikation

Maschinelles Lernen

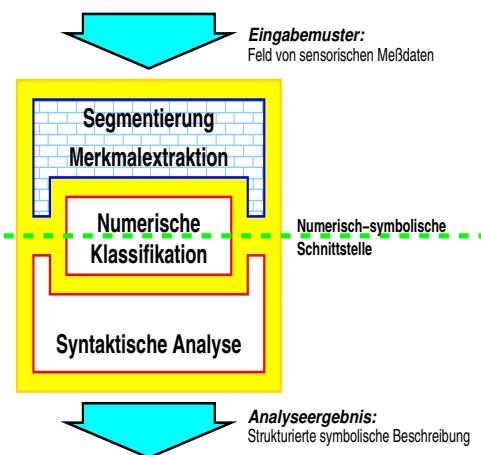
Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

Architektur von Musteranalysesystemen



- **Eingabe**
 - einfache Muster
 - komplexe Muster (1D, 2D, ...)
- **Ausgabe**
 - Klassenname
 - symbolische Beschreibung
- **Segmentierung**
Zerlegung in Intervalle oder Regionen
- **Klassifikation**
Kategorisierung atomarer Segmente
- **Aggregierung**
geometrieverträgliche Verschachtelung
Attributierung grammikgesteuerte Filterung/Korrektur

Sequentielles Verarbeitungsschema der Musteranalyse

-
- (Algorithmus)
- 1 VORVERARBEITUNG
Diskretisierung, Filterung
 - 2 SEGMENTIERUNG
Das komplexe Muster wird (hierarchisch) in Teilmuster zerlegt
 - 3 IDENTIFIZIERUNG
Die elementaren Teilmuster werden maschinell klassifiziert
~~ Filterung, Merkmale, Klassifikator
 - 4 RESTRUKTURIERUNG
Die Klassennamen werden vermöge (2) serialisiert/verschachtelt
 - 5 NACHBEARBEITUNG
Korrekturregeln, Vereinfachungsregeln, Inferenzprozesse
 - 6 MUSTERVERGLEICH
optional: Klassifikation durch syntaktische Vergleichsoperationen
-
- (zumdingIA)

Segmentierung komplexer Muster

Was sind eigentlich Segmente?

Intervalle (1D) · Kreise, Rechtecke, Polygone, einfach zusammenhängende Gebiete (2D)

- **Grenzen finden**

Dynamische Programmierung (1D)

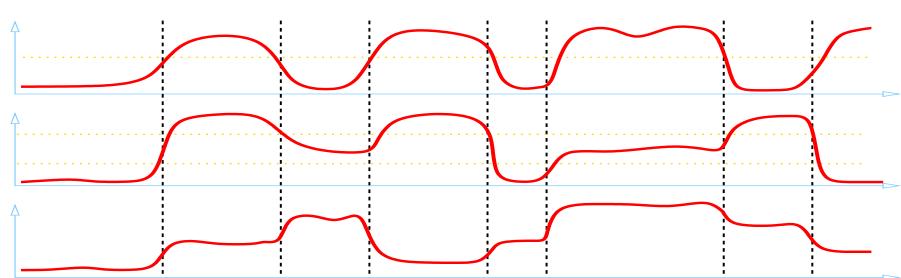
Hough-Methoden für Linien- und Kurvensegmente (2D)

Iterative Energieminimierer (2D ezg)

- **Gebiete finden**

Schwellwerte · Klassifizierer · Level-Sets

Relaxationsverfahren ('temporal/spatial K-means')



Schwachstellen der sequentiellen Systemarchitektur

- **Ungehinderte Fehlerfortpflanzung**

Klassifikationsfehler \rightsquigarrow weiche Entscheidung

Segmentierungsfehler \rightsquigarrow Zerlegungsalternativen ?!

- **Ad hoc Techniken bei Segmentierung/Nachverarbeitung**

systematische Methodenentwicklung ??

- **1:1-Kopplung zwischen Zerlegung und Beschreibung**

unrevidierbare Frühentscheidungen ...

- **Dekodierziel nicht präzise definiert**

Metriken zur Beurteilung des Analyseerfolgs ?

- **Lernen aus Beispieldaten nur rudimentär**

überwachtes Lernen nur auf der Klassifikationsebene

Nachverarbeitungsverfahren

- **Vereinfachung**

Kompression sukzessiv wiederholter Klassennamen

$$(m,m,m,m,u,u,s,s,t,a,a,a) \Rightarrow (m,u,s,t,a)$$

- **Glättung**

Tilgung/Umsetzung sporadischer Ausreißersymbole

$$(0,1,0,0,0,1,1,1,1,0,0,1,0,0) \Rightarrow (0,0,0,0,0,1,1,1,1,0,0,0,0,0)$$

- **Korrektur I**

Kontextuelle Filterung mit N-Grammen

$$(S,q,h,u,k,d,t,-,T,a,1,a,m,a,z,z,i,n,i)$$

$$(S,c,h,u,k,a,t,-,T,a,l,a,m,a,t,z,i,n,i)$$

- **Korrektur II**

Kontextuelle Filterung mit Wörterbuch („T9“)

$$(9,e,b,O,r,c,n, ,j,n, ,6,e,n,f) \Rightarrow (g,e,b,o,r,e,n, ,i,n, ,S,e,n,f)$$

- **Inferenz**

„Rechnen“ mit partiellen symbolischen Beschreibungen

Numerische Klassifikation

Maschinelles Lernen

Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

Symbolische Beschreibung und Musteranalyse

Definition

Es sei Ω eine Menge, \mathcal{A} ein endlicher Symbolvorrat (Alphabet) und G eine formale Grammatik über \mathcal{A} . Eine Abbildung

$$\delta : \begin{cases} \Omega & \rightarrow \mathcal{L}(G) \\ \mathbf{X} & \mapsto \mathbf{b} = \delta(\mathbf{X}) \end{cases}$$

heißt **syntaktische Entscheidungsregel**. Ein Element

$\mathbf{b} = \delta(\mathbf{X}) \in \mathcal{L}(G)$ heißt **symbolische Beschreibung des komplexen Musters** $\mathbf{X} \in \Omega$.

Bemerkung

Damit liegen bis auf die Tatsache $|\mathcal{L}(G)| = \infty$ die bekannten Voraussetzungen für die Herleitung der **optimalen** (syntaktischen) Entscheidungsregel vor.

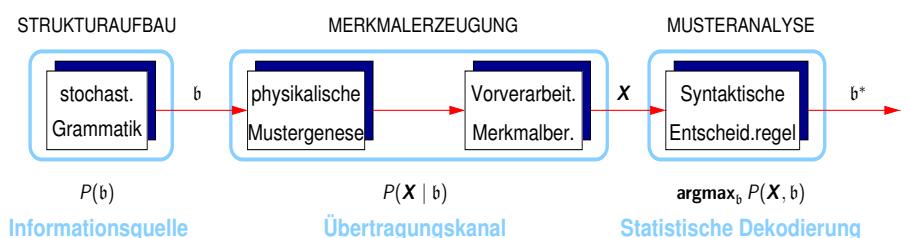
Musteranalyse mit der Bayesregel

Satz

Die **syntaktische Bayes-Entscheidungsregel**

$$\delta^{MAP}(\mathbf{X}) \stackrel{\text{def}}{=} \underset{\mathbf{b} \in \mathcal{L}(G)}{\operatorname{argmax}} P(\mathbf{b} | \mathbf{X}) = \underset{\mathbf{b} \in \mathcal{L}(G)}{\operatorname{argmax}} \frac{P(\mathbf{b}) \cdot P(\mathbf{X} | \mathbf{b})}{P(\mathbf{X})}$$

ist **optimal**, d.h., sie garantiert den kleinstmöglichen erwarteten Analysefehler.



Kontextuelle Klassifikation — \mathbf{b} sequentiell & synchron

$$\mathbf{X} = (x_1, \dots, x_T) \xrightarrow{\text{MA}} \ell = (\ell_1, \dots, \ell_T)$$

• Quellmodell

Unabhängige vs. markovsche Komponentenklassen:

$$P(\ell) = \prod_{t=1}^T \pi_{\ell_t} \quad \text{versus} \quad P(\ell) = \prod_{t=1}^T a_{\ell_{t-1}, \ell_t}$$

• Übertragungsmodell

Klassenbedingt unabhängige Musterkomponentenerzeugung

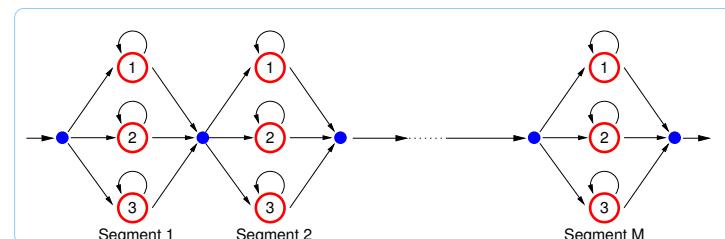
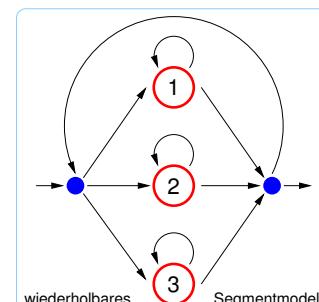
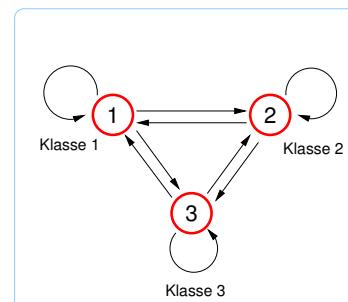
$$P(\mathbf{X} | \ell) = \prod_{t=1}^T P(x_t | \ell_t) =: \prod_{t=1}^T b_{\ell_t}(x_t)$$

• Dekodierungsregel

Wiederholte Bayesregel vs. HMM Viterbi-Decoder

$$\hat{\ell}_t = \underset{k}{\operatorname{argmax}} P(k | x_t) \quad \text{versus} \quad \hat{\ell} = \underset{\ell}{\operatorname{argmax}} P^*(\mathbf{X}, \ell | \pi, \mathbf{A}, \mathbf{B})$$

Hidden Markov Modelle zur Klassifikation und Segmentierung



Segmentelle Klassifikation — \mathfrak{b} sequentiell

$$\mathbf{X} = (x_1, \dots, x_T) \xrightarrow{\text{MA}} \ell = (\ell_1, \dots, \ell_M)$$

- **Quellmodell**

Markovsche Komponentensegmente:

$$P(M, \ell, t) = P(\text{Länge} = M) \cdot P(t_0, \dots, t_M) \cdot \prod_{m=1}^M a_{\ell_{m-1}, \ell_m}$$

- **Übertragungsmodell**

Kbu. Mustersegmenterzeugung durch wiederholbare Zustände s_ℓ

$$P(\mathbf{X} | M, \ell, t) = \prod_{m=1}^M \prod_{t=t_{m-1}+1}^{t_m} b_{\ell_m}(x_t)$$

- **Potentielle Dekodierungsziele**

$$\underset{\ell \in \mathcal{M}^M}{\operatorname{argmax}} P(\mathbf{X}, \ell) \quad \text{versus} \quad \underset{\ell \in \mathcal{M}^*}{\operatorname{argmax}} P(\mathbf{X}, \ell) \quad \text{versus} \quad \underset{\ell \in \mathcal{M}^*, t \in \mathbb{N}^*}{\operatorname{argmax}} P(\mathbf{X}, \ell, t)$$

Hierarchische Zerlegung — \mathfrak{b} Ableitungsbaum

$$\mathbf{X} = (x_1, \dots, x_T) \xrightarrow{\text{MA}} \mathfrak{b} \text{ Ableitung in } G = (\mathcal{A}, \mathcal{N}, S, \mathfrak{R})$$

- **Quellmodell**

Reguläre stochastische Phrasenstrukturgrammatik mit Regeln

$$R_i \in \mathcal{N} \times (\mathcal{A} \cup \mathcal{N})^* \times [0, 1]$$

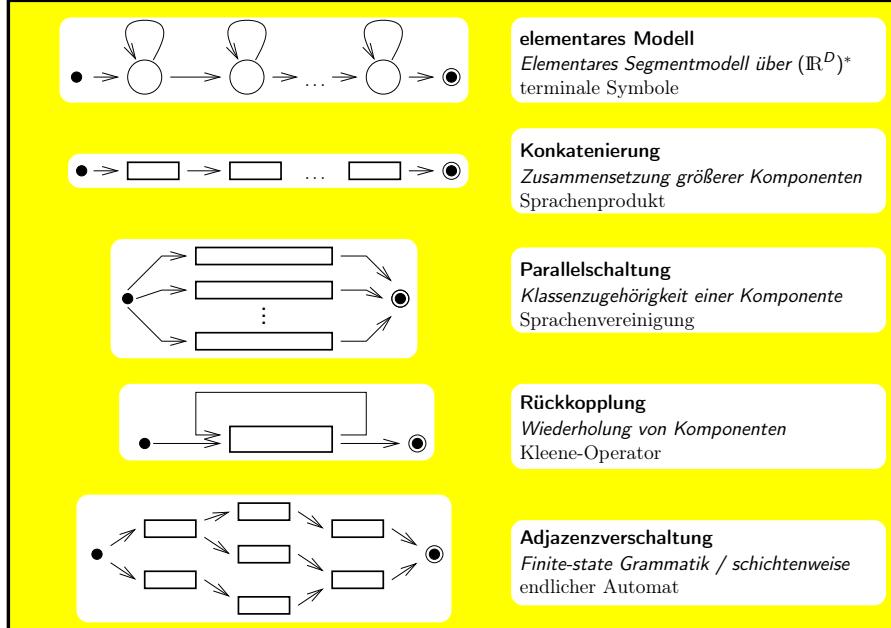
- **Wahrscheinlichkeit** einer **Ableitung** / einer **Oberfläche**:

$$P(\mathfrak{b}) = \prod_{i=1}^{|\mathfrak{b}|} P(R_i^\mathfrak{b}), \quad \mathfrak{b} = (R_1^\mathfrak{b}, R_2^\mathfrak{b}, R_3^\mathfrak{b}, \dots)$$

$$P(\ell) = \sum_{\sigma(\mathfrak{b})=\ell} \prod_{i=1}^{|\mathfrak{b}|} P(R_i^\mathfrak{b}), \quad \ell \in \mathcal{A}^*$$

Es bezeichne $\sigma(\mathfrak{b})$ die „Oberfläche“ (Blattzeichenfolge) von \mathfrak{b} .

Hierarchische Verschaltung von HMMs



Numerische Klassifikation

Maschinelles Lernen

Komplexe Musteranalyseaufgaben

Sequentielle Musteranalyseverfahren

Statistische Musteranalyseverfahren

Vorlesungsinhalte

Übersicht über den Vorlesungsinhalt

1. Klassifikation und Musteranalyse

- Resümee und Motivation

2. Automatische Erkennung gesprochener Sprache

- Aufgabenstellung · Lautsprache · Merkmalgewinnung
- Wortstruktur/modelle · Satzgrammatik

3. Vertiefte Theorie der Hidden Markov Modelle

- Hidden Markov Modelle (DD/CD, F/B, MAP/VA, BW/VT)
- Dekodierverfahren mit HMMs
- RMM — das hierarchisch verschachtelte HMM
- Diskriminative HMM-Derivate: *conditional maximum entropy*

4. Das ISADORA-Paket — RMM-Implementierung

- Spezialisierte RMM-Zustände
- Ausgabeverteilungen, Modellinventar & Vererbung
- Dekodierung, Gedächtnis & Opazität
- Einfache Analysetechniken:
Klassifikation · Clustering · Segmentierung · Spotting · Parsing · Motiventdeckung
- Konfiguration: Kraftfahrzeug-Kennzeichen
- Konfiguration: Spracherkennung

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit
Hidden-Markov-Modellen

Teil II

Automatische Spracherkennung

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Motivation

Sprachverstehen

Taxonomie

Schwierigkeiten

Systemaufbau

Motivation

Sprachverstehen

Taxonomie

Schwierigkeiten

Systemaufbau

Durchschnittliche Übertragungsraten

Wozu automatische Spracherkennung?

Was ist maschinelles Sprachverstehen?

Taxonomie sprachverstehender Systeme

Warum ist Spracherkennung schwierig ?

Architektur eines Spracherkennungssystems

Tastenfeld trainiert 100–150 W/min
untrainiert 10–25 W/min
Tastenzahl & Tastenbelegung
100% Erkennung (Tippfehler!)

Handschrift 25 W/min m/o Übung
· automat. Erkennung von Blockschrift gelöst
· automat. Erkennung von Kursivschrift ungelöst

Lautsprache 120–250 W/min m/o Übung
Diktiermaschine 40 W/min
automatische Erkennung ??

Vorteile gesprochener MM-Kommunikation

- hohe **Datenrate**
zusätzlicher Kommunikationskanal
- **Hände & Augen** sind frei für andere Aktivitäten
- Nutzung **existierender Übertragungskanäle** (Telefon)
- **Bewegungsfreiheit**
keine mitzuführenden Armaturen
- geringer **Raumbedarf** des Endgeräts (Mikrofon)
- funktioniert auch im **Dunkeln**
- unterstützt effizient **kollektives Problemlösen**
- **natürliche** Kommunikationsform
- wenig **Übung** erforderlich
mnemonisch · keine Kürzel

Anwendungsgebiete maschineller Spracherkennung

Haushalt	Beleuchtung, Unterhaltungselektronik, Anrufbeantworter
Büro	Aktenhaltung, Informationsabfrage, Gerätbedienung, akustische Schreibmaschine
Industrie	Qualitätskontrolle, Inventur, Versand
Zahlungsverkehr	telefonischer Bankauftragsdienst, Börsenhandel, Kreditkartenwesen
Personentransport	Fahrzeugbedienung, Fahrplanauskunft, Reservierung
Informationsdienste	Wetterbericht, Veranstaltungskalender, Gelbe Seiten
Ausbildung	Fremdsprachenerwerb, rechnergestütztes Lernen
Medizin	Diagnosesysteme, Mikroskopie, Patientenrufanlage
Militär	Waffensystemkontrolle, Flugzeugbedienung, nachrichtendienstliche Observation
Behindertenhilfe	Sprechtraining für Gehörlose, Fahrzeugbedienung, Filmuntertitelung
Sprachkommunikation	maschinelle Telefonvermittlungen, automatische Dolmetschergeräte
Datenerfassung · Gerätesteuerung · Informationsgewinnung	

Wozu automatische Spracherkennung?

Was ist maschinelles Sprachverstehen?

Taxonomie sprachverstehender Systeme

Warum ist Spracherkennung schwierig ?

Architektur eines Spracherkennungssystems

Auskunftsdialogsysteme

System: Hier ist die automatische InterCity-Auskunft. Was kann ich für Sie tun?

Nutzer: Ich will morgen abend nach Frankfurt.

System: Sie können ab Bonn fahren um [...]

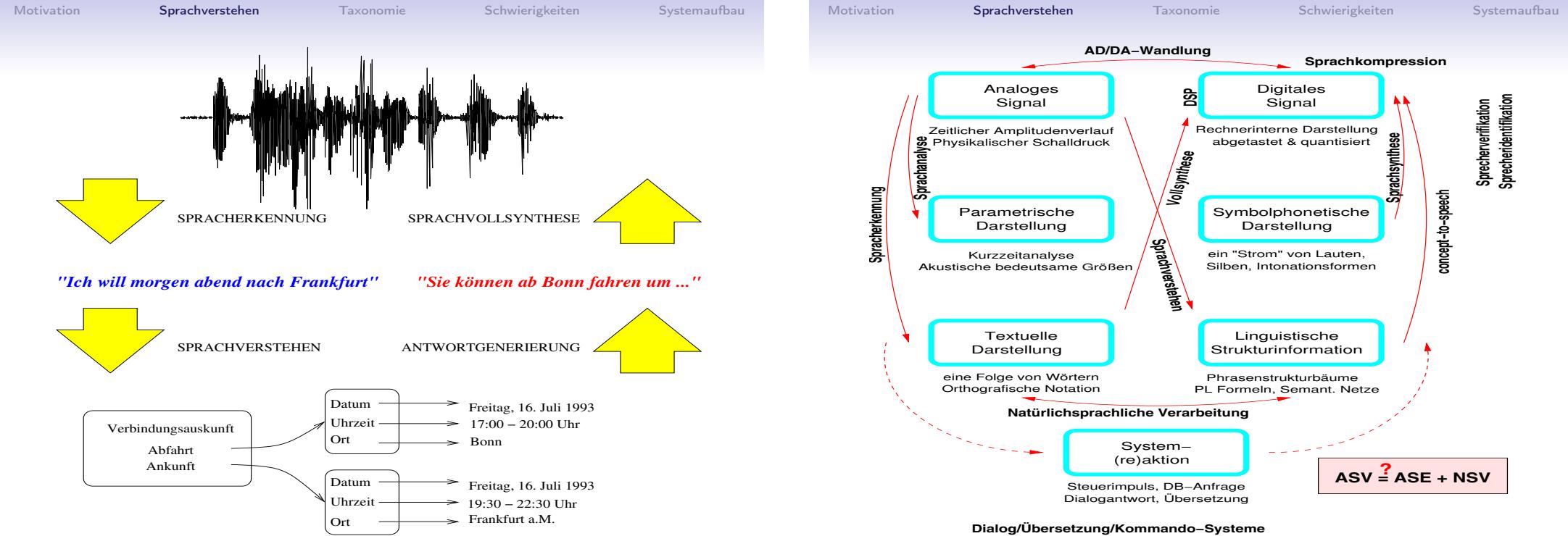
Nutzer: Gibt es auch noch einen früheren Zug?

System: Bis wann möchten Sie spätestens in Frankfurt ankommen?

Nutzer: Bis einundzwanzig Uhr.

System: Sie können ab Bonn fahren um [...]

Nutzer: Vielen Dank. Auf Wiedersehen.



Wozu automatische Spracherkennung?

Was ist maschinelles Sprachverstehen?

Taxonomie sprachverstehender Systeme

Warum ist Spracherkennung schwierig ?

Architektur eines Spracherkennungssystems

Darstellungsform isolierte Einzelwörter
kontinuierliche Sätze oder Passagen

Kommunikationsmodus Kommandos
Menü
Dialog (wechsel-/gegensprechend)
Übersetzung (MeMaMe)

Wortschatz Umfang
Schwierigkeitsgrad
Adaptivität

Wieviele Wortformen braucht der Mensch ?

Alarmstoppschalter	1					
Menü–Steuerung (J/N)	2					
Zahlen/Ziffern	10 + n					
Gerätebedienung	20 – 200					
Auskunftsdialog	500 – 2000					
Alltagssprache	8 000 – 20 000					
Diktiermaschine	20 000 – 50 000					
Deutsch ohne Fremdwörter			ca. 300 000			

Taxonomie sprachverstehender Systeme II

Sprachumfang Kommando-Set
stark formalisierte Kunstsprache
schriftsprachlich
spontansprachlich
Überdeckungsgrad
Verzweigungsfaktor, Perplexität

Diskursbereich klein · überschaubar · mittel · umfangreich · utopisch
Versandbestellung, „home banking“
Bahnauskunft, Flugreservierung
Terminabsprache, „telephone rosé“
Patentverzeichnis, ärztliche Diagnose
Telefonseelsorge

Taxonomie sprachverstehender Systeme III

Sprecherabhängigkeit ein Sprecher
feste Sprechergruppe
Sprechertypus (Geschlecht, Dialekt)
beliebige Sprecher
adaptiv

Sprecherverhalten Diszipliniertheit
Kooperativität
Vertrautheitsgrad
Stress, Disposition

Sprachsignalqualität Bandbreite
Störgeräusche
Raumakustik

Wozu automatische Spracherkennung?

Was ist maschinelles Sprachverstehen?

Taxonomie sprachverstehender Systeme

Warum ist Spracherkennung schwierig ?

Architektur eines Spracherkennungssystems

Warum ist Spracherkennung schwierig ?

Guten Morgen, Herr Hauptkomissar Thanner.
Gibt es irgendetwas Neues im Fall "Verbmobil"?

Morgen, Thanner.
Irgendwas Neues im Fall "Verbmobil"?

morgen thanner irgendwas neues im fall verbmobil

morgenthannerirgendwasneuesimfallverbmobil

moangtannairgnwasneuesimfalwerpmobiehl

moangtannairgnwasneuesimfalwerpmobiehl

moangtannairgnwasneuesimfalwerpmobiehl

moangtannairgnwasneuesimfalwerpmobiehl

der Text in "Schönschrift"
spontan gesprochene Sprache

Großschreibung?
Satzzeichen?

kontinuierliche Sprache

Aussprachevarianten

artikulatorische Verschleifung

Störungen und Verzerrungen

Fremdstimmen
„Cocktailparty“

Vier Problemfelder

KONTINUITÄT

Wahrnehmung Folge von Wörtern

Folge von Silben

Folge von Lauten

keinerlei akustische Grenzmarkierungen

Sprachsignal

KOMPLEXITÄT

Datenmengen z.B. 16 000 Abtastwerte/Sekunde

Inventare 40–50 Phoneme,

> 10 000 Silben,

100–250 k Wörter

Kombinatorik exponentielles Wachstum:

Anzahl möglicher Sätze

Restriktionen Grammatik versus Suchraum

VARIABILITÄT

Sprecher Anatomie, Dialekt, Idiolekt

Sprechweise Tempo, Lautstärke,

Kooperation, Anspannung

Wozu automatische Spracherkennung?

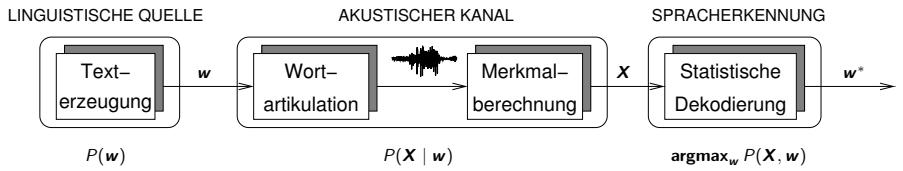
Was ist maschinelles Sprachverstehen?

Taxonomie sprachverstehender Systeme

Warum ist Spracherkennung schwierig ?

Architektur eines Spracherkennungssystems

Fundamentalformel der ASE

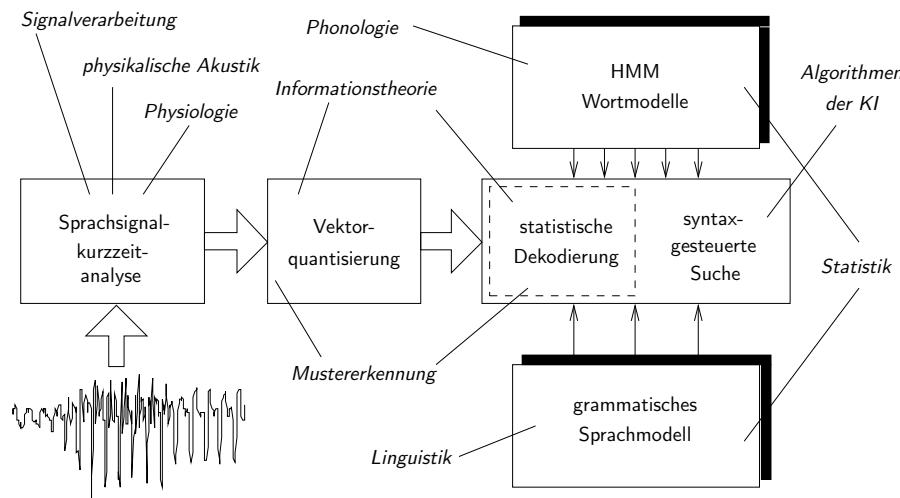


BAYES'sche Entscheidungsregel:

Suche diejenige Wortfolge w mit maximaler *a posteriori* Wahrscheinlichkeit

$$P(w|X) \stackrel{\text{def}}{=} \frac{P(w) \cdot P(X|w)}{P(X)}$$

Systemarchitektur eines Spracherkenners



Stand der Technik

- es gibt kommerzielle Systeme für die Erkennung isoliert gesprochener Wörter
10 bis wenige 100 Wörter
mit kurzer Anpassungsphase *oder* sprecherunabhängig in ruhiger Umgebung *oder* robust gegen Fremdschall
- es gibt kommerzielle Diktiermaschinen
 $\geq 20\,000$ Wörter · sprecherabhängig · isolierte Wörter
- es gibt Laborsysteme, die kontinuierlich gesprochene Sprache verstehen und eine sinnvolle Reaktion geben
1000 oder mehr Wörter
mit restriktivem Sprachmodell (Perplexität < 100)
bei sehr guter Sprachqualität
- es gelten einschneidende Beschränkungen hinsichtlich Wortschatz · Syntax · Dialekt · Problemkreis

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil III

Gesprochene Sprache

Motivation	Phonetik oooooooooooo	Phonologie oooooooooo	Akustik ooooooooooooooo	Wahrnehmung oooooooooo
------------	--------------------------	--------------------------	----------------------------	---------------------------

Motivation	Phonetik oooooooooooo	Phonologie oooooooooo	Akustik ooooooooooooooo	Wahrnehmung oooooooooo
------------	--------------------------	--------------------------	----------------------------	---------------------------

Warum ist menschliche Sprachkommunikation interessant ?

Motivation

Artikulatorische Phonetik

Phonologische Spracheinheiten und ihre Realisierung

Akustische Phonetik

Schallwahrnehmung und Lautwahrnehmung

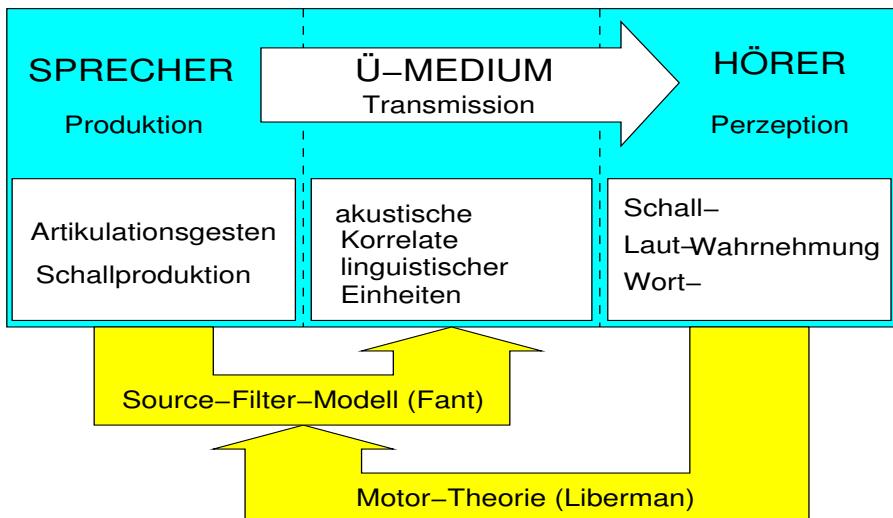
Fakt

Das menschliche Zentralnervensystem liefert den *Existenzbeweis* für einen Mechanismus zum Verstehen gesprochener Sprache.

Fakt

Es gibt aber keine Garantie für das Vorhandensein eines davon abweichenden Verfahrens.

Der Prozeß menschlicher Sprachkommunikation



Motivation

Artikulatorische Phonetik

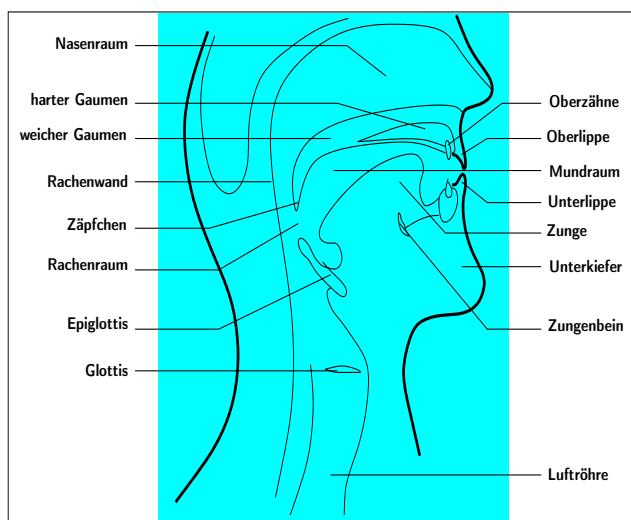
Artikulationsorgane
Artikulatorische Kategorien
Phonetisches Beschreibungssystem

Phonologische Spracheinheiten und ihre Realisierung

Akustische Phonetik

Schallwahrnehmung und Lautwahrnehmung

Die menschlichen Artikulationsorgane



- Vokaltrakt
 - Rachenraum
 - Mundraum
 - Nasenraum
- Stimmritze (Glottis)
- Lunge & Zwerchfell

Sprechen ist eine Kombination von ...

Ausatmen	Muskelbewegung
<i>Schallanregung an der Stimmritze (Glottis)</i>	<i>Ausformung des Anregungsschalls im Vokaltrakt</i>
<i>periodische</i> Schwingung bei stimmhaften Lauten durch alternierendes Öffnen & Schließen der Glottis	die Form des Ansatzrohrs (Rachenraum & Mundraum) bestimmt die Resonanzbildung
<i>Rauschen</i> bei stimmlosen Lauten durch Turbulenzen an der geöffneten Glottis	Zuschalten des Nasenraums bewirkt Auftreten von Antiresonanzen
<i>Schlaggeräusch</i> durch explosives Öffnen der Glottis	Engebildung führt zu Reibegeräusch, abrupter Änderung oder Vibration
	Lippenrundung

Artikulatorische Kategorien I

- **Luftstrommechanismus**
Im Deutschen ausschließlich **exhalatorisch** (Ausatmung)
 - **Phonation**
Je nach Grad der Verengung der Stimmritze gibt es drei Möglichkeiten:
 1. **stimmlos**
die Glottis ist offen
 2. **stimmhaft**
die Glottis wechselt periodisch zwischen offen und geschlossen
 3. **Glottisschlag**
sie wird einmalig explosionsartig geöffnet

Artikulatorische Kategorien II

- **Nasalität**
 { Heben } des **Velums** → { Abkoppeln } des Nasalbereichs
 { Senken } (Resonanzveränderung; Velum = weicher Gaumen)
 - **Engebildung**
Durch (unterlassene) Engebildung werden vier Öffnungsgrade erzielt:
 1. **völliger Verschluß**
z.B. [p] oder [n]
 2. **Friktionsenge**
z.B. [f] oder [s]
 3. **Friktionlose Enge**
z.B. [a] oder [u]
 4. **Vibration**
Alternieren zwischen 1&3, z.B. gerolltes [r]

Artikulatorische Kategorien III

- **Form des Ansatzrohrs**
Sie bestimmt wesentlich die Resonanzen bei Lauten mit frictionsloser Enge:
 1. **vorne–mittcn–hinten**
betrifft den höchsten Zungenpunkt
 2. **geschlossen–halbgeschlossen–halboffen–offen**
betrifft die vertikale Distanz zwischen Zunge und Gaumen
 3. **gerundet–ungerundet**
betrifft die Formung der Lippen

Artikulatorische Kategorien IV

- **Artikulationsort**
Er sorgt für die Enge- oder Verschlußbildung:
 1. **Aktive Artikulationsorte**
Unterlippe (*labial*),
Zungenspitze (*apikal*),
Zungenrücken (*dorsal*)
 2. **Passive Artikulationsorte**
Oberlippe (*labial*),
Zähne (*dental*),
Zahnfortsatz (*alveolar*),
harter Gaumen (*palatal*),
weicher Gaumen (*velar*),
Zäpfchen (*uvular*),
Rachenwand (*pharyngal*),
Stimmritze (*glottal*)

Artikulatorische Kategorien V

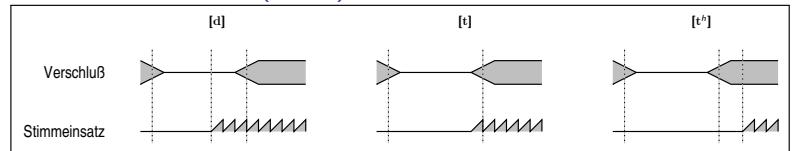
- **Zentral-lateral**

Zentrale Öffnung wie [z] oder [j] in „sagen“ und „ja“
Seitliche Öffnung wie [l] wie in „Leber“

- **Gerillt-flach**

Unterscheidet [s] in „Wasser“ von [ʃ] in „Tasche“

- **Voice Onset Time (VOT)**

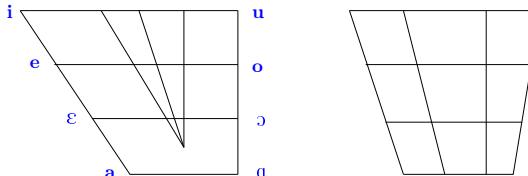


Bei den Phonen [p], [t], [k] vor einem Vokal verstreicht im Gegensatz zu [b], [d], [g] eine nicht unerhebliche Zeitspanne von **Verschlußöffnung** bis **Wiedereinsetzen** der Stimmbandschwingung

System der Vokalphone I

Definition

Vokalphone (oder Vokoide) sind zentrale, orale Laute ohne Engebildung.



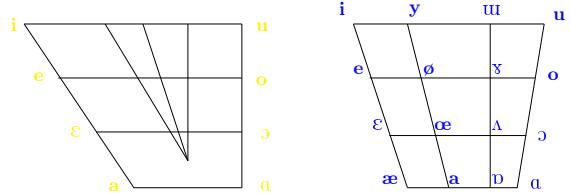
Vokalviereck mit 8 Kardinalvokalen

$$\left\{ \begin{array}{c} \text{vorn} \\ \text{hinten} \end{array} \right\} \times \left\{ \begin{array}{c} \text{offen} \\ \text{halb offen} \\ \text{halb geschlossen} \\ \text{geschlossen} \end{array} \right\}$$

System der Vokalphone II

Vokalviereck mit 16 Kardinalvokalen

$$\left\{ \begin{array}{c} \text{vorn} \\ \text{hinten} \end{array} \right\} \times \left\{ \begin{array}{c} \text{offen} \\ \text{halb offen} \\ \text{halb geschlossen} \\ \text{geschlossen} \end{array} \right\} \times \left\{ \begin{array}{c} \text{gerundet} \\ \text{ungerundet} \end{array} \right\}$$



Definition

Diphthonge sind kontinuierliche Bewegungen im Vokalraum.

Beispiel

Die Wörter „Haus“, „Beil“, „Heu“, aber auch „Tier“, „kurz“ usw.

System der Konsonantphone I

Definition

Konsonantphone (oder Kontoide) heißen alle Sprachlaute, die keine Vokalphone sind. Wir kategorisieren sie nach den Merkmalen

$$\boxed{\text{Phonation}} \times \boxed{\text{Artikulationsart}} \times \boxed{\text{Artikulationsort}}$$

Artikulationsarten:

- **Plosiv** = oraler Verschluß & Heben des Velums
- **Nasal** = oraler Verschluß & Senken des Velums
- **Frikativ** = Frikitionsenge („Reibelaut“)
- **Lateral** = friktionslose Enge & lateral
- **Halbvokal** = friktionslose Enge & Dauerlaut

System der Konsonantphone II

	bilabial	labiodental	alveolar	palatal	velar	uvular	glottal
PLOSIVE	p b		t d		k g		?
NASALE	m	n̊	n		ŋ		
FRIKATIVE	ɸ β	f v	s z / ʃ ʒ	ç j	x γ	χ ʁ	h
laterale Frikative			ɬ				
LATERALE			l				
Vibranten			r			R	
Anschläge			f			R	
HALBVOKALE	v		j	γ	ʁ		

Verschlußlösung in ein homogenes Phon:

- nasale Plosion [pm], [tn], [kj]
- frikative Plosion [ts], [tʃ], [pf], [kx]
- laterale Plosion [tl], [dl]

Minimalpaaranalyse — europäischer Strukturalismus

Definition

Zwei Wörter bilden ein **Minimalpaar**, wenn sie lediglich in einem einzigen Lautsegment voneinander abweichen.

Folgerung

Je zwei Phone, welche das Unterscheidungsmerkmal für ein Minimalpaar stellten, gehören zwangsläufig zu verschiedenen Phonemen.

Beispiel

Die Wörter „*Fisch*“ und „*Tisch*“ bilden ein Minimaalpaar. Folglich führen die Phone [f] und [t] zur Differenzierung der Phoneme /f/ und /t/.

Motivation

Artikulatorische Phonetik

Phonologische Spracheinheiten und ihre Realisierung

- Phonemsysteme
- Ausspracheverschleifung
- Prosodie

Akustische Phonetik

Schallwahrnehmung und Lautwahrnehmung

Ein Phonemsystem für das Deutsche I: die Konsonanten

Allgemein gilt für Phonemsysteme:

- endlicher** Vorrat diskreter Symbole
- minimales** System zur abstrakten lautsprachlichen Beschreibung
- sprachenspezifische** Lautinventare
- Phoneme ≠ Grapheme (Buchstaben)

	labial	alveolar	velar/palatal	glottal
PLOSIVLAUTE	p, b	t, d	k, g	h, r
NASALLAUTE	m	n	ŋ	
FRIKATIVLAUTE	f, v	s, z; ʃ, ʒ	x, j	
LATERALE		l		

Motivation

Phonetik
ooooooooooooPhonologie
oooooooo●○○Akustik
ooooooooooooooooWahrnehmung
ooooooooooo

Schwache Formen

Häufige Funktionswortkombinationen erleiden in Folge ihrer hohen Redundanz extreme Verschleifungsverluste — bis hin zur vollständigen Löschung.

Beispiele

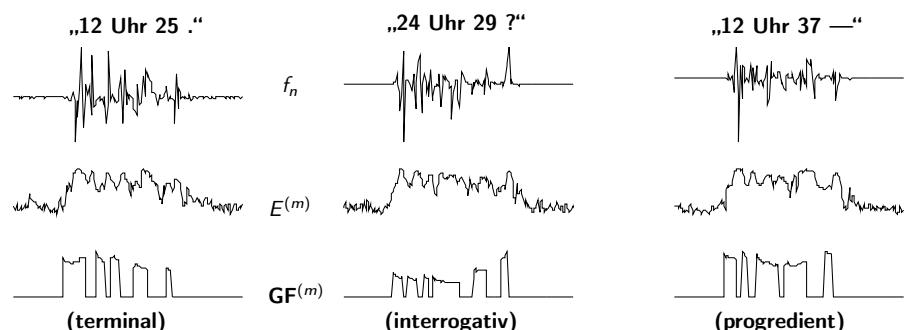
„laß ihn meckern“	[lasn]	„laß sie zetern“	[lasə]
„auf der Couch“	[aʊfə]	„auf die Couch“	[aʊfə]
„eine Frau“	[nɛ]	„was macht denn Theo?“	[n]
„fünfundzwanzig“	[fYm]	„er ist in den Käfig gesperrt“	[eəsIn]
„haben wir noch Bier?“	[hamvə]	„pack ihn ein“	[pakn]

Motivation

Phonetik
ooooooooooooPhonologie
oooooooo●○○Akustik
ooooooooooooooooWahrnehmung
ooooooooooo

Satzmodusunterscheidung

- **terminal:** „Das ist also der Dank!“
- **progredient:** „Feldberg plus 4 ... Kahler Asten minus 12 ...“
- **interrogativ:** „Hast Du auch den Hals gewaschen?“



Motivation

Phonetik
ooooooooooooPhonologie
oooooooo●○○Akustik
ooooooooooooooooWahrnehmung
ooooooooooo

Akzentuierung

- **Angriffspunkt**
betonte vs. unbetonte Silben

- Wortakzent
- Phrasenakzent
- Satzakzent

primärer / sekundärer Akzent

- **Wortunterscheidung**

„Be'leidigung“
„nicht aus A'polda“
„Ich denke also 'bin ich“
„Ak'zentver'schiebung“
„um'fahren“ ⇔ „'umfahren“

- **Hervorhebung**

emphatisch
kontrastiv
normal ('out of the blue')

„das ist ja 'un'er'hört!“
„Frank'furt, nicht Frank'reich“

Motivation

Phonetik
ooooooooooooPhonologie
oooooooo●○○Akustik
ooooooooooooooooWahrnehmung
ooooooooooo

Motivation

Phonetik
ooooooooooooPhonologie
oooooooo●○○Akustik
ooooooooooooooooWahrnehmung
ooooooooooo

Motivation

Artikulatorische Phonetik

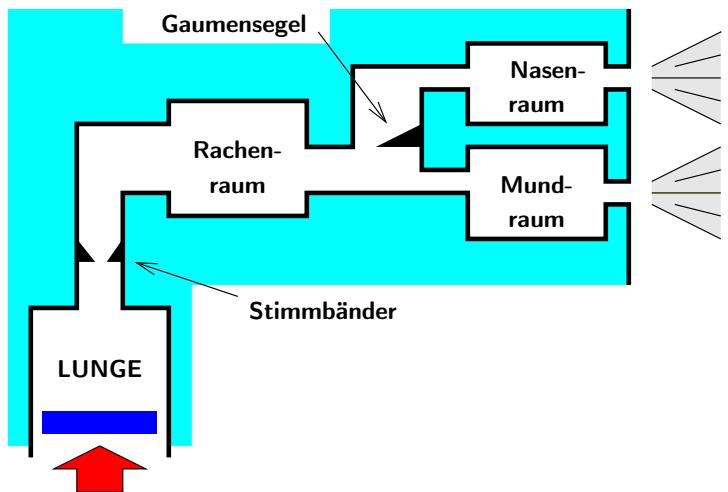
Phonologische Spracheinheiten und ihre Realisierung

Akustische Phonetik

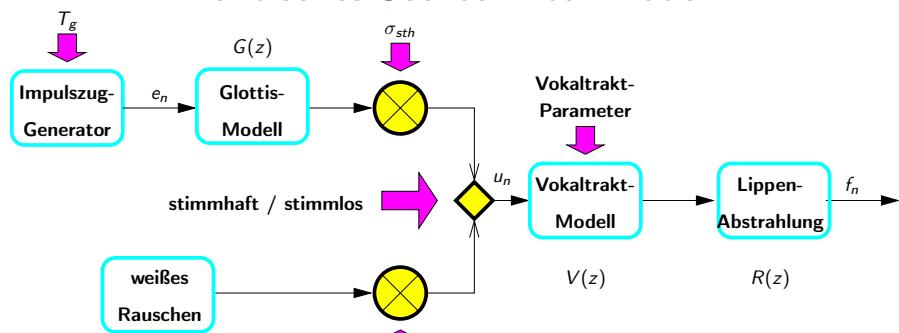
Akustische Lauterzeugungsmodelle
Akustische Phonetik

Schallwahrnehmung und Lautwahrnehmung

Schematisches Vokaltraktmodell



Fant'sches Source-Filter-Modell



FALTUNGSFORMEL im Zeit/Frequenzbereich:

$$f = u * v * r$$

$\left\{ \begin{array}{l} f \text{ diskretisiertes Schallsignal} \\ u \text{ sth/stl Anregung (excitation)} \\ v \text{ Impulsantwort des Vokaltrakts} \\ r \text{ Impulsantwort der Lippenabstrahlung} \end{array} \right.$

$$F(z) = U(z) \cdot V(z) \cdot R(z)$$

Schallerzeugungsmechanismen

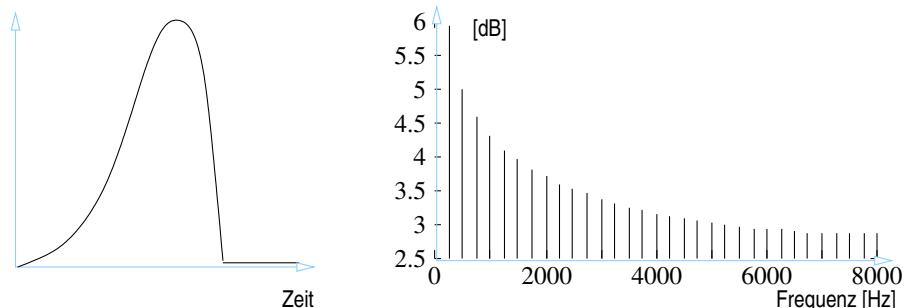
- Schallanregung an der Glottis**
Periodische Stimmbandschwingung oder turbulenzbedingtes Rauschen
- Zeitliche Veränderung der Vokaltraktform**
 $Formanten \triangleq$ Resonanzen des Vokaltrakts
- Verluste an den Vokaltraktwänden (??)**
Elastizität · Wärmeleitung · Flüssigkeitenverformung · Reibungseffekte
- Zuschaltung des Nasenraums (?)**
 $Antiformanten \triangleq$ Antiresonanzen des Nasaltrakts
- Abstrahlung des Schalls von den Lippen**
Hochpassfilter \rightsquigarrow Unterdrückung niederfrequenter Signalanteile

Glottismodell

- stimmlos:**
weißes Rauschen mit Verstärkung: $U(z) \equiv \sigma_{stl} \cdot C$
- stimmhaft:**
periodisches Signal (50–400 Hz) mit Verstärkung:

$$u = \sigma_{sth} \cdot s \star g$$

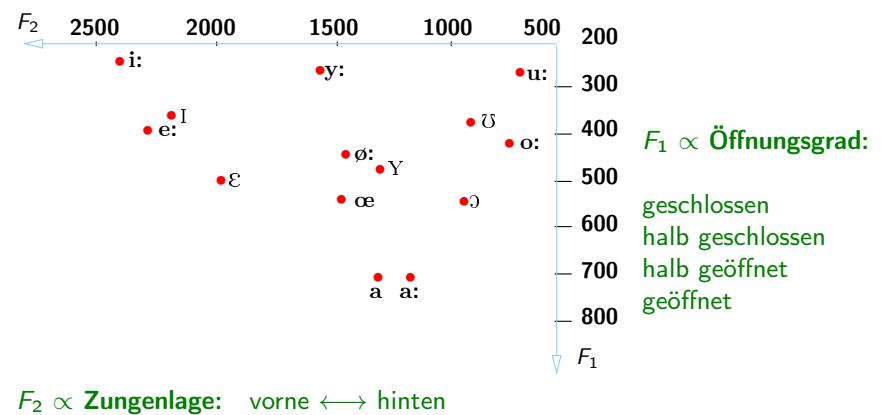
$\left\{ \begin{array}{l} s \text{ Impulszug von } F_0 = 1/T_g \text{ Hz} \\ g \text{ Grundperiodenform, z.B. s.u.} \end{array} \right.$



Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooo●oooooWahrnehmung
oooooooooo

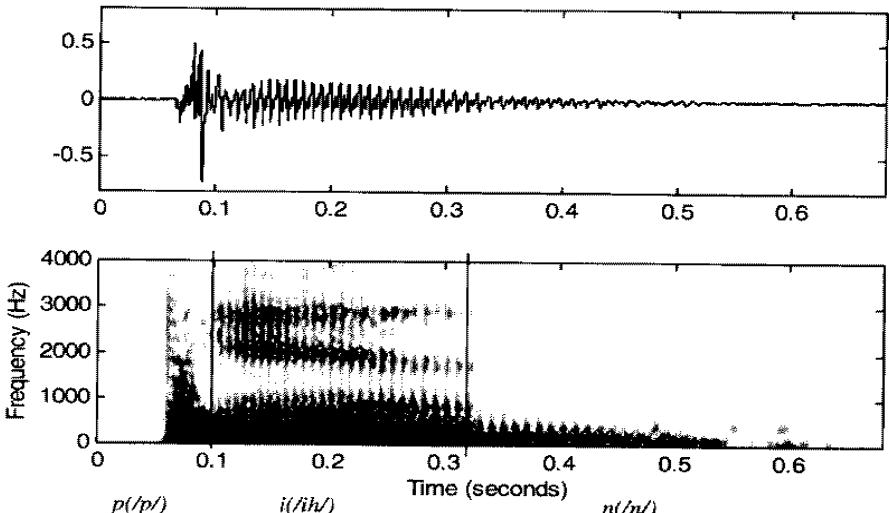
Vokalphoneme und ihre mittleren Resonanzfrequenzen



Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooo●ooooWahrnehmung
oooooooooo

Zeitsignal und Spektrogramm

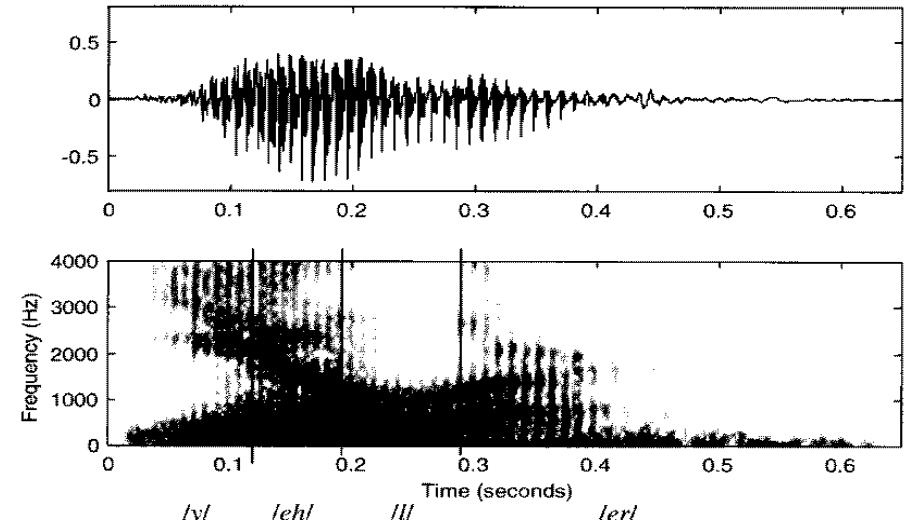


Gesprochen (engl.): „pin“ — [pIn]

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooo●oooooWahrnehmung
oooooooooo

Formantverlauf bei Vokalen und Halbvokalen

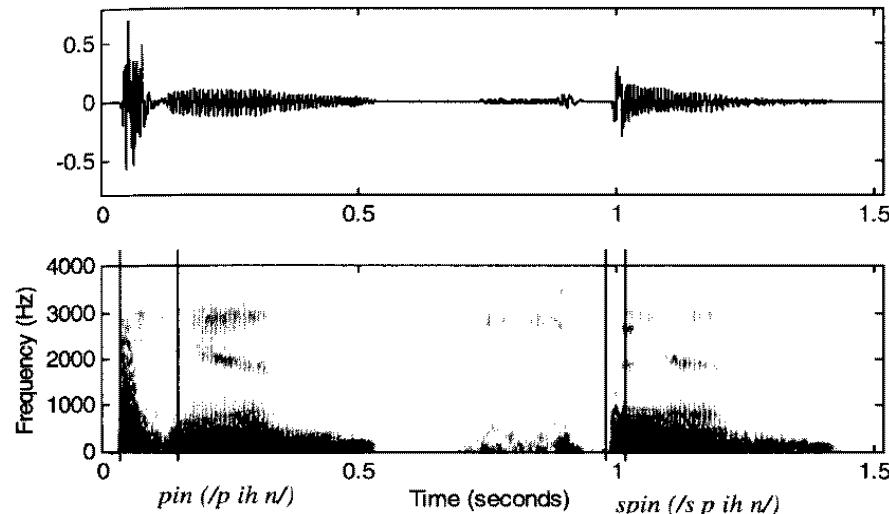


Gesprochen (engl.): „yeller“ — [jɛlə]

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooo●ooooWahrnehmung
oooooooooo

Verschlußlösungsdauer und Plosivkontext

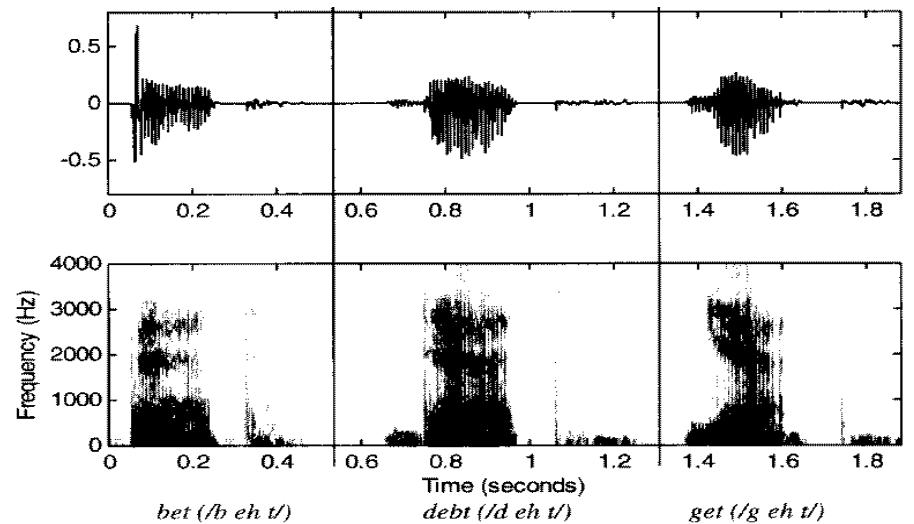


Gesprochen (engl.): „pin“ vs. „spin“ — [pIn] vs. [spIn]

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooooo●Wahrnehmung
ooooooooooo

Formantverlauf bei Plosiv-Vokal-Folgen



engl.: „bet“ vs. „debt“ vs. „get“ — [bɛt] vs. [dɛt] vs. [gɛt]

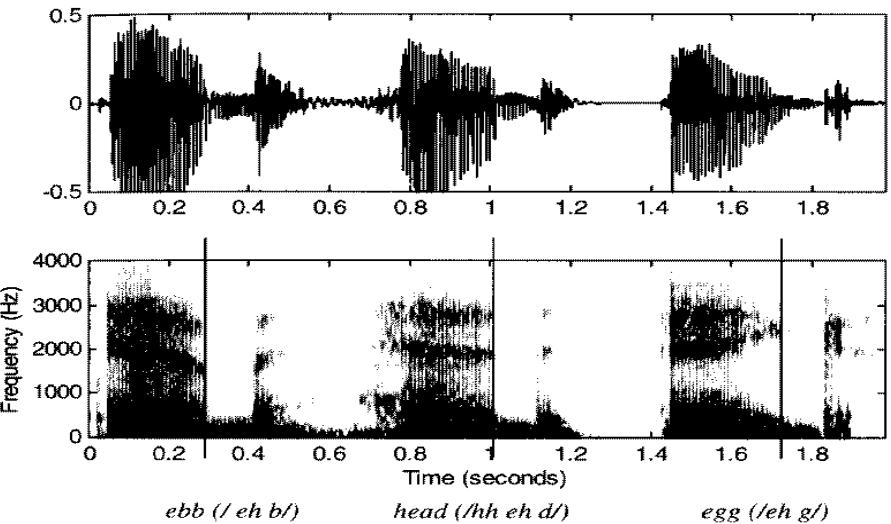
Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooWahrnehmung
ooooooooooo

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
oooooooooooooo●Wahrnehmung
ooooooooooo

Formantverlauf bei Vokal-Plosiv-Folgen



engl.: „ebb“ vs. „head“ vs. „egg“ — [ɛb] vs. [hɛd] vs. [ɛg]

Motivation

Artikulatorische Phonetik

Phonologische Spracheinheiten und ihre Realisierung

Akustische Phonetik

Schallwahrnehmung und Lautwahrnehmung

Gehörphysiologie

Menschliche Schallwahrnehmung

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooWahrnehmung
ooooooooooo

Untersuchungsgegenstand der Perzeptionsforschung

Meßtechnisch erfassbarer akustischer Reiz



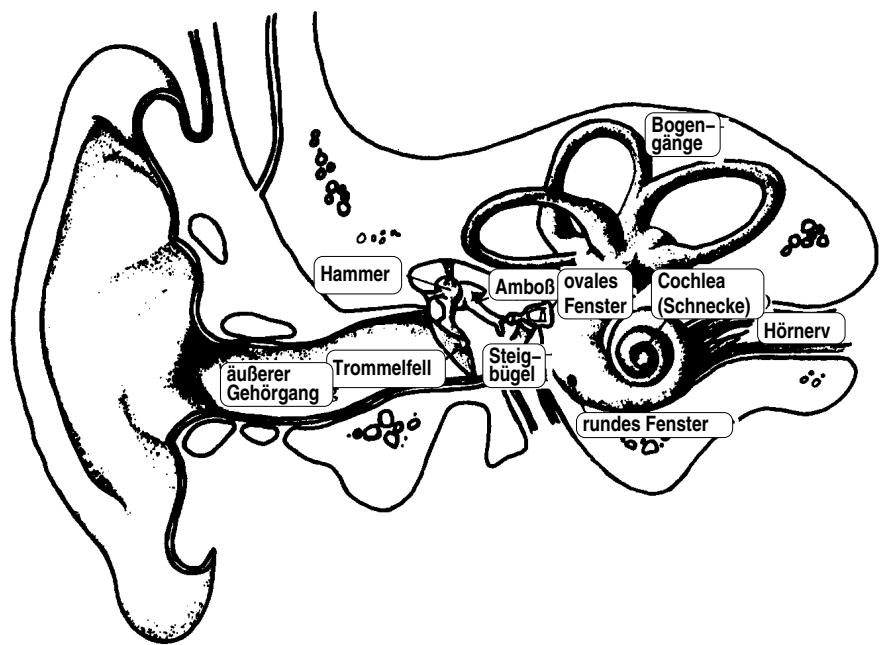
Psychoakustische Empfindungsgrößen

Reiz-Skala	Vergleichs-Skala	Verhältnis-Skala
Pegel [dB]	Lautstärke [phon]	Lautheit [sone]
Frequenz [Hz]	krit. Bandbreite [bark]	Tonheit [mel]

Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooooWahrnehmung
●ooooooooo

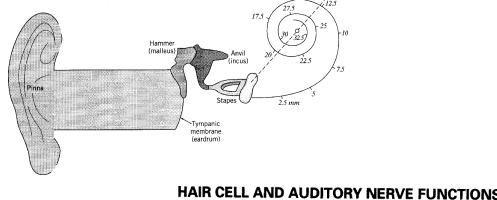
Äußeres Gehörsystem



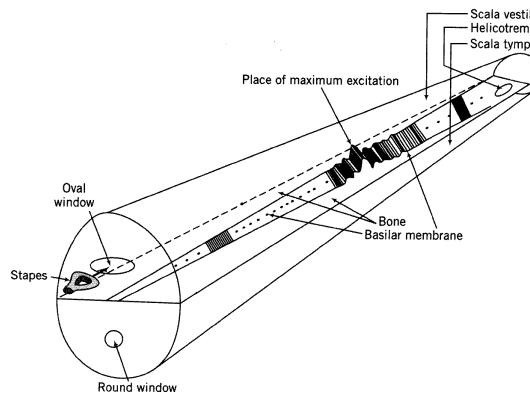
Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooooooooWahrnehmung
○●ooooooooo

Cochlea — die Hörschnecke



HAIR CELL AND AUDITORY NERVE FUNCTIONS



Schematischer Aufbau des peripheren Gehörsystems

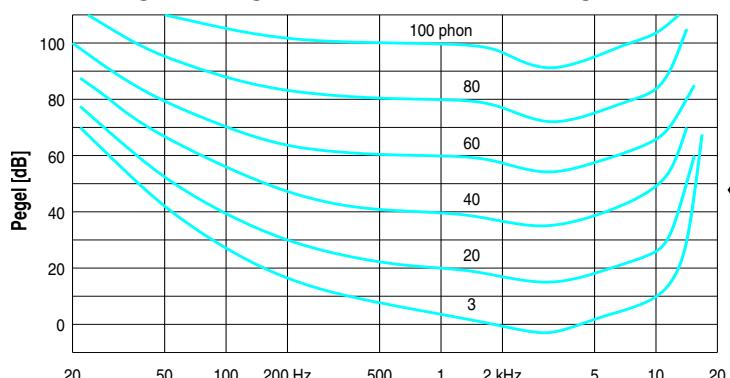
Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooooooooWahrnehmung
○○●oooooooo

Frequenzabhängige Lautstärkeempfindung

Definition

Der reine Sinuston von 1 kHz mit einem Pegel von p Dezibel besitzt eine **Lautstärke** von p Phon. Die Phonzahl aller anderen Schallereignisse ergibt sich aus dem Hörvergleich.



Motivation

Phonetik
ooooooooooooPhonologie
ooooooooooAkustik
ooooooooooooooooooooWahrnehmung
○○○●oooooooo

Wahrgenommenes Lautstärkeverhältnis

Definition

Der reine Sinuston von 1 kHz und 40 phon besitzt eine **Lautheit** von 1 sone. Ein doppelt so laut wahrgenommenes Schallereignis besitzt eine Lautheit von 2 sone usw.

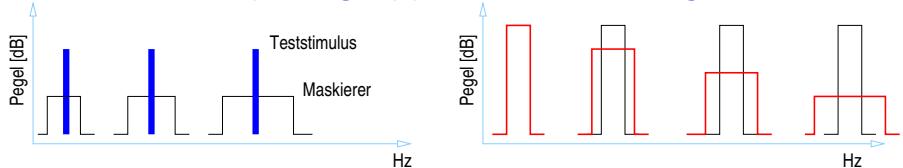
Bemerkung

Es gilt — für reine Sinustöne fester Frequenz — die folgende Faustregel:

$$\text{Lautheit [sone]} \propto I^{0.3}$$

Achtung! Im Proportionalitätsfaktor verbirgt sich noch die frequenzabhängige Pegel-Lautstärke-Umrechnung.

Frequenzgruppenwahrnehmung I



- Die Schallwahrnehmung integriert die Energien nahe beieinander liegender Frequenzanteile (Geometrie der Basilmembran!)

Frequenzgruppen, kritische Bänder

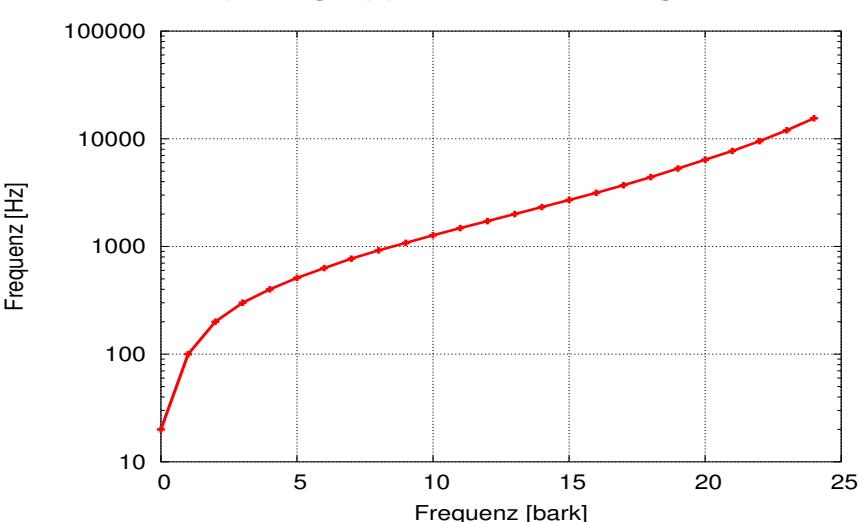
Mittenfrequenzen F_g , Bandbreiten B_g

Hörbarer Bereich (16–20 000 Hz) zerfällt in ca. 24 nicht überlappende Gruppen

- In höheren Frequenzbereichen sind auch die Bandbreiten größer:

$$B_g \approx \begin{cases} 100 \text{ Hz} & F_g \leq 1000 \text{ Hz} \\ 0.15 \cdot F_g \text{ Hz} & F_g \geq 1000 \text{ Hz} \end{cases}$$

Frequenzgruppenwahrnehmung II



Definition

Für jede beliebige Mittenfrequenz besitze die umgebende kritische Frequenzgruppe eine Bandbreite von genau **1 bark**.

Wahrgenommenes Tonhöhenverhältnis

Definition

Die **Tonheit** (*melody-frequency*) gibt die wahrgenommenen Tonhöhenverhältnisse wieder und wird in der Skaleneinheit **[mel]** angegeben. Eichpunkt ist das ungestrichene 'C':

$$131 \text{ Hz} \cong 131 \text{ mel}$$

Für doppelt so hoch empfundene Töne ist die Tonheit zu verdoppeln usw.

Bemerkung

Die Melody-Skala und die Barkhausen-Skala sind annähernd proportional:

$$1000 \text{ mel} \cong 8 \text{ bark}$$

Psychoakustische Näherungsformeln I

- Hörschwelle [dB]**
für einen reinen Sinuston von ω Hertz

$$\theta(\omega) = 3.64 \cdot \left(\frac{\omega}{1000} \right)^{-0.8} - 6.5 \cdot \exp \left\{ -0.6 \cdot \left(\frac{\omega}{1000} - 3.3 \right)^2 \right\} + 10^{-3} \cdot \left(\frac{\omega}{1000} \right)^4$$

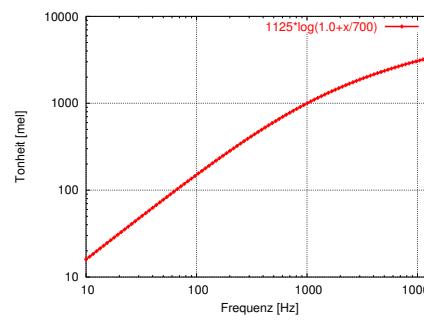
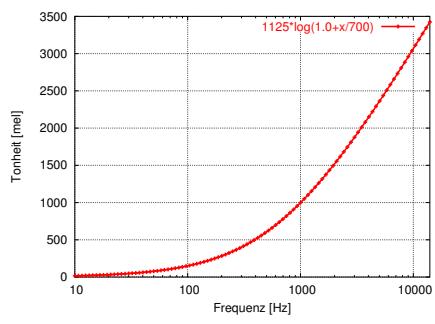
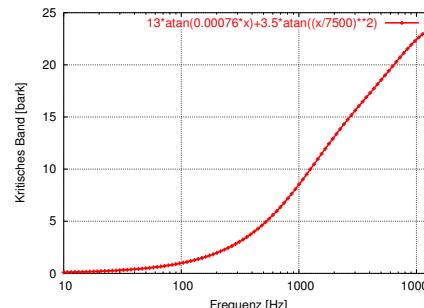
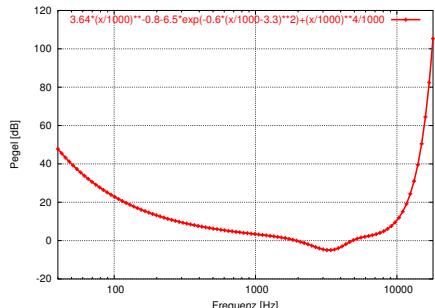
- Kritische-Band-Skala [bark]**
für die Bandmittelfrequenz ω [Hz]

$$F_{\text{bark}}(\omega) = 13 \cdot \arctan(0.00076 \cdot \omega) + 3.5 \cdot \arctan \left\{ \left(\frac{\omega}{7500} \right)^2 \right\}$$

- Tonheit [mel]**
eines reinen Tons der Höhe ω [Hz]

$$F_{\text{mel}}(\omega) = 1125 \cdot \log \left(1 + \frac{\omega}{700} \right)$$

Psychoakustische Näherungsformeln II



SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil IV

Sprachsignalanalyse

Motivation	AD/DA oooo	KZ-Analyse oooooooo	Zeitbereich ooooooo	Spektrum oooooooo	Cepstrum ooooo	$\Delta & \Delta^2$ oooo	Σ
------------	---------------	------------------------	------------------------	----------------------	-------------------	-----------------------------	----------

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

Motivation	AD/DA oooo	KZ-Analyse oooooooo	Zeitbereich ooooooo	Spektrum oooooooo	Cepstrum ooooo	$\Delta & \Delta^2$ oooo	Σ
------------	---------------	------------------------	------------------------	----------------------	-------------------	-----------------------------	----------

Zweck der Sprachsignalverarbeitung

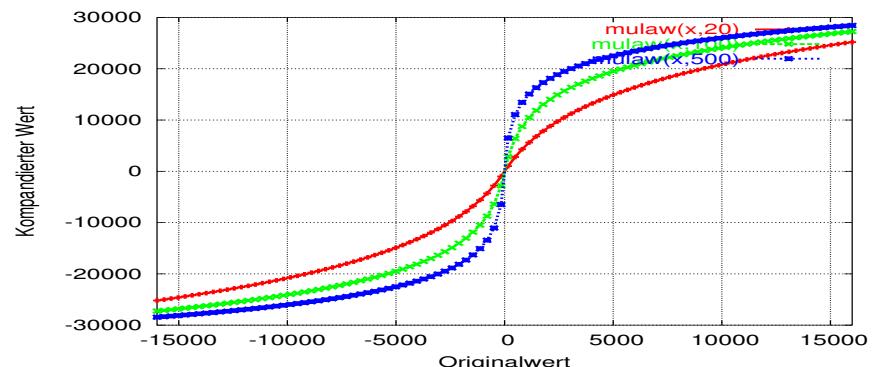
- eine **digitale Repräsentation** des Sprachschalls
- eine **Reduktion** des Datenvolumens
- die **Hervorhebung** von Eigenschaften, die zur Identifikation des Äußerungsinhaltes hilfreich sind
→ ±Sprache · Allophone · Wörter · Sätze
- die **Ausblendung** irrelevanter Einflüsse:
 - Vokaltraktanatomie
 - Sprechweise
 - Umgebungsgeräusch
 - akustischer und elektrischer Übertragungskanal

Motivation AD/DA KZ-Analyse Zeitbereich Spektrum Cepstrum Δ & Δ^2 Σ

Logarithmische Kompondierung (μ -law)

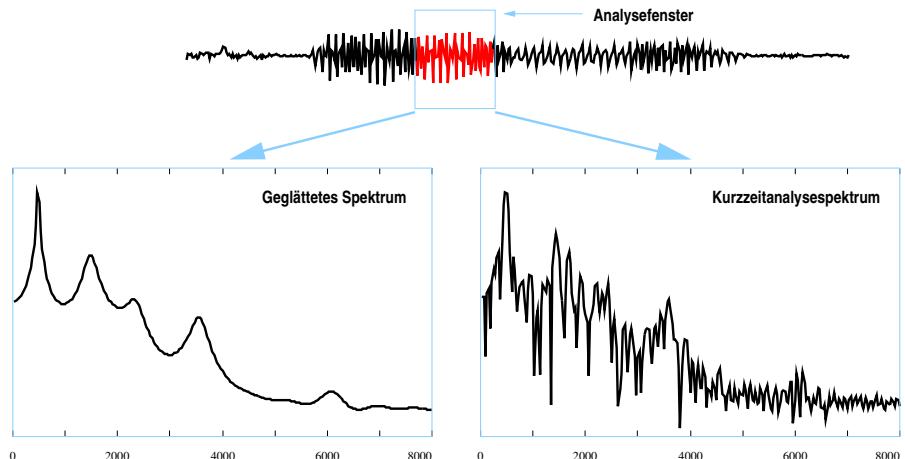
Schallsignalabtastwerte sind betragsmäßig nahezu exponentialverteilt.

$$f_n \mapsto f_{\max} \cdot \text{sign}(f_n) \cdot \frac{\log(1 + \mu \cdot \frac{|f_n|}{f_{\max}})}{\log(1 + \mu)}, \quad \mu = 100..500$$



Motivation AD/DA KZ-Analyse Zeitbereich Spektrum Cepstrum Δ & Δ^2 Σ

Analyse quasistationärer Schallsignaleigenschaften



Spektrale Analyse kurzzeitiger artikulatorischer Ereignisse durch „Herausschneiden“ kleiner Signalfenster (windowing)

Motivation AD/DA KZ-Analyse Zeitbereich Spektrum Cepstrum Δ & Δ^2 Σ

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals
Ausblenden gewichteter Signalzeitfenster

Zeitbereichsmerkmale

Spektralzerlegung

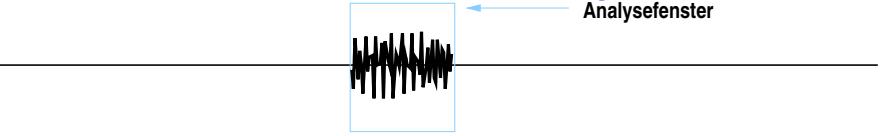
Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

Motivation AD/DA KZ-Analyse Zeitbereich Spektrum Cepstrum Δ & Δ^2 Σ

Gewichtetes Ausblenden von Signalfenstern



- Fenster an Position $m \in \mathbb{Z}$

$$\{f_n\}_n \mapsto \left\{ f_n^{(m)} \right\}_n \quad \text{mit} \quad f_n^{(m)} \stackrel{\text{def}}{=} f_n \cdot w_{m-n}$$

- Kurzzeitspektrum an Position m

$$F^{(m)}(e^{i\omega}) = \sum_{n=-\infty}^{+\infty} f_n \cdot w_{m-n} \cdot e^{-i\omega n}$$

- Faltungssatz, Verschiebung, Spiegelung:

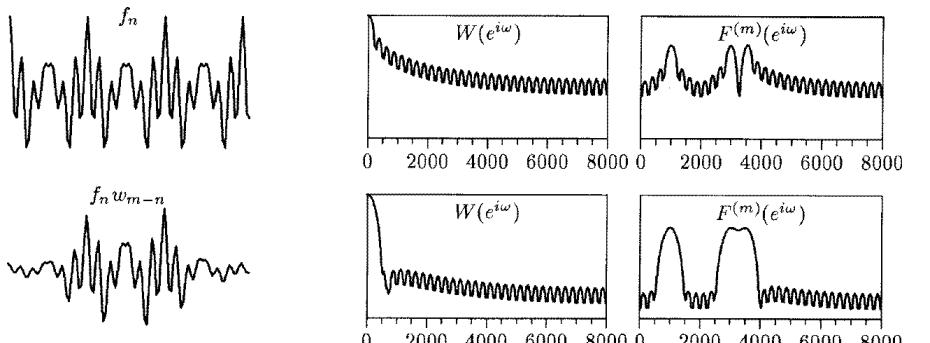
$$F^{(m)}(e^{i\omega}) = \frac{1}{2\pi} \int_{-\pi}^{+\pi} W(e^{-i\phi}) \cdot e^{-i\phi m} \cdot F(e^{i(\omega-\phi)}) d\phi$$

⇒ Jedes unvollständig konzentrierte $W(\cdot)$ „verschmiert“ $F(\cdot)$

Motivation

AD/DA
ooooKZ-Analyse
ooo●ooooZeitbereich
ooooooooSpektrum
ooooooooooooCepstrum
oooooo $\Delta & \Delta^2$
oooo Σ

Spektraler Aliasing-Effekt



Beispiel

Das synthetische Zeitsignal (links) besitzt ein Linienspektrum (Sinustöne von 1000, 3000 und 3500 Hertz), aber sowohl das *hart* (Rechteck, oben) wie auch das *weich* (Hamming, unten) herausgeschnittene Fenster besitzt ein verschmiertes Spektrum (rechts), weil die Fensterfunktionsspektren nicht vollständig auf 0 Hz konzentriert sind.

Motivation

AD/DA
ooooKZ-Analyse
ooo●ooooZeitbereich
ooooooooSpektrum
ooooooooooooCepstrum
oooooo $\Delta & \Delta^2$
oooo Σ

Fensterfunktionen I

Zeitfunktion	Fenstertyp	Dämpfung
$w_n^R = 1$	Rechteckfenster	13 dB
$w_n^P = 4 \frac{n}{N} \left(1 - \frac{n}{N}\right)$	Parabel-Fenster	22 dB
$w_n^H = 0.50 - 0.50 \cos\left(\frac{2\pi n}{N-1}\right)$	Hanning-Fenster	32 dB
$w_n^M = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$	Hamming-Fenster	43 dB
$w_n^G = \exp\left(-0.5 \cdot \left(\frac{n - N/2}{1/3 \cdot N/2}\right)^2\right)$	Gauß-Fenster	58 dB

Definition

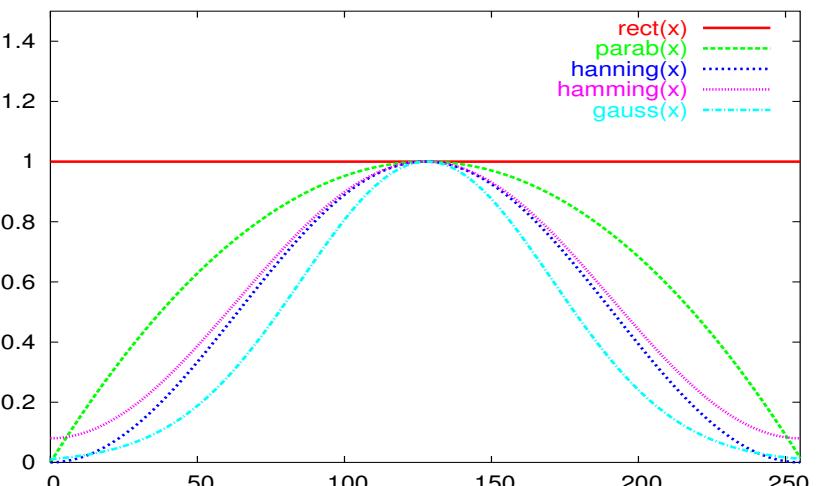
Als **spektrale Dämpfung** einer Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ bezeichnen wir das Spektralleistungsverhältnis vom Gipfel (0 Hz) zum ersten Seitenband:

$$r = 10 \cdot \log_{10} (F(\omega_0) / F(\omega_1)) \quad [\text{dB}]$$

Motivation

AD/DA
ooooKZ-Analyse
oooo●ooooZeitbereich
ooooooooSpektrum
ooooooooooooCepstrum
oooooo $\Delta & \Delta^2$
oooo Σ

Fensterfunktionen II

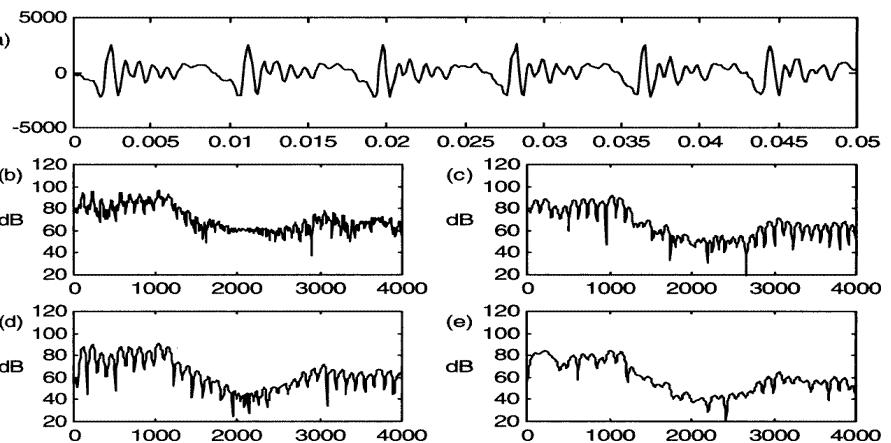


Verschiedene Fensterfunktionen, dargestellt im Zeitbereich, auf dem Trägerintervall [0, 256]

Motivation

AD/DA
ooooKZ-Analyse
ooo●ooooZeitbereich
ooooooooSpektrum
ooooooooooooCepstrum
oooooo $\Delta & \Delta^2$
oooo Σ

Spektrale Verschmierung & Harmonische Oszillation I



Schallsignal und Kurzzeitspektren des Vokals [a]

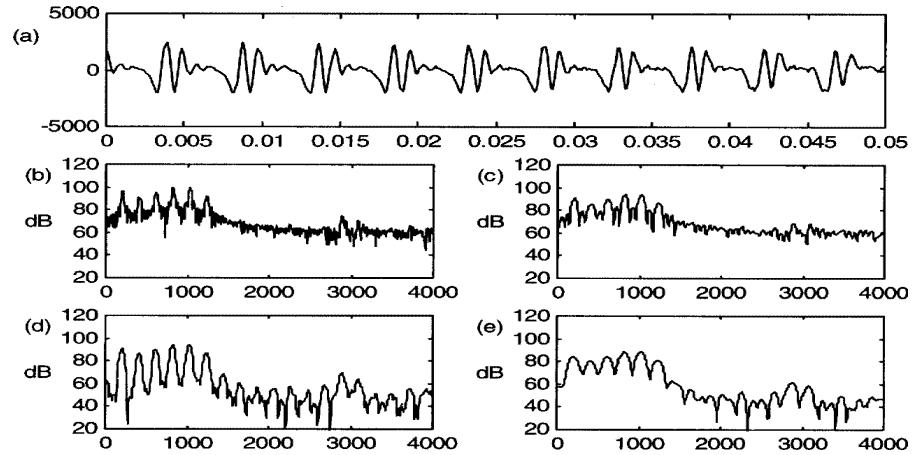
Rechteck/Hammingfenster (o/u) der Breite 30/15 ms (l/r)

- oben links: starke Verschmierung

- unten rechts: kaum Oszillation ($15 \text{ ms} \leq 2/110 \text{ Hz}$)

(männlich 110 Hz)

Spektrale Verschmierung & Harmonische Oszillation II



Schallsignal und Kurzzeitspektren des Vokals [a]

Rechteck/Hammingfenster (o/u) der Breite 30/15 ms (l/r)

· oben links und rechts: starke Verschmierung

· überall: mächtig Oszillation (Fensterbreite $\geq 2/200$ Hz)

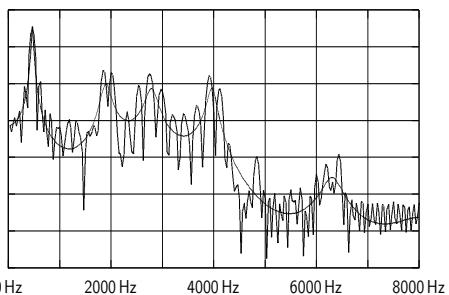
(weiblich 200 Hz)

Ziel — Spektrale Dekonvolution

$$f(t) = e(t) * h(t)$$

Sprachschall Source Filter

$$F(z) = E(z) \cdot H(z)$$



Beispiel

Logarithmiertes Kurzzeitspektrum eines Signalabschnitts

1. Spektrale Neigung
2. Formantstruktur
3. Harmonische Struktur

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Energien, Periodizitäten und Nulldurchgänge

Spektralzerlegung

Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

Schallsignalenergie

Definition

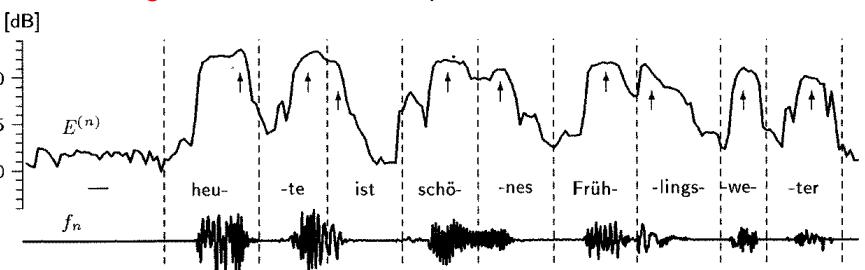
Sei $\{f_n\}_{n \in \mathbb{Z}}$ eine Abtastfolge und $\{w_n\}_{n \in \mathbb{Z}}$ eine geeignete Fensterfunktion.
Dann heißt

$$E \stackrel{\text{def}}{=} \sum_{n=-\infty}^{+\infty} |f_n|^2$$

die Langzeitenergie von f und

$$E^{(m)} \stackrel{\text{def}}{=} \sum_{n=-\infty}^{+\infty} |f_n^{(m)}|^2 = \sum_{n=-\infty}^{+\infty} |f_n w_{m-n}|^2, \quad m = 0, \pm 1, \pm 2, \dots$$

die Kurzzeitenergie von f zum Abtastzeitpunkt m .



Kurzzeitenergiefenster

- **Endliches Fenster**
der Breite N

$$E^{(m)} = \sum_{n=m}^{m+N} |f_n w_{m-n}|^2 = \sum_{k=0}^N \alpha_k \cdot |f_{m+k}|^2$$

mit den positiven Koeffizienten $\alpha_k = w_k^2$

- **Rechteckfenster**
der Breite N

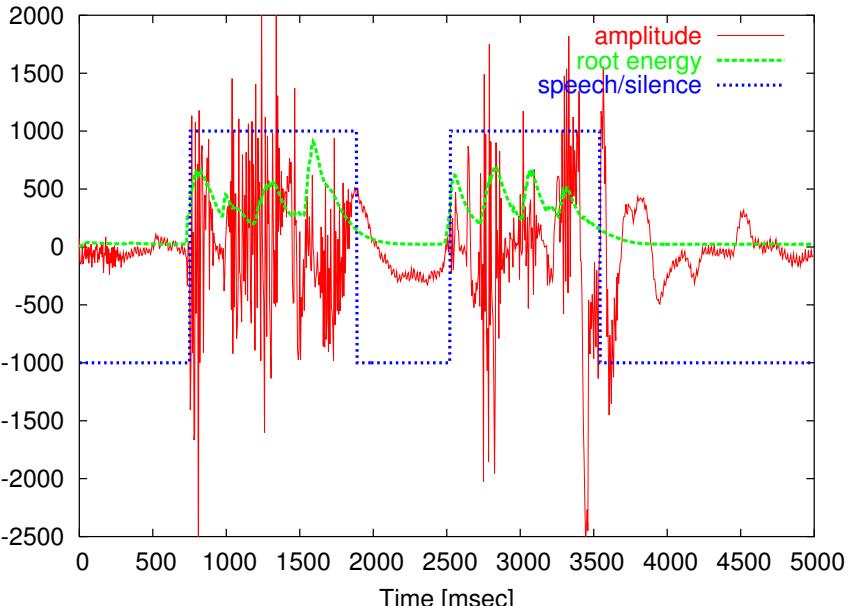
$$E^{(m)} = \sum_{k=0}^N |f_{m+k}|^2 \quad \text{bzw.} \quad E^{(m)} = \frac{1}{N} \cdot \sum_{k=0}^N |f_{m+k}|^2$$

- **Exponentiell abklingendes Fenster**
mit der Abklingrate β

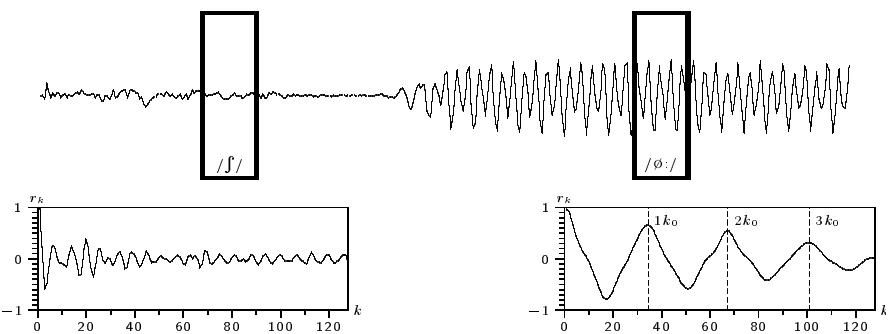
$$E^{(m)} = \frac{1}{1-\beta} \cdot \sum_{n=0}^{\infty} \beta^n \cdot |f_{m-n}|^2 \quad \Rightarrow \quad E^{(m)} = \beta \cdot E^{(m-1)} + (1-\beta) \cdot |f_m|^2$$

d.h., die Koeffizienten sind $w_k^2 = (1-\beta)^{-1} \beta^k$

Sprach-Stille-Segmentierung in Echtzeit



Periodizitäten des Sprachsignals



Definition

Sei $\{f_n\}_{n \in \mathbb{Z}}$ eine Abtastfolge. Dann heißt $\{r_k\}_{k \in \mathbb{Z}}$ mit

$$r_k \stackrel{\text{def}}{=} \sum_{n=-\infty}^{+\infty} f_n \cdot f_{n+k}, \quad k = 0, \pm 1, \pm 2, \dots$$

die **Langzeit-Autokorrelationsfunktion** von f .

Der Index k heißt **Versatz** ('lag') der Zeitachse.

Kurzzeit-Autokorrelationsfunktion

- **Allgemeine Fensterfunktion**

$$r_k^{(m)} = \sum_{n=-\infty}^{+\infty} f_n^{(m)} \cdot f_{n+k}^{(m)} = \sum_{n=-\infty}^{+\infty} f_n f_{n+k} \cdot w_{m-n} w_{m-n-k}$$

- **AKF mit Rechteckfenster**

$$r_k^{(m)} = \sum_{n=m}^{m+N-k-1} f_n \cdot f_{n+k} = \langle \mathbf{f}_m^{m+N-k-1}, \mathbf{f}_{m+k}^{m+N-1} \rangle$$

$$\text{Normierte AKF} = \bar{r}_k^{(m)} = \frac{r_k^{(m)}}{r_0^{(m)}}$$

- **Autokovarianzfunktion**

$$r_k^{(m)} = \sum_{n=m}^{m+N-1} f_n \cdot f_{n+k} = \langle \mathbf{f}_m^{m+N-1}, \mathbf{f}_{m+k}^{m+k+N-1} \rangle$$

- **Normierte Kreuzkovarianz**

$$\varrho_k^{(m)} = \frac{\langle \mathbf{f}_{m-k}^m, \mathbf{f}_m^{m+k} \rangle}{|\mathbf{f}_{m-k}^m| \cdot |\mathbf{f}_m^{m+k}|} = \cos_{\triangleleft} (\mathbf{f}_{m-k}^m, \mathbf{f}_m^{m+k})$$

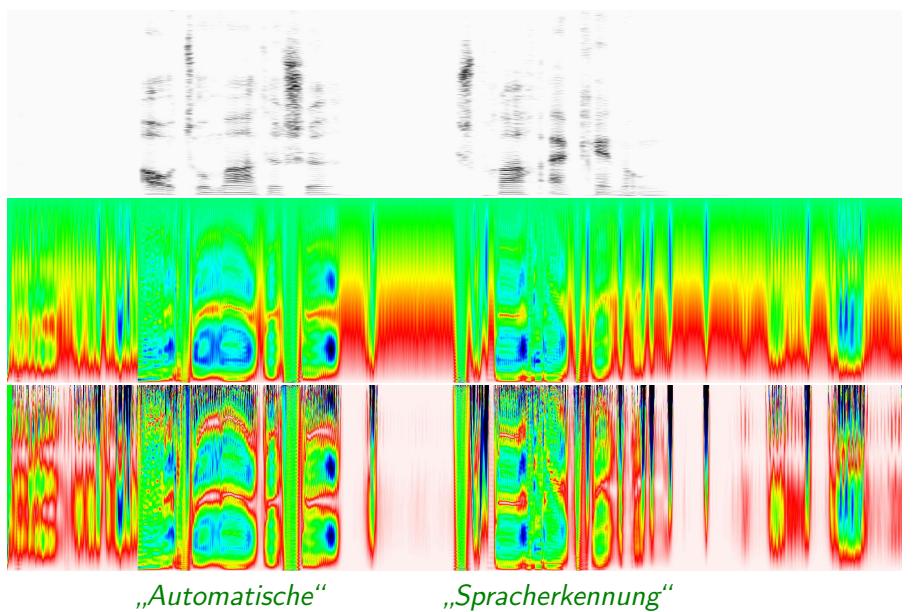
Motivation

AD/DA
ooooKZ-Analyse
ooooooooZeitbereich
oooooo●Spektrum
ooooooooCepstrum
ooooo $\Delta & \Delta^2$
oooo Σ

Motivation

AD/DA
ooooKZ-Analyse
ooooooooZeitbereich
oooooo●Spektrum
ooooooooCepstrum
ooooo $\Delta & \Delta^2$
oooo Σ

Periodogrammdarstellung



Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

DFT, Filterbank und Spektrogramm

Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

Wiener-Khintchin-Identität

Beobachtung

Die AKF-Koeffizienten sind unendliche Skalarprodukte:

$$r_k = \langle \{f_n\}_n, \{f_{n+k}\}_n \rangle$$

Bezeichnet \overleftarrow{f} die gespiegelte Folge, so gilt die Faltungsformel:

$$r = \{r_n\}_n = \{f_n\}_n * \{f_{-n}\}_n = f * \overleftarrow{f}$$

Satz

Sind $F(z)$ und $R(z)$ die z-Transformierten der Abtastfolge f und ihrer AKF r , so gilt die Identität

$$R(z) = |F(z)|^2, \quad z \in \mathbb{C}.$$

Bemerkungen

Der Satz folgt aus dem Faltungssatz und der Beziehung $\overleftarrow{F}(z) = F^*(z)$. Die AKF lässt sich effizient — mit FFT und FFT^{-1} in $O(N \log N)$ Zeit — durch Rücktransformation des Betragsquadratspektrums berechnen.

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

DFT, Filterbank und Spektrogramm

Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

Motivation

AD/DA
ooooKZ-Analyse
ooooooooZeitbereich
oooooo●Spektrum
●ooooooooCepstrum
ooooo $\Delta & \Delta^2$
oooo Σ

Die Parsevalsche Gleichung

Definition

Die **Fourier-Transformierte** $\text{FT}\{f_n\}$ einer Abtastfolge lautet

$$F(e^{i\omega}) = F(z) |_{z=e^{i\omega}} = \sum_{n=-\infty}^{+\infty} f_n \cdot e^{-i\omega n}$$

Die reellwertigen Verläufe

$$|F(e^{i\omega})| \quad \text{bzw.} \quad |F(e^{i\omega})|^2 \quad \text{bzw.} \quad \log |F(e^{i\omega})|^2$$

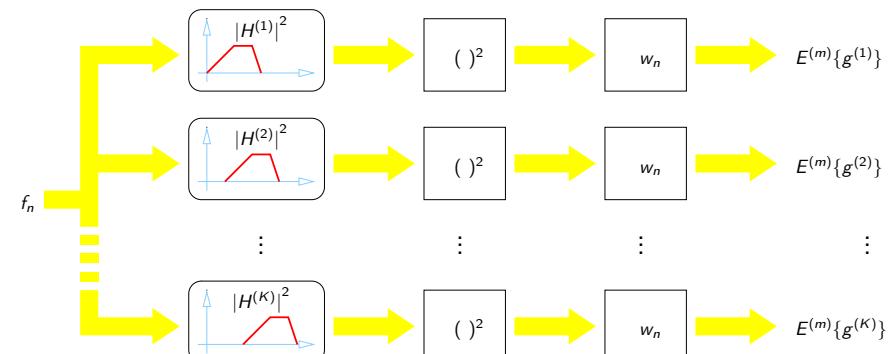
heißen (logarithmiertes) **Betrags(quadrat)spektrum** von $\{f_n\}$.

Satz

Für die spektrale Zerlegung einer Abtastfolge gilt wegen $r_0 = \text{FT}^{-1}\{|F|^2\}$ die Parsevalsche Gleichung

$$E(f) = \sum_{n=-\infty}^{+\infty} |f_n|^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |F(e^{i\omega})|^2 d\omega$$

Spektralanalyse durch eine Bank von Bandpaßfiltern



Zerlegung in Zeit- und Frequenzbereich

$$\sum_{k=1}^K H^{(k)}(z) = 1 \Rightarrow F(e^{i\omega}) = \sum_{k=1}^K G^{(k)}(e^{i\omega}) \Rightarrow f_n = \sum_{k=1}^K g_n^{(k)}$$

\rightsquigarrow je Abtastzeitpunkt K Kanalenergiwerte

Das DFT-Kurzeitspektrum

... durch gedachte periodische Fortsetzung des Fensterinhalts:

$$F_\nu^{(m)} = \sum_{n=0}^{N-1} f_{m-n} \cdot w_n \cdot e^{-2\pi i \nu n / N}, \quad \nu = 0, \dots, N-1$$

f_A	Abtastfrequenz	16 000	Hz
$1/f_A$	Abtastperiode	0.0625	ms
N	Fensterbreite	256	Δt
N/f_A	— dto. —	16	ms
f_A/N	Frequenzauflösung	62.5	Hz
$f_A/2$	Grenzfrequenz	8 000	Hz
$N/2 + 1$	Anzahl DFT-Koeffizienten	129	
$ F_\nu $	Amplitudenhöhe		N/M^2
$\arg F_\nu$	Phasenverschiebung		rad
M	Fortschaltrate	80	Δt
M/f_A	— dto. —	5	ms

Das DFT-Spektrum

Fakt

Für eine N -periodische Folge $\{f_n\}$ verschwinden alle Spektralkomponenten außer den Vielfachen der Frequenz f_A/N (\rightsquigarrow „Linienspektrum“).

Definition

Für die N -periodische Abtastfolge $\{f_n\}$ definiert

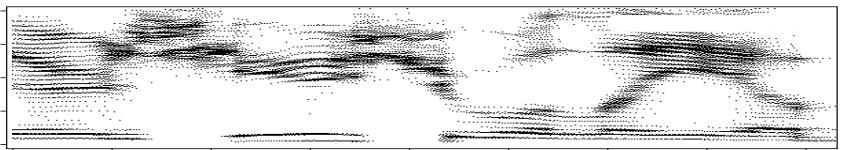
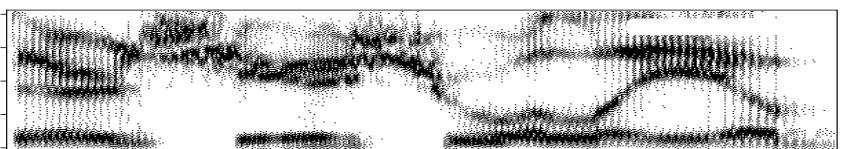
$$F_\nu = \sum_{n=0}^{N-1} f_n \cdot e^{-2\pi i \nu n / N}, \quad \nu = 0, \dots, N-1$$

die **Diskrete Fouriertransformierte** (DFT) und

$$f_n = \frac{1}{N} \sum_{\nu=0}^{N-1} F_\nu \cdot e^{2\pi i \nu n / N}, \quad n = 0, \dots, N-1$$

die **Rücktransformation** (DFT^{-1}).

Das Fourier-Spekrogramm



Breitband (oben)

- geringe Frequenzauflösung
- hohe Zeitauflösung
- senkrechte Balken im Abstand der Grundperiode

Schmalband (unten)

- hohe Frequenzauflösung
- geringe Zeitauflösung
- waagerechte Balken im Abstand der Grundfrequenz

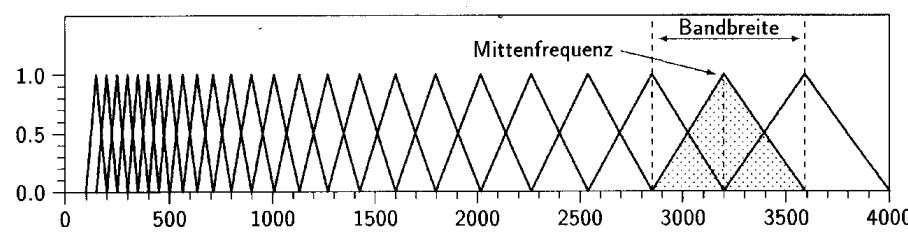
Filterbanksimulation mit der DFT

Näherungsformel

Wegen der Faltungsformel $G^{(k)} = F \cdot H^{(k)}$ für die Filterbankkanäle gilt in guter Näherung die Mittelung

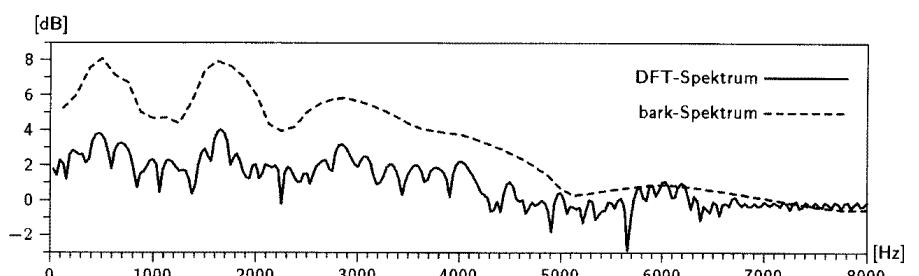
$$E^{(m)}\{g^{(k)}\} = \sum_{\nu=0}^{N-1} \eta_{k\nu} \cdot |F_\nu^{(m)}|^2$$

mit den Gewichten $\eta_{k\nu} = |H_\nu^{(k)}|^2$.

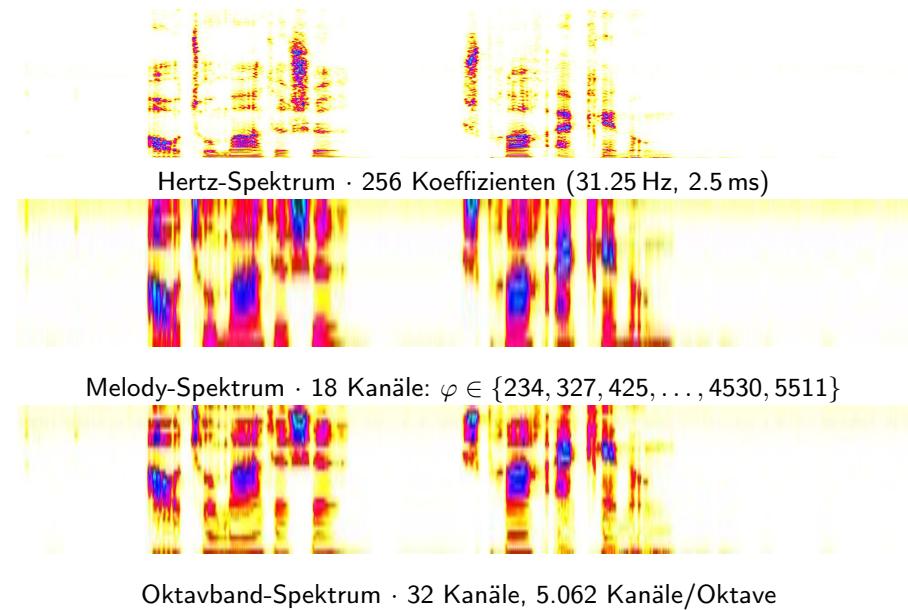


Gehörrichtige Spektralanalyse I

- Rechteckfilter mit kritischer Bandbreite (1 bark)
- Überlappende Dreieckfilter mit äquidistanten Mittenfrequenzen auf verzerrter Frequenzskala
 - $\frac{1}{3}$ -Oktav-Filter
 - Barkhausen-Skala
 - Melody-Skala
- Bandpässe nach der Basilarmembran-Abstimmkurve



Gehörrichtige Spektralanalyse II



Lautheitstransformation

- **Logarithmierung** \rightsquigarrow Pegel [dB]

$$L_k = 10 \cdot \log_{10} E_k$$

- **Potenzgesetz** (Stevens 1957)

$$L_k = (E_k)^{0.23}$$

- **Gesamtlautheit & Schallfülle** (Ruske 1984)

$$L = \sum_{k=1}^K L_k \quad \text{sowie} \quad L_{SF} = \sum_{k \leq \kappa} L_k - \sum_{k \geq \lambda} L_k$$

- **Adaptives Lautheitsmodell** (Cohen 1989)

$$\ell_k = 120 \cdot \frac{10 \log_{10} E_k - \theta_{k\downarrow}}{\theta_{k\uparrow} - \theta_{k\downarrow}} \quad [\text{phon}]$$

$$L_k = \text{const} \cdot \sqrt[3]{10^{\ell_k/10}} \quad [\text{sone}]$$

mit den frequenzgruppenspezifischen Schmerz/Ruhehörschwellen $\theta_{k\uparrow}, \theta_{k\downarrow}$
(Quantile $p = 0.975$ bzw. $p = 0.01$)

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

Homomorphe Analyse

Komplexes & reelles Cepstrum, Quefrenz, Lifter

Dynamische Merkmale

Beispielaufbau

Das reelle Cepstrum

Berechnung der Kurzzeitanalysekoeffizienten:

$$c_q^{(m)} = \frac{1}{N} \sum_{\nu=0}^{N-1} \log |F_\nu^{(m)}| \cdot e^{i2\pi\nu q/N}, \quad 0 \leq q < N$$

Lemma

Die reellen Cepstrumkoeffizienten lassen sich durch die **diskrete Kosinustransformation (DCT)** berechnen.

$$c_0^{(m)} = \sqrt{\frac{2}{N}} \sum_{\nu=0}^{N/2-1} \log |F_\nu^{(m)}|$$

$$c_q^{(m)} = \sqrt{\frac{4}{N}} \sum_{\nu=0}^{N/2-1} \log |F_\nu^{(m)}| \cdot \cos \frac{\pi q(2\nu+1)}{N}, \quad q = 1, \dots, \frac{N}{2}$$

Grund

Mit dem Betragsspektrum ist auch das reelle Cepstrum *reellwertig und symmetrisch*.

Dekonvolution durch homomorphe Analyse

$$\{f_n\} = \{e_n\} * \{h_n\}$$

$$\text{FT}\{f_n\} = \text{FT}\{e_n\} \cdot \text{FT}\{h_n\}$$

$$\log \text{FT}\{f_n\} = \log \text{FT}\{e_n\} + \log \text{FT}\{h_n\}$$

$$\text{FT}^{-1}\{\log \text{FT}\{f_n\}\} = \text{FT}^{-1}\{\log \text{FT}\{e_n\}\} + \text{FT}^{-1}\{\log \text{FT}\{h_n\}\}$$

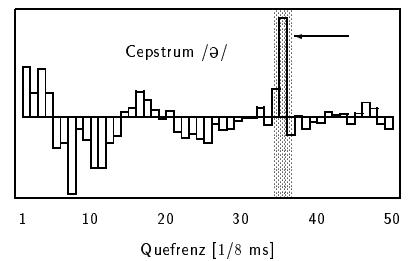
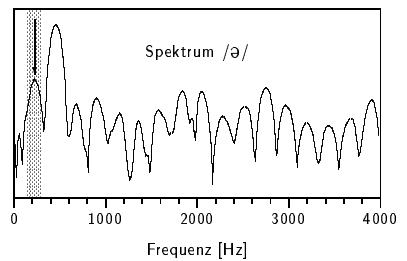
Definition

Ist $\{f_n\}_{n \in \mathbb{Z}}$ eine Abtastfolge, so heißt

$$\text{FT}^{-1}\{\log \text{FT}\{f_n\}\} \quad \text{bzw.} \quad \text{FT}^{-1}\{\log |\text{FT}\{f_n\}|\}$$

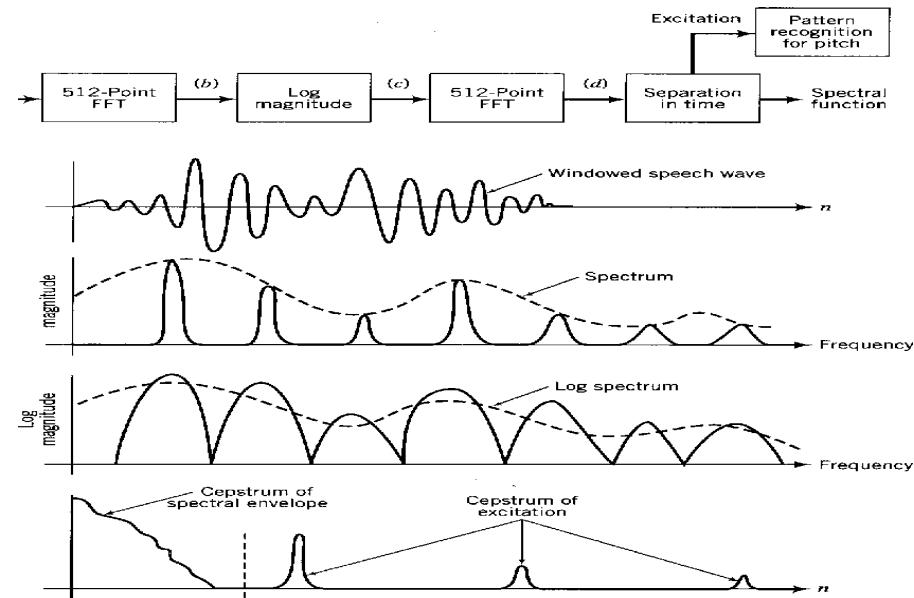
das **Cepstrum** bzw. das **reelle Cepstrum** des Signals.

Das Cepstrum zur Periodizitätsanalyse des Spektrums



Abtastrate	f_A	8 000	Hz
Frequenzauflösung	f_A/N	?	Hz
Grenzfrequenz	$f_A/2$	4 000	Hz
$\{F_\nu\}$ -Schwingungsperiode bei c_q	N/q	$\frac{N}{37}$	
— dto. —	f_A/q	ca. 216	Hz
Quefrenz des Koeffizienten c_q	q/f_A	4.625	ms
Quefrenzauflösung	$1/f_A$	1/8	ms

Grundfrequenzschätzung mit dem reellen Cepstrum



Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

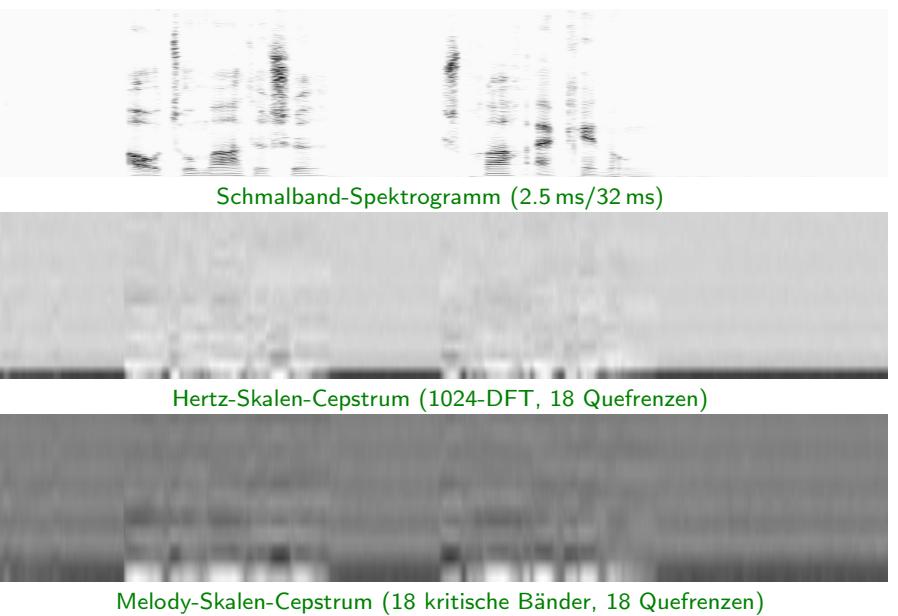
Homomorphe Analyse

Dynamische Merkmale

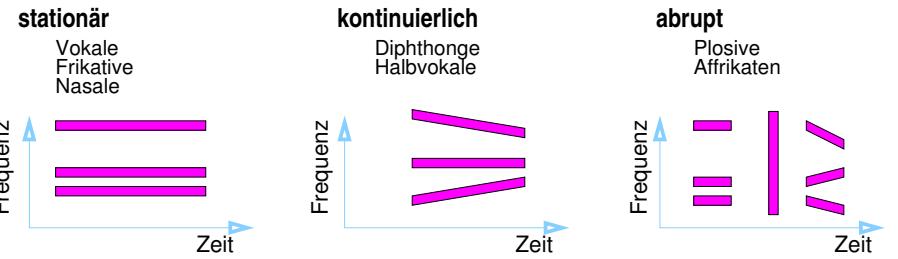
Differenzen, Ableitungen

Beispielaufbau

Spektrogramm versus Cepstrogramm



Zeitliche Veränderung des Sprachsignals

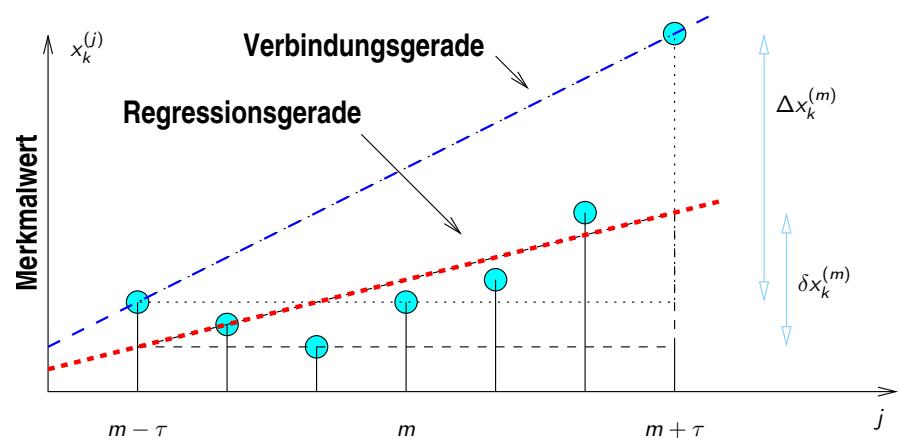


- Statische Merkmale**
Momentane spektrale Eigenschaften des Signals
- Dynamische Merkmale**
Zeitlicher Verlauf der Kurzzeitparameter
- {Sensibilisierung, Immunisierung} gegenüber {schnellen, langsamem} Parameterbewegungen**

Steigung des Merkmalverlaufs

Statische Sprachsignalmerkmale im Zeitfenster m :

$$\mathbf{x}^{(m)} = (x_1^{(m)}, x_2^{(m)}, x_3^{(m)}, \dots, x_{d-1}^{(m)}, x_d^{(m)})$$



Steigung der Ausgleichsgeraden

Lemma

Es sei $x_k^{(m)}$ der Wert des k -ten Merkmals zur Zeit m und $\tau \in \mathbb{N}$. Für die Steigung der Ausgleichsgeraden durch den Verlauf

$$x_k^{(m-\tau)}, \dots, x_k^{(m-1)}, x_k^{(m)}, x_k^{(m+1)}, \dots, x_k^{(m+\tau)}$$

des k -ten Merkmals gilt die Formel

$$\delta x_k^{(m)} = \frac{\sum_{j=-\tau}^{\tau} j \cdot x_k^{(m+j)}}{\sum_{j=-\tau}^{\tau} j^2}.$$

Algorithmus

Der obige Zählerausdruck lässt sich durch folgende Doppelrekursion auf hocheffiziente Weise **inkrementell** auswerten:

$$\begin{aligned} y_k^{(m)} &= y_k^{(m-1)} + \tau \cdot (x_k^{(m+\tau)} - x_k^{(m-\tau-1)}) - z_k^{(m-1)} \\ z_k^{(m)} &= z_k^{(m-1)} + (x_k^{(m+\tau)} - x_k^{(m-\tau)}) \end{aligned}$$

Differenzen als diskretisierte Ableitungen

Definition

Es sei $x_k^{(m)}$ der Wert des k -ten Merkmals zur Zeit m und $\tau \in \mathbb{N}$.

Die Werte

$$\Delta x_k^{(m)} = x_k^{(m+\tau)} - x_k^{(m-\tau)}$$

heißen **erste Differenzen** der Ordnung τ .

Die Werte

$$\begin{aligned} \Delta^2 x_k^{(m)} &= \Delta x_k^{(m+\tau)} - \Delta x_k^{(m-\tau)} \\ &= x_k^{(m+2\tau)} - 2x_k^{(m+\tau)} + x_k^{(m-2\tau)} \end{aligned}$$

heißen **zweite Differenzen** der Ordnung τ .

Die Werte

$$\begin{aligned} \Delta^3 x_k^{(m)} &= \Delta^2 x_k^{(m+\tau)} - \Delta^2 x_k^{(m-\tau)} \\ &= x_k^{(m+3\tau)} - 3x_k^{(m+\tau)} + 3x_k^{(m-\tau)} - x_k^{(m-3\tau)} \end{aligned}$$

heißen **dritte Differenzen** der Ordnung τ .

Motivation

Diskretisierung

Kurzzeitanalyse des Schallsignals

Zeitbereichsmerkmale

Spektralzerlegung

Homomorphe Analyse

Dynamische Merkmale

Beispielaufbau

An Stelle einer Zusammenfassung

EXEMPLARISCHE BERECHNUNGSFOLGE ZUR MERKMALGEWINNUNG

Delta-Melody-Cepstrum

1 Diskretisierung

16 kHz Abtastung, 16 bit Quantisierung

2 Spektrale Kurzzeitanalyse

25.6 msec Hammingfenster, 10 msec Fortschaltung, 512-Punkte FFT

3 Frequenzskalenverzerrung

Betragsspektrumintegration über 24-kanalige Oktav-Dreieckfilterbank

4 Lautheitstransformation

Logarithmierung der Kanalenergien

5 Dekonvolution

Berechnung der ersten 12 reellen Cepstrumkoeffizienten durch DCT

6 Zeitliche Dynamik

Hinzufügen der ersten und zweiten Differenzen

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 7. Dezember 2018

Teil V

Hidden Markov Modelle

Motivation	DTW	HMM/Definition	FA/BA	MAP/Viterbi	$\mathcal{N}(x \mu, S)$	Baum-Welch	Robustheit	Σ
Motivation	oooooooooooo	oooooooooooo	ooooooo	ooooooo	oooooooooooo	oooooooooooo	ooooooo	ooooooo

Dynamic Time Warping
Hidden Markov Modell

Produktionswahrscheinlichkeiten
Aufdeckung der verborgenen Zustandsfolge

Gaußsche Mischverteilungen
Lernen der HMM-Parameter
Robuste Schätzverfahren

Motivation	DTW	HMM/Definition	FA/BA	MAP/Viterbi	$\mathcal{N}(x \mu, S)$	Baum-Welch	Robustheit	Σ
Erkennung isoliert gesprochener Wörter	oooooooooooo	oooooooooooo	ooooooo	ooooooo	oooooooooooo	oooooooooooo	ooooooo	ooooooo

GEGEBEN:

- Erkennungswortschatz: $\mathcal{V} = \{W_1, \dots, W_L\}$
- Merkmalstrom $\mathbf{X} = x_1, \dots, x_T$ des Eingabeschalls



GESUCHT:

- das mutmaßlich gesprochene Wort $W_\ell \in \mathcal{V}$

ZIELVORGABE:

- Minimierung der **Wortfehlerrate**
- Echtzeitverarbeitung

Motivation

Dynamic Time Warping

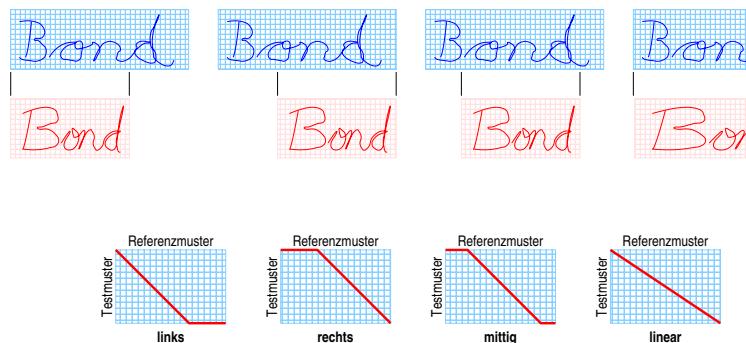
Einzelworterkenner · Minimum-Abstand-Klassifikation · DTW-Algorithmus

Hidden Markov Modell

Gaußsche Mischverteilungen

Lernen der HMM-Parameter

Skalenausrichtung zwischen Sequenzdaten



Akkumulation lokaler Distanzen

entlang einem Gitterpfad (**Skalenverzerrungsfunktion**):

$$D_\phi(\mathbf{X}, \mathbf{Y}) = \sum_{\tau=1}^{T_\phi} d(\mathbf{x}_{\phi_1(\tau)}, \mathbf{y}_{\phi_2(\tau)}) , \quad \phi : [1, T_\phi] \rightarrow [1, T_X] \times [1, T_Y]$$

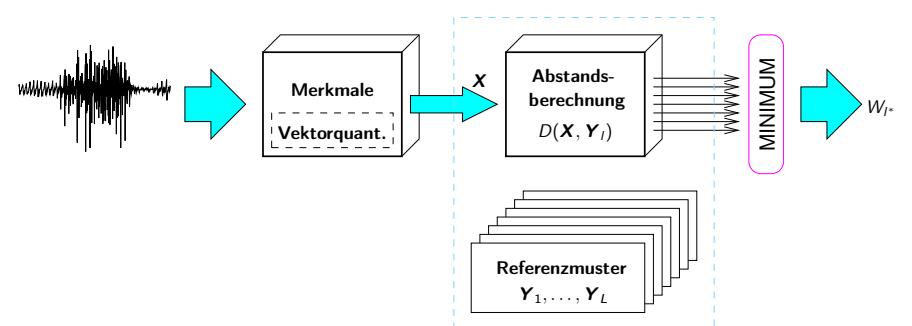
Einzelworterkennung durch Referenzmustervergleich

- Minimum-Abstand-Klassifikation:

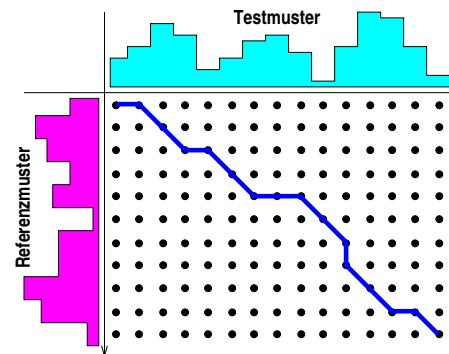
$$\ell^*(\mathbf{X}) = \operatorname*{argmin}_{\ell=1, \dots, L} D(\mathbf{X}, \mathbf{Y}_\ell)$$

- ### • Multireferenz-Worterkenner:

$$\ell^*(\mathbf{X}) = \operatorname{argmin}_{\ell=1..L} \min_{m=1..M_\ell} D(\mathbf{X}, \mathbf{Y}_{\ell,m})$$



Mustervergleich zwischen Sequenzdaten



Dynamic Time Warping

Kumulative Distanz bezüglich optimaler Zeitverzerrungsfunktion

$$D(\mathbf{X}, \mathbf{Y}) \stackrel{\text{def}}{=} \min_{\phi \in \Phi} D_\phi(\mathbf{X}, \mathbf{Y})$$

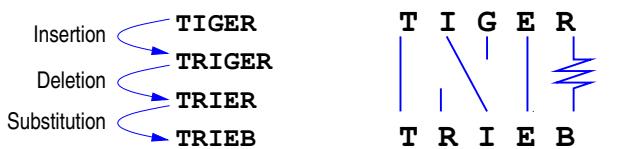
Kombinatorische Suche — Aufwand $O(3^T)$

Levenshtein-Abstand

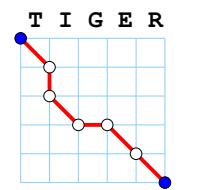
Mustervergleich zwischen Zeichenketten

Elementaroperationen auf Zeichenketten

- Ersetzung eines Zeichens durch ein anderes
- Löschung eines Zeichens
- Einfügung eines Zeichens



substitution
deletion
insertion



Definition

Ist \mathcal{A} ein endliches Alphabet und sind v, w zwei Zeichenfolgen aus \mathcal{A}^* , so bezeichnet der **Levenshtein-Abstand** $d^{\text{lev}}(v, w)$ die minimale Anzahl von Elementaroperationen, mit denen v in w überführt werden kann.

Dynamic Time Warping Abstand

Rekursives Berechnungsschema (Itakura 1975 und Sakoe 1978)

	y_1	y_2	y_3	y_4
x_1	1	4	5	8
x_2	4	3	2	7
x_3	7	4	9	0

lokale Distanzen

	y_1	y_2	y_3	y_4
x_1	1	5	10	18
x_2	5	4	6	13
x_3	12	8	13	6

kumulative Distanzen

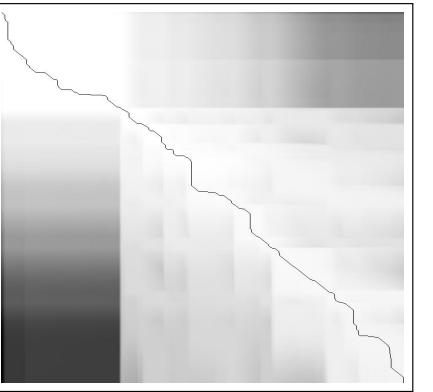
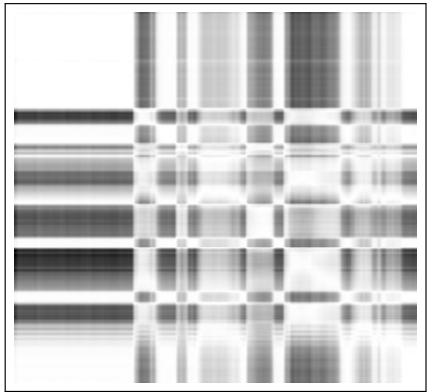
	y_1	y_2	y_3	y_4
x_1				
x_2				
x_3				

Rückwärtszeiger

	y_1	y_2	y_3	y_4
x_1				
x_2				
x_3				

lokale Transitionen

Lokale vs. kumulative Abstände & optimale Ausrichtung



lokale Distanzen:

$$d_{st} = \|x_s - y_t\|$$

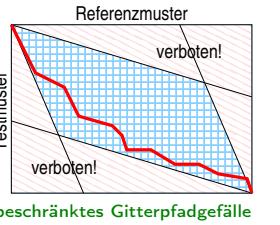
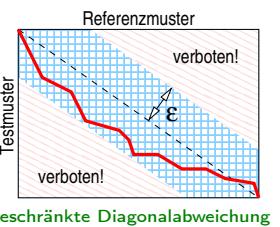
kumulative Distanzen:

$$\tilde{d}_{st} = D(\mathbf{X}_1^s, \mathbf{Y}_1^t)$$

Beispielwort: „Edmund Stoiber“ (2×) / 'ɛtmʊnt'ʃtɔɪ̯bɐ̯/

Zulässige Skalenverzerrungsfunktionen

Verbot unerwünschter Ausrichtungen — Einsparung von Zeit und Speicher

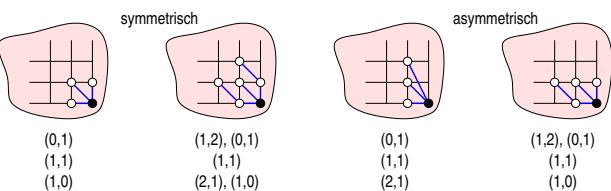


Globale Einschränkungen

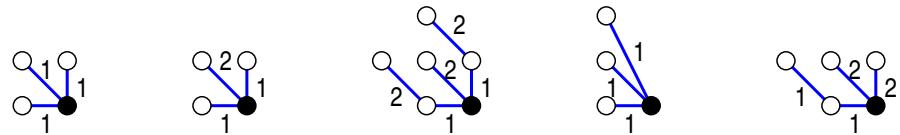
Welche Gitterpunkte werden für den ϕ -Verlauf gesperrt?

Lokale Einschränkungen

Welche Nachbarkonfigurationen eines Pfadknotens $\phi(\tau)$ sind erlaubt?



Gewichtete Skalenverzerrungsfunktionen



Problem

Diagonalferne Ausrichtungen ϕ besitzen größere Lauflänge T_ϕ ; ihre Distanzsumme nimmt tendenziell höhere Werte an.

Lösung

Minimiere Distanz**mittel** statt Distanzsumme

Optimalitätsprinzip (\rightsquigarrow DP)

nicht gültig für Mittelwertbildung !

Lokal gewichtete kumulative Distanz

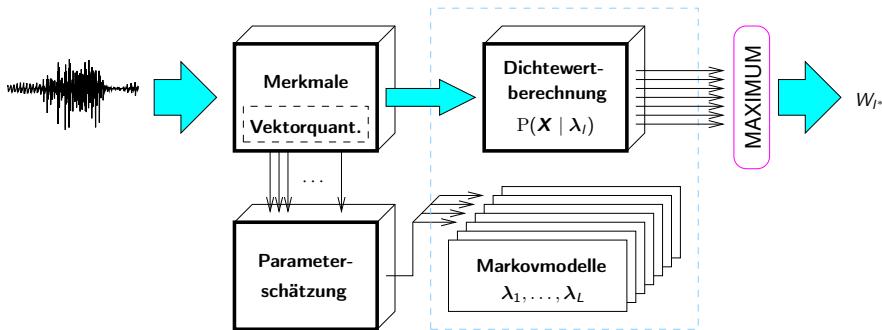
Die Gewichtsumme aller Pfade ist konstant:

$$D_\phi(\mathbf{X}, \mathbf{Y}) = \sum_{\tau=1}^{T_\phi} w_{\phi(\tau), \phi(\tau-1)} \cdot d(x_{\phi_1(\tau)}, y_{\phi_2(\tau)})$$

Einzelworterkennung mit Wort-HMMs

- Erkennung mit der Bayesregel:

$$\ell^* = \underset{\ell=1..L}{\operatorname{argmax}} P(W_\ell | \mathcal{X}) = \underset{\ell=1..L}{\operatorname{argmax}} \frac{P(W_\ell) \cdot P(\mathcal{X} | \lambda_\ell)}{P(\mathcal{X})}$$



Motivation

Dynamic Time Warping

Hidden Markov Modell

Einzelworterkenner · Definition eines HMM · Topologien für die ASE

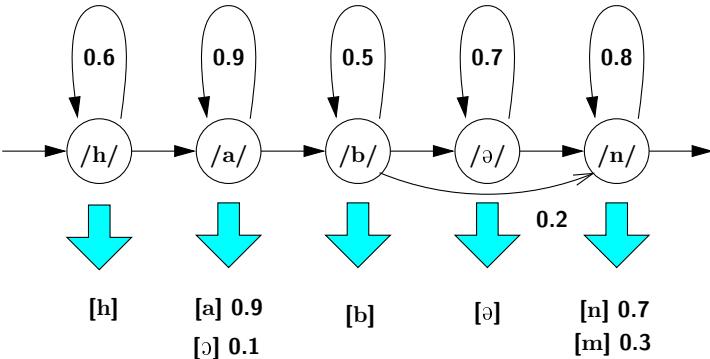
Produktionswahrscheinlichkeiten

Gaußsche Mischverteilungen

Lernen der HMM-Parameter

Robuste Schätzverfahren

Das HMM als Wortaussprachemodell



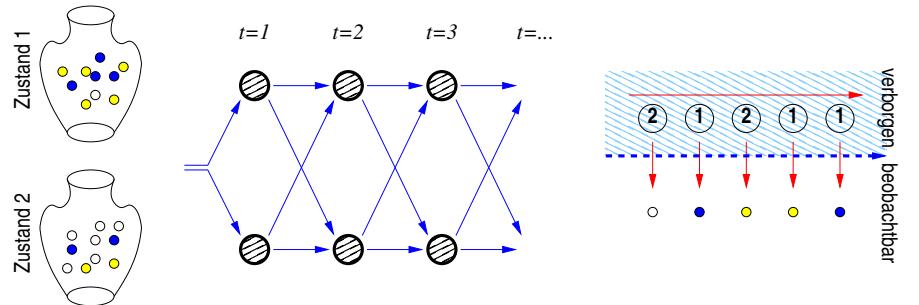
- Modellzustände $\hat{=}$ Artikulationsgesten
 - Zustandswiederholung $\hat{=}$ längere Lautdauer
 - Zustand überspringen $\hat{=}$ Lautereignis elidieren
 - zufallsgesteuerte Ausgabe $\hat{=}$ Ausspracheverschleifung

Was ist eigentlich *verborgen* im Hidden Markov Modell ?

- ... die Folge $q_1, q_2, \dots, q_t, q_{t+1}, \dots$
(die inneren „Systemzustände“)

Und was ist *beobachtbar* im Hidden Markov Modell ?

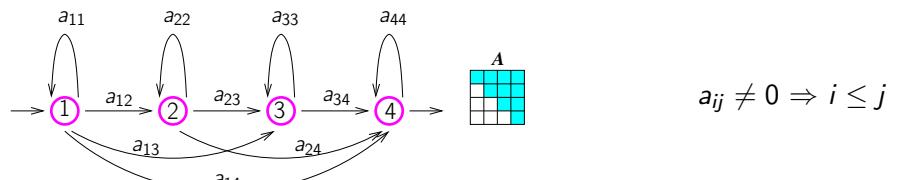
- ... die Folge $o_1, o_2, \dots, o_t, o_{t+1}, \dots$
(die „Ausgabezeichen“ des Zufallsprozesses)



Fakt

Jeder Zustand kann grundsätzlich jedes Zeichen erzeugen !

Verbindungsstruktur einer Markovkette



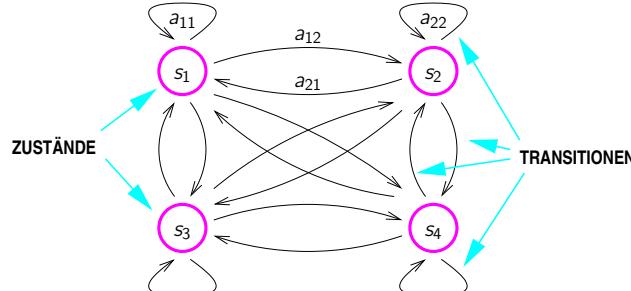
$$a_{ij} \neq 0 \Rightarrow i \leq j$$

$$ij \neq 0 \Rightarrow j-i \in \{0, 1, 2\}$$

The diagram illustrates a linear model with four states, labeled 1 through 4, arranged horizontally. State 1 is the initial state, indicated by an incoming arrow from the left. Transitions between states are labeled a_{ij} , where i is the current state and j is the next state. The transitions are: a_{12} from state 1 to state 2, a_{23} from state 2 to state 3, a_{34} from state 3 to state 4, and a_{41} from state 4 back to state 1. Each transition is represented by a curved arrow pointing from one state circle to the next. To the right of the states is a 4x4 matrix labeled A , representing the transition probabilities or weights for this linear model.

$$a_{ii} \neq 0 \Rightarrow j - i \in \{0, 1\}$$

Markovkette = einfache stationäre Markovquelle

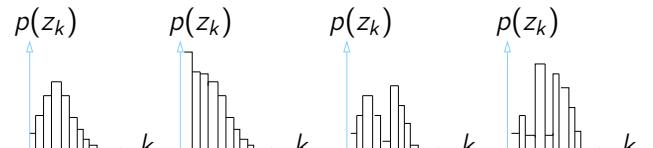


- Endliches Zustandsalphabet $\mathcal{S} = \{s_1, \dots, s_N\}$
 - Diskreter stochastischer Prozess $q_1, q_2, \dots, q_t, \dots \quad q_t \in \mathcal{S}$
 - Erste Markoveigenschaft $P(q_t | q_1, \dots, q_{t-1}) = P(q_t | q_{t-1})$
 - Stationäre Übergangswahrscheinlichkeiten $a_{ij} \stackrel{\text{def}}{=} P(q_t = s_j | q_{t-1} = s_i)$
 - Anfangswahrscheinlichkeiten $\pi_i \stackrel{\text{def}}{=} P(q_1 = s_i)$

\rightsquigarrow Parameter (π, A) $\in \mathbb{R}^N \times \mathbb{R}^{N \times N}$

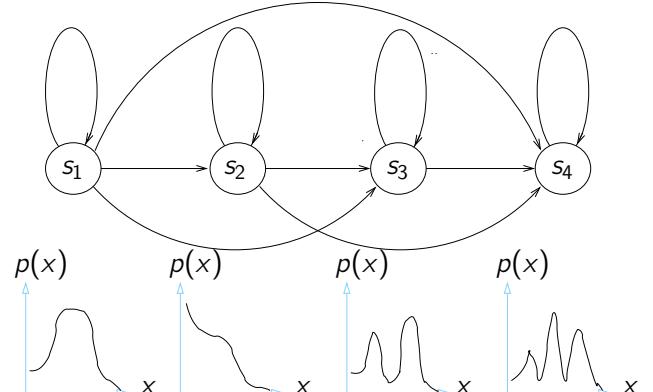
Ausgabeverteilungen eines HMM

diskrete Modellierung



Links-Rechts

kontinuierliche Modellierung



Diskrete Ausgabeverteilungen

- Endliches Zeichenalphabet

$$\mathcal{K} = \{v_1, \dots, v_K\}$$

- Folge beobachteter Ausgabezeichen

$$o_1, o_2, \dots, o_t, \dots \quad o_t \in \mathcal{K}$$

- Zweite Markoveigenschaft

$$P(o_t | q_1, \dots, q_t, o_1, \dots, o_{t-1}) = P(o_t | q_t)$$

- Stationäre Ausgabewahrscheinlichkeiten

$$b_{jk} \stackrel{\text{def}}{=} P(o_t = v_k | q_t = s_j)$$

$$\rightsquigarrow \text{Parameter } (\pi, A, B) \in \mathbb{R}^N \times \mathbb{R}^{N \times N} \times \mathbb{R}^{N \times K}$$

Stochastische Normierungsbedingungen

- Anfangswahrscheinlichkeiten

$$\sum_{i=1}^N \pi_i = 1$$

- Übergangswahrscheinlichkeiten

$$\sum_{j=1}^N a_{ij} = 1, \quad i = 1, \dots, N$$

- Diskrete Ausgabewahrscheinlichkeiten

$$\sum_{k=1}^N b_{jk} = 1, \quad j = 1, \dots, N$$

- Kontinuierliche Ausgabedichtefunktionen

$$\int_{\mathbb{R}^D} b_j(x) dx = 1, \quad j = 1, \dots, N$$

Stetige (kontinuierliche) Ausgabeverteilungen

- Folge beobachteter Ausgabevektoren

$$x_1, x_2, \dots, x_t, \dots \quad x_t \in \mathbb{R}^D$$

- Zweite Markoveigenschaft

$$P(x_t | q_1, \dots, q_t, x_1, \dots, x_{t-1}) = P(x_t | q_t)$$

- Stationäre Ausgabewahrscheinlichkeiten

$$b_j(y) \stackrel{\text{def}}{=} P(\mathbb{X}_t = y | q_t = s_j)$$

$$\rightsquigarrow \text{Parameter } (\pi, A, [b_j]) \in \mathbb{R}^N \times \mathbb{R}^{N \times N} \times (\mathbb{R}^D \rightarrow \mathbb{R})^N$$

Drei offene Fragen zum Thema HMM

?

Berechnung der Datenerzeugungswahrscheinlichkeit

$$P(o|\lambda) = \sum_q P(q, o | \lambda)$$

?

Aufdeckung der wahrscheinlichsten Zustandsfolge

$$P(q, o | \lambda) \xrightarrow{!} \text{MAX}$$

?

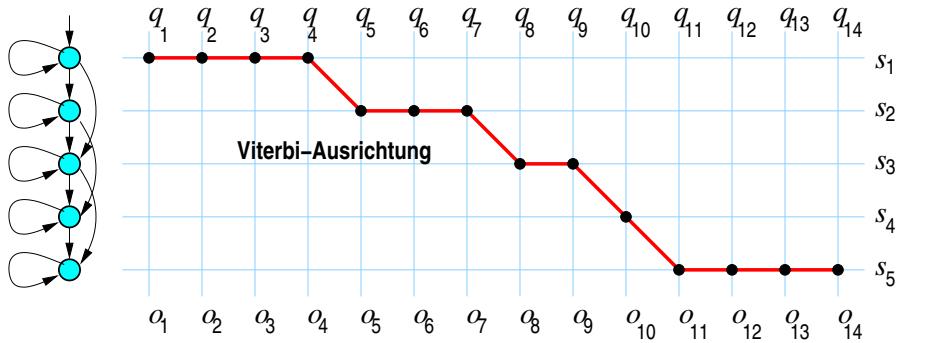
Schätzung der bestpassenden Modellparameter

$$P(o|\hat{\lambda}) = \max_{\lambda} P(o|\lambda)$$

Viterbi-Ausrichtung & Entscheidungsüberwachtes Lernen

Sind die Viterbi-Ausrichtungen der Sequenzen einer Probe bekannt, so lassen sich verbesserte HMM-Parameter als relative aus absoluten Häufigkeiten gewinnen („Viterbi-Training“):

$$\begin{aligned}\hat{a}_{ij} &\propto \#(i \rightarrow j) & \stackrel{\text{def}}{=} & \{t \mid q_{t-1} = s_i, q_t = s_j\} \\ \hat{b}_{jk} &\propto \#(j \downarrow k) & \stackrel{\text{def}}{=} & \{t \mid q_t = s_j, o_t = v_k\}\end{aligned}$$



Motivation

Dynamic Time Warping

Hidden Markov Modell

Produktionswahrscheinlichkeiten

Gaußsche Mischverteilungen

Multivariate Normalverteilungsdichten Identifikation von Mischverteilungsdichten

Lernen der HMM-Parameter

Viterbi-Training — ein Flip-Flop-Algorithmus

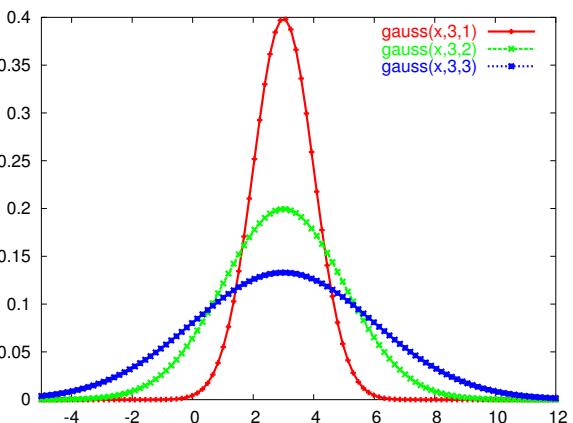
- 1 INITIALISIERUNG** Startkonfiguration $\lambda^0 = (\pi^0, \mathbf{A}^0, \mathbf{B}^0)$
- 2 VITERBI-AUSRICHTUNG(EN)** $\mathbf{q}^* = \text{argmax} P(\mathbf{q}, \mathbf{o} | \lambda^i)$
- 3 PARAMETER-AKTUALISIERUNG** $\lambda^i = (\pi^i, \mathbf{A}^i, \mathbf{B}^i)$
- 4 ABBRUCHBEDINGUNG** ENDE oder $i \mapsto i + 1$ und weiter bei 2

Iterbi-Training ist ein monotoner Gradientenaufstieg

$$\begin{aligned} \text{P}^*(\boldsymbol{o}|\boldsymbol{\lambda}^i) &= \max_q \text{P}(\boldsymbol{o}, \boldsymbol{q} \mid \boldsymbol{\lambda}^i) \\ &= \text{P}(\boldsymbol{o}, \boldsymbol{q}^i \mid \boldsymbol{\lambda}^i) \\ &\leq \text{P}(\boldsymbol{o}, \boldsymbol{q}^i \mid \boldsymbol{\lambda}^{i+1}) \\ &\leq \text{P}(\boldsymbol{o}, \boldsymbol{q}^{i+1} \mid \boldsymbol{\lambda}^{i+1}) \\ &= \text{P}^*(\boldsymbol{o}|\boldsymbol{\lambda}^{i+1}) \end{aligned}$$

Univariate Normalverteilungsdichten

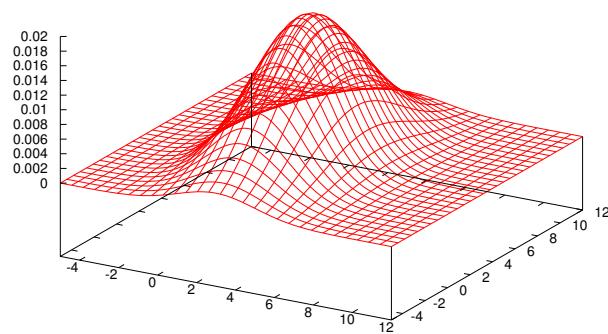
$$\mathcal{N}(x \mid \mu, \sigma^2) \stackrel{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left\{-\frac{1}{2} \cdot \frac{(x - \mu)^2}{\sigma^2}\right\}$$



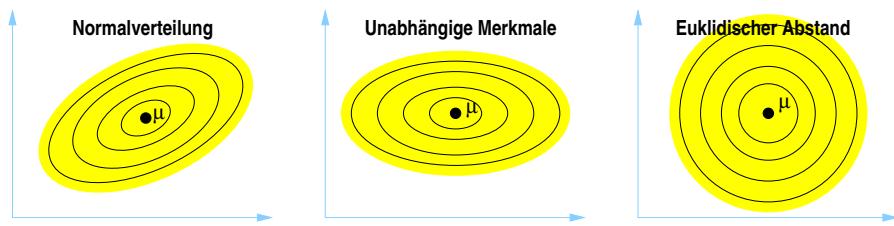
Bivariat unkorrelierte Normalverteilungsdichten

$$\mathcal{N}(x \mid \mu, \sigma) \stackrel{\text{def}}{=} \frac{1}{2\pi\sigma_1\sigma_2} \cdot \exp \left\{ -\frac{1}{2} \cdot \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right) \right\}$$

gauss(x,y,3,2,4) —



Parameterreduzierte Normalverteilungsdichten



**Symmetrisch
positiv-definit**

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1D} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2D} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{D1} & \sigma_{D2} & \dots & \sigma_{DD} \end{pmatrix} \quad \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_D^2 \end{pmatrix}$$

allgemeines
Hyperellipsoid

Diagonalmatrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Trägheitsachsen
parallel zu
Koordinatenachsen

Einheitsmatrix

normierte
Hypersphäre

Multivariate Normalverteilungsdichten

Definition

Ein Zufallsvektor $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_D)^\top$ heißt **multivariat normalverteilt**, falls er der D -dimensionalen Verteilungsdichtefunktion

$$\mathcal{N}(x \mid \mu, S) \stackrel{\text{def}}{=} \frac{1}{\sqrt{\det(2\pi S)}} \cdot \exp \left\{ -\frac{1}{2} \cdot (x - \mu)^\top S^{-1} (x - \mu) \right\}$$

gehorcht. Es ist $\mu \in \mathbb{R}^D$ der **Erwartungswertvektor** der Verteilung; die positiv-definite, symmetrische Matrix $S \in \mathbb{R}^{D \times D}$ heißt **Kovarianzmatrix** der Normalverteilung.

Mischverteilungsdichtefunktionen

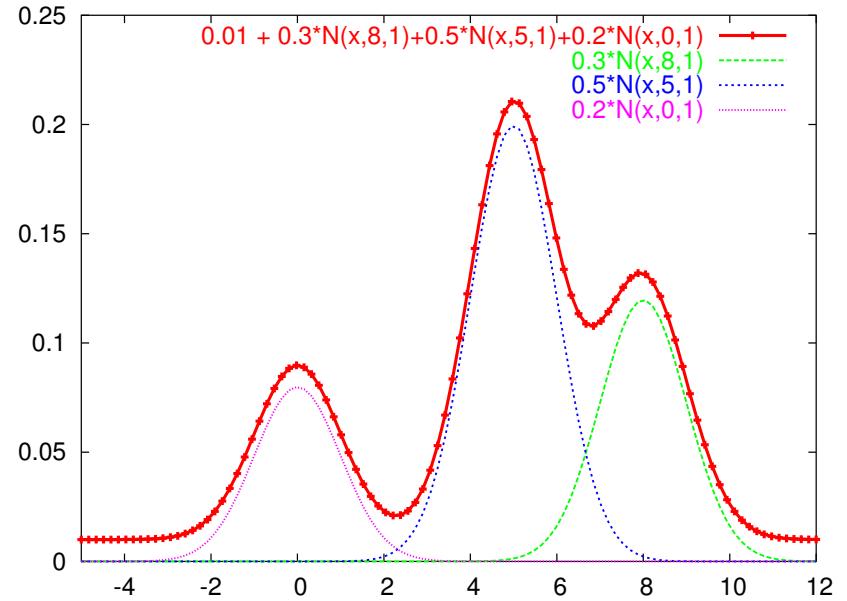
Definition

Ein Zufallsvektor $\mathbb{X} = (\mathbb{X}_1, \dots, \mathbb{X}_D)^\top$ heißt (multivariat normal) **mischverteilt** mit der Ordnung K , falls er einer Verteilungsdichtefunktion

$$f(x) = \sum_{k=1}^K c_k \cdot f_k(x) = \sum_{k=1}^K c_k \cdot \mathcal{N}(x | \mu_k, S_k)$$

mit $c_k \geq 0$ und $\sum c_k = 1$ gehorcht. Die Koeffizienten c_k heißen **Mischungsgewichte**, die Dichtefunktionen $f_k(\cdot)$ heißen **Mischungskomponenten** von $f(\cdot)$.

Mischung von univariaten Normalverteilungsdichten



EM-Algorithmus zur Identifikation gaußscher Mischungen

1 INITIALISIERUNG

Wähle eine geeignete Mischungsordnung $K \in \mathbb{N}$

Wähle Startparameter $(c_k^{(0)}, \mu_k^{(0)}, S_k^{(0)})$, $k = 1..K$; setze $i = 1$

2 ERWARTUNGSWERT-SCHRITT

Bestimme die $T \cdot K$ a posteriori Auswahlwahrscheinlichkeiten

$$\gamma_t^{(i)}(k) \stackrel{\text{def}}{=} P^{(i-1)}(\Omega_k \mid \mathbf{x}_t) = \frac{c_k^{(i-1)} \cdot \mathcal{N}(\mathbf{x}_t \mid \boldsymbol{\mu}_k^{(i-1)}, \boldsymbol{\Sigma}_k^{(i-1)})}{\sum_l c_l^{(i-1)} \cdot \mathcal{N}(\mathbf{x}_t \mid \boldsymbol{\mu}_l^{(i-1)}, \boldsymbol{\Sigma}_l^{(i-1)})}$$

3 MAXIMIERUNGS-SCHRITT

Berechne neue Parameter mit maximaler Kullback-Leibler-Statistik

$$c_k^{(i)} = \sum \gamma_t^{(i)}(k) / T$$

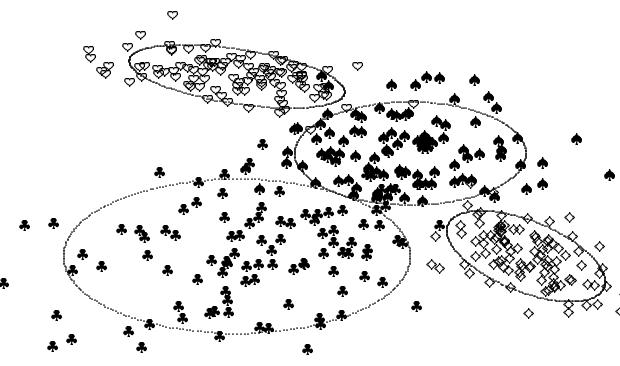
$$\mu_k^{(i)} = \sum \gamma_t^{(i)}(k) \cdot x_t \Big/ \sum \gamma_t^{(i)}(k)$$

$$\boldsymbol{s}_k^{(i)} = \overline{\sum \gamma_t^{(i)}(\boldsymbol{k}) \cdot \boldsymbol{x}_t \boldsymbol{x}_t^\top} / \sum \gamma_t^{(i)}(\boldsymbol{k}) - (\boldsymbol{\mu}_k^{(i)}) (\boldsymbol{\mu}_k^{(i)})^\top$$

4 TERMINIERUNG

Weiter mit $i \leftarrow i + 1$ oder gehe \approx ENDE

Identifikation von Mischverteilungen



Problem

Angenommen, obige Daten sind gemäß $\sum_{k=1}^K c_k f_k(\mathbf{x})$ mischverteilt. Wie lauten die **bestpassenden** Parameter (Maximum-Likelihood) ?

$$K \in \mathbb{N}, \quad (c_1, \mu_1, S_1), (c_2, \mu_2, S_2), \dots, (c_K, \mu_K, S_K)$$

Motivation

Dynamic Time Warping

Hidden Markov Modell

Produktionswahrscheinlichkeiten

nen der HMM-Parameter
Baum-Welch-Algorithmus · DDHMM, CDHMM, GMHMM,
GHMM

EM-Prinzip & Baum-Welch-Trainingsalgorithmus

Definition

Für ein HMM mit Parametern λ (bzw. $\hat{\lambda}$) und eine Lernsequenz $o \in \mathcal{K}^T$ bezeichne

$$\ell_{\text{ML}}(\lambda) \stackrel{\text{def}}{=} \log P(o|\lambda) = \log \sum_{q \in \mathcal{S}^T} P(o, q | \lambda)$$

die logarithmierte Likelihood-Zielgröße und

$$Q(\lambda, \hat{\lambda}) \stackrel{\text{def}}{=} \mathcal{E}[\log P(o, q \mid \hat{\lambda}) \mid o, \lambda]$$

die Kullback-Leibler-Statistik.

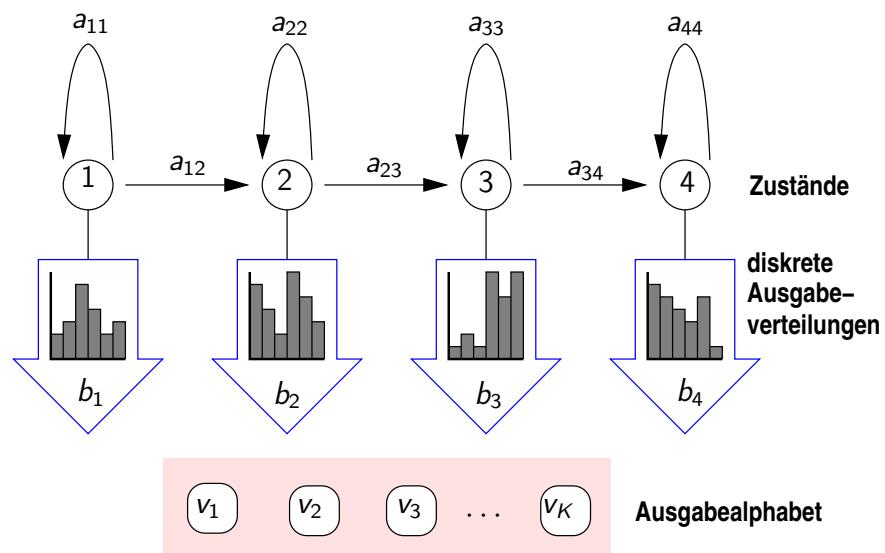
Satz (Expectation-Maximization-Prinzip)

Für alle HMM-Parameterfelder λ , $\hat{\lambda}$ gilt

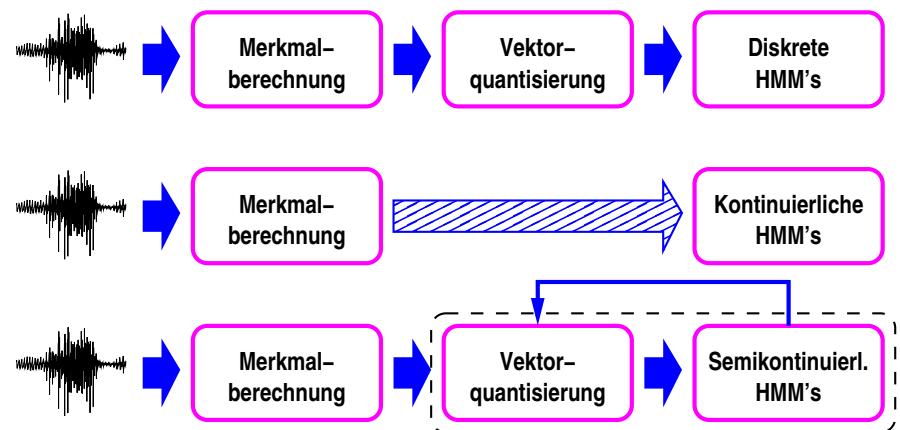
$$Q(\lambda, \hat{\lambda}) \geq Q(\lambda, \lambda) \quad \Rightarrow \quad \ell_{\text{ML}}(\hat{\lambda}) \geq \ell_{\text{ML}}(\lambda)$$

mit Gleichheit nur an stationären Stellen λ von $\ell_{\text{ML}}(\cdot)$.

HMMs mit diskreten Ausgabeverteilungen



Architekturen von HMM-Spracherkennungssystemen



diskret · normalverteilt · mischverteilt · semikontinuierlich

Baum-Welch-Algorithmus für diskrete Ausgabeverteilungen

- ## 1 INITIALISIEREN WEITERSCHALTEN ABBRUCH TESTEN

2 EXPECTATION

A posteriori Übergangswahrscheinlichkeiten für $s_i \rightarrow s_i$ in t

$$\xi_t(i, j) \stackrel{\text{def}}{=} P(q_t = i, q_{t+1} = j \mid \boldsymbol{o}, \boldsymbol{\lambda}) = \frac{\alpha_t(i) \cdot a_{ij} \cdot b_j(o_{t+1}) \cdot \beta_{t+1}(j)}{\sum_{i=1}^N \alpha_t(i) \cdot \beta_t(i)}$$

A posteriori Zustandswahrscheinlichkeiten für s_i in t

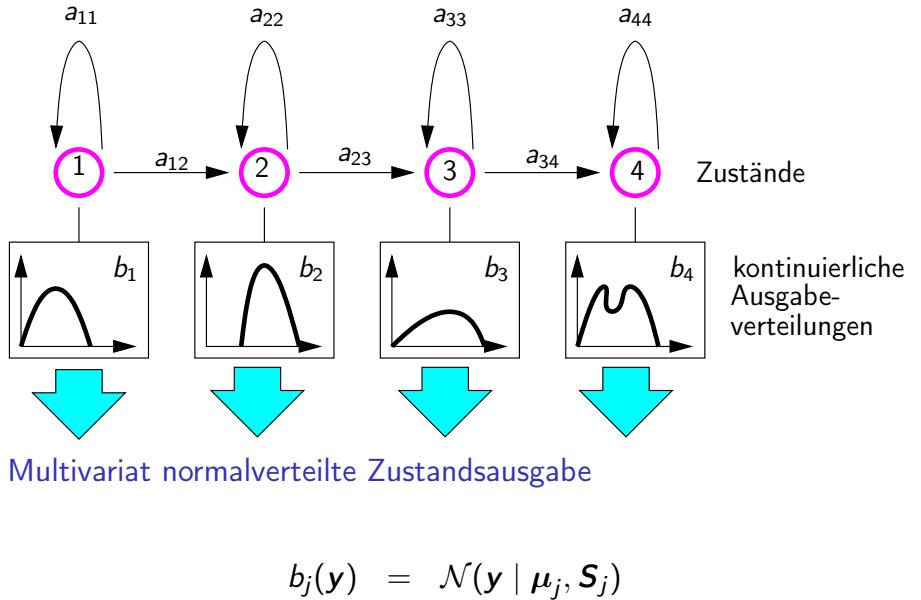
$$\gamma_t(i) \stackrel{\text{def}}{=} P(q_t = i \mid \boldsymbol{o}, \boldsymbol{\lambda}) = \frac{\alpha_t(i) \cdot \beta_t(j)}{\sum_{j=1}^N \alpha_t(j) \cdot \beta_t(j)}$$

- ## 3 MAXIMIZATION

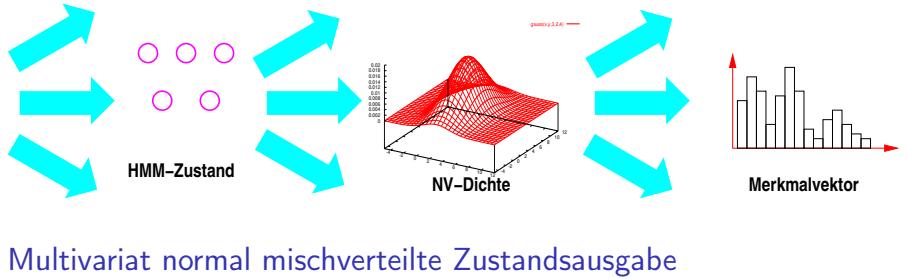
Neuberechnung der $Q(\lambda, \hat{\lambda})$ -optimalen Parameter

$$\hat{\pi}_i = \gamma_1(i), \quad \hat{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, \quad \hat{b}_{jk} = \frac{\sum_{t=1}^T \mathbf{1}_{o_t=v_k} \cdot \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)}$$

HMMs mit stetigen Ausgabeverteilungen



HMMs mit Mischverteilungen



$$b_j(\mathbf{y}) = \sum_{m=1}^{M(j)} c_{jm} \cdot \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_{jm}, \mathbf{S}_{jm})$$

Bemerkung

NM bzw. $\sum_j M(j)$ Mischungskoeffizienten

NMD bzw. $NMD^2/2$ Normalverteilungsparameter

Baum-Welch-Algorithmus für normalverteilte Ausgaben

- ## 1 INITIALISIEREN, WEITERSCHALTEN, ABBRUCH TESTEN

- ## 2 EXPECTATION

A posteriori Zustandswahrscheinlichkeiten und Übergangswahrscheinlichkeiten

$$\gamma_t(i) \ , \quad \xi_t(i,j) \ , \qquad t = 1..T, i = 1..N, j = 1..N$$

- ## 3 MAXIMIZATION

Neuberechnung der $Q(\lambda, \hat{\lambda})$ -optimalen Parameter $\{\hat{\pi}_i\}$, $\{\hat{a}_{ij}\}$ und

$$\hat{\mu}_j = \frac{\sum_{t=1}^T \gamma_t(j) \cdot \mathbf{x}_t}{\sum_{t=1}^T \gamma_t(j)}, \quad \hat{\mathbf{s}}_j = \frac{\sum_{t=1}^T \gamma_t(j) \cdot (\mathbf{x}_t - \hat{\mu}_j)(\mathbf{x}_t - \hat{\mu}_j)^\top}{\sum_{t=1}^T \gamma_t(j)}$$

Baum-Welch-Algorithmus für mischverteilte Ausgaben

- 1 INITIALISIEREN, WEITERSCHALTEN, ABBRUCH TESTEN

- ## 2 EXPECTATION

A posteriori Zustandswahrscheinlichkeiten $\gamma_t(i)$, Übergangswahrscheinlichkeiten $\xi_t(i,j)$ sowie Selektionswahrscheinlichkeiten

$$\zeta_t(j, m) = \text{P}(q_t = j, k_t = m \mid \boldsymbol{X}, \boldsymbol{\lambda}) = \gamma_t(j) \cdot c_{jm} \mathcal{N}_{jm}(\boldsymbol{x}_t) \left/ \sum_{l=1}^{M(j)} c_{jl} \mathcal{N}_{jl}(\boldsymbol{x}_t) \right.$$

- ## 3 MAXIMIZATION

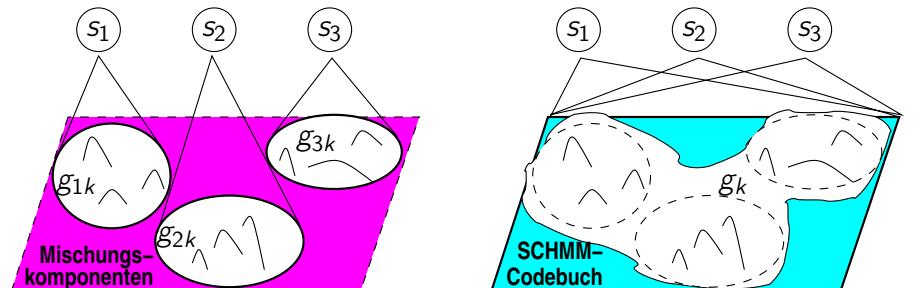
Neuberechnung der $Q(\lambda, \hat{\lambda})$ -optimalen Parameter $\{\hat{\pi}_i\}$, $\{\hat{a}_{ij}\}$ und

$$\hat{c}_{jm} = \sum_{t=1}^T \zeta_t(j, m) \Bigg/ \sum_{m=1}^{M(j)} \sum_{t=1}^T \zeta_t(j, m) = \sum_{t=1}^T \zeta_t(j, m) \Bigg/ \sum_{t=1}^T \gamma_t(j)$$

$$\lambda_{jm} = \sum_{t=1}^T \zeta_t(j, m) \cdot x_t \Bigg/ \sum_{t=1}^T \zeta_t(j, m)$$

$$\hat{\mathbf{s}}_{jm} = \sum_{t=1}^T \zeta_t(j, m) \cdot \mathbf{x}_t \mathbf{x}_t^\top \Bigg/ \sum_{t=1}^T \zeta_t(j, m) - \hat{\mu}_{jm} \hat{\mu}_{jm}^\top$$

Semikontinuierliche HMMs



Multivariat normalverteilte gemeinsame Dichten ('tied mixtures')

$$b_j(\mathbf{y}) = \sum_{k=1}^K c_{jk} \cdot \mathcal{N}(\mathbf{y} \mid \boldsymbol{\mu}_k, \mathbf{S}_k)$$

Bemerkung

NK Mischungskoeffizienten

KD bzw. $KD^2/2$ Normalverteilungsparameter

Motivation

Dynamic Time Warping

Hidden Markov Modell

Produktionswahrscheinlichkeiten

Gaußsche Mischverteilungen

Lernen der HMM-Parameter

Robuste Schätzverfahren

Mehrfachheit · Verklebung · Interpolation · Dauer

Baum-Welch-Algorithmus für semikontinuierliche Ausgaben

- ## 1 INITIALISIEREN, WEITERSCHALTEN, ABBRUCH TESTEN

- ## 2 EXPECTATION

A posteriori Übergangswahrscheinlichkeiten $\gamma_t(i)$, Zustandswahrscheinlichkeiten $\xi_t(i, j)$ sowie Selektionswahrscheinlichkeiten

$$t(j, k) = \text{P}(q_t = j, k_t = k \mid \boldsymbol{X}, \boldsymbol{\lambda}) = \gamma_t(j) \cdot c_{jk} \mathcal{N}_k(\boldsymbol{x}_t) \Bigg/ \sum_{l=1}^K c_{lk} \mathcal{N}_k(\boldsymbol{x}_t)$$

- ## 3 MAXIMIZATION

Neuberechnung der $Q(\lambda, \hat{\lambda})$ -optimalen Parameter $\{\hat{\pi}_i\}$, $\{\hat{a}_{ij}\}$ und

$$\begin{aligned}\hat{c}_{jk} &= \sum_{t=1}^T \zeta_t(j, k) \Bigg/ \sum_{t=1}^T \gamma_t(j) \\ \hat{\mu}_k &= \sum_{j=1}^N \sum_{t=1}^T \zeta_t(j, k) \cdot \mathbf{x}_t \Bigg/ \sum_{j=1}^N \sum_{t=1}^T \zeta_t(j, k) \\ \hat{\mathbf{s}}_k &= \sum_{j=1}^N \sum_{t=1}^T \zeta_t(j, k) \cdot \mathbf{x}_t \mathbf{x}_t^\top \Bigg/ \sum_{j=1}^N \sum_{t=1}^T \zeta_t(j, k) - \hat{\mu}_k \hat{\mu}_k^\top\end{aligned}$$

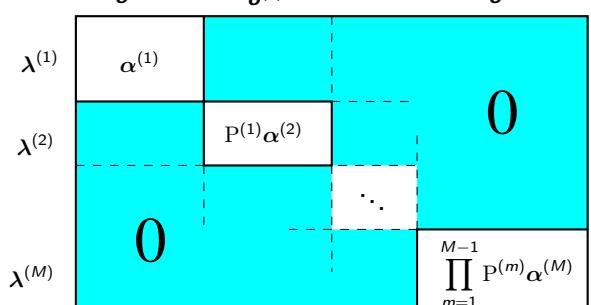
Robuste Parameterschätzung ?

zu viele $\left\{ \begin{array}{l} \text{Wortmodelle} \\ \text{HMM-Zustände} \\ \text{freie Parameter} \end{array} \right\}$ \Leftrightarrow zu wenige $\left\{ \begin{array}{l} \text{Äußerungen} \\ \text{Wortvorkommen} \\ \text{Lautereignisse} \end{array} \right\}$

Problematik des Parameterlernens aus Daten:

- zu große **Varianz** der geschätzten Parameterwerte
 - starke **Zerklüftung** der Zielfunktion $\ell_{\text{ML}}(\lambda)$
 - systematisches **Verschwinden** der Statistiken $\gamma_t(i)$, $\xi_t(i,j)$, $\zeta_t(j,k)$
 - **nullwertige** Parameter \hat{a}_{ij} , \hat{b}_{jk} , $\hat{\mathbf{S}}_k$ etc.
 - Nullwertigkeit ist **reproduzierend** !

Mehrfache Modelle — mehrfache Probemuster

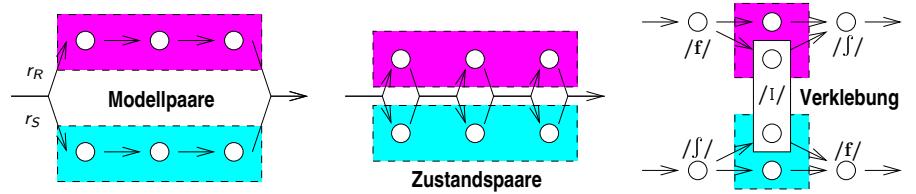


Modifizierte Schätzformel (exemplarisch):

$$\hat{a}_{ij} = \frac{\sum_{\ell=1}^L \sum_{m=1}^{M_\ell} \left(\sum_{t=1}^{T_{\ell,m}-1} \xi_t^{(\ell,m)}(i,j) \right)}{\sum_{\ell=1}^L \sum_{m=1}^{M_\ell} \left(\sum_{t=1}^{T_{\ell,m}-1} \gamma_t^{(\ell,m)}(i) \right)}$$

Strukturinterpolation

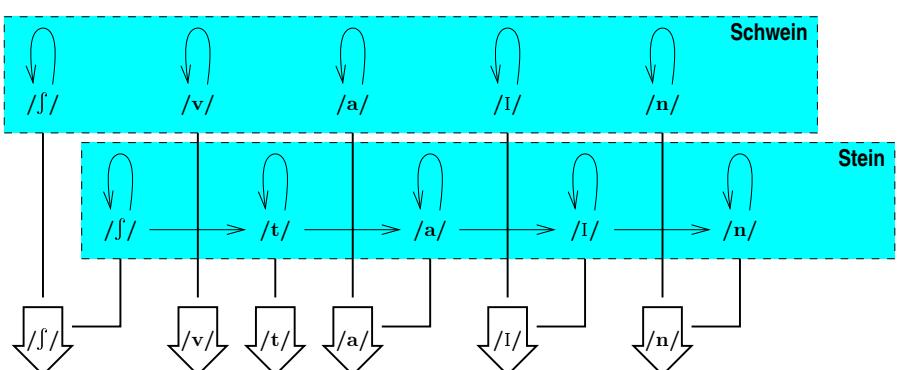
- Interpolation zweier HMMs
- Interpolation zweier Zustände



$$P(o | \{\lambda_\ell\}, \{r_\ell\}) = \sum_{\ell=1}^{\ell_{\max}} r_\ell \cdot P(o | \lambda_\ell), \quad \sum_{\ell=1}^{\ell_{\max}} r_\ell = 1$$

Parameterverklebung (Gleichschaltung, 'tying')

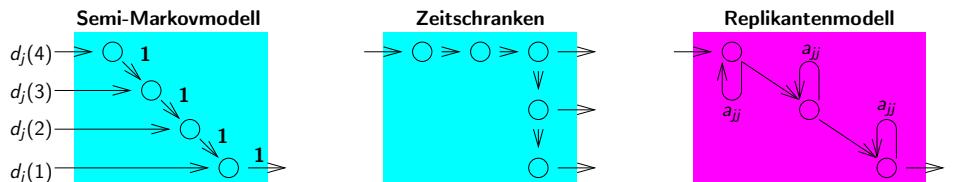
- Paarweise Identifikation von Verteilungsparametern
- ... erzwingt fort dauernde Wertegleichheit
- ... reduziert Anzahl der Freiheitsgrade des Modells
- ... realisiert via gemeinsam genutzter ('pooled') Statistiken



Zustandsdauerverteilung im HMM

HMMs sind lausig schlechte Dauermodelle !

$$d_i(\tau) \stackrel{\text{def}}{=} P(\text{"noch genau } (\tau-1)\text{-mal in } s_i \text{ bleiben"} \mid q_{t-\tau} = s_i, \lambda) \\ = a_{ii}^{\tau-1} \cdot (1 - a_{ii})$$

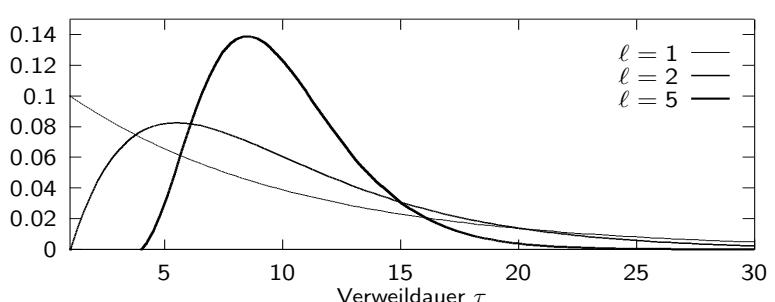


- **Semi-HMM** — explizite Dauerverteilung innerhalb $[1 : \ell]$
- **Min-Max-HMM** — Dauergleichverteilung innerhalb $[\ell_0 : \ell]$
- **Replikanten-HMM** — implizite Dauerverteilung innerhalb $[\ell : \infty)$

Replikantenmodelle

Für die Dauerverteilung eines ℓ -fachen Zustandes
(Original zzgl. $\ell - 1$ Kopien)
gilt die Faltungsdarstellung

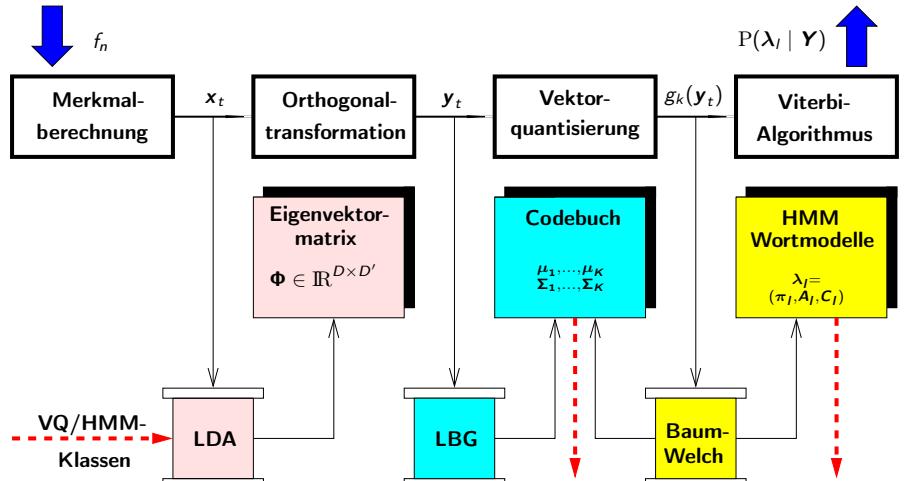
$$d_{i,\ell}(\tau) = \binom{\tau-1}{\ell-1} \cdot a_{ii}^{\tau-\ell} \cdot (1-a_{ii})^\ell$$



An Stelle einer Zusammenfassung

EXEMPLARISCHE BERECHNUNGSFOLGE ZUM HMM-TRAINING

Cepstrum · LDA · VQ · SCHMM



Dynamic Time Warping

Gaußsche Mischverteilungen

Lernen der HMM-Parameter

Robuste Schätzverfahren

Beispielaufbau

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Teil VI

Wortmodelle

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 21. Dezember 2018

Motivation	Ganzwort-HMM oooooooo	Wortuntereinheiten ooooooo	CD-PLUS oooooo	Subphone oooo	Wortgrenzen oooo	Neue Wörter ooooo	A-Varianten ooo	Σ
------------	--------------------------	-------------------------------	-------------------	------------------	---------------------	----------------------	--------------------	----------

Motivation

Wortbezogene Hidden Markov Modelle

Modellierungseinheiten unterhalb der Wortebene

Kontextabhängige Phone

Subphonemische Modellierung

Modellierung phonetischer Effekte an den Wortgrenzen

Ad hoc Modellierung unbekannter Wörter

Modellierung von Ausspracheverarianten

Motivation	Ganzwort-HMM oooooooo	Wortuntereinheiten ooooooo	CD-PLUS oooooo	Subphone oooo	Wortgrenzen oooo	Neue Wörter ooooo	A-Varianten ooo	Σ
------------	--------------------------	-------------------------------	-------------------	------------------	---------------------	----------------------	--------------------	----------

HMMs als Wortmodelle

Bayesregel zur Wortkettenerkennung

$$w^* = \underset{w \in \mathcal{V}^*}{\operatorname{argmax}} P(w|X) = \underset{w \in \mathcal{V}^*}{\operatorname{argmax}} \frac{\overbrace{P(w)}^{LSM} \cdot \overbrace{P(X|w)}^{ASM}}{P(X)}$$

GEGEBEN: eine Wortfolge

$$w = w_1 \dots w_m$$

GESUCHT: ein HMM $\lambda(w)$ mit

$$P(X | \lambda(w)) \approx P(X|w)$$

Analysis-by-Synthesis

sequentielle Verkettung von Wortmodellen

$$\lambda(w) = \lambda(w_1) \circ \lambda(w_2) \circ \lambda(w_3) \circ \dots \circ \lambda(w_m)$$

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooo oooooooo ooooooo oooo oooo oooo ooo

Motivation

Wortbezogene Hidden Markov Modelle

HMM-Struktur · Initialisierung · Stichprobe · Lernen

Modellierungseinheiten unterhalb der Wortebene

Kontextabhängige Phone

Subphonemische Modellierung

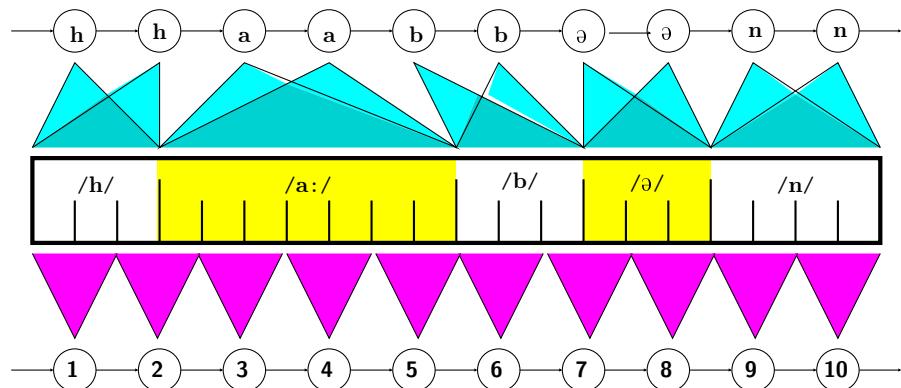
Modellierung phonetischer Effekte an den Wortgrenzen

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooooooooo oooooooo ooooooo oooo oooo oooo ooo

Datengetriebene Vorbesetzung der HMM-Parameter



Wortmodell $\lambda(/haben/$ mit $N = 10$ Zuständen

- explizite phonetische Segmentierung
- lineare Zeitverzerrung (Daten vs. HMM)

(oben)
(unten)

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ●oooooooooooo oooooooo ooooooo oooo oooo oooo ooo

Erzeugung von Ganzwortmodellen

Konfiguration

- Links-Rechts-HMM (linear, Bakis)
- $\lambda(w_\ell)$ besitzt N_ℓ Zustände
z.B. $N_\ell = 5$ (Ziffernwörter) oder $N_\ell \propto$ „Anzahl der Phoneme in W_ℓ “

Initialisierung

- π_i, a_{ij} unkritisch; b_{jk}, c_{jm} gleichverteilt
- SCHMM-Parameter μ_k, S_k via LBG/EM
- CD-HMM und GM-HMM $\mu_{jm}, S_{jm} = ?$

Parametertraining

- Baum-Welch-Iteration auf etikettierter Sprachdatensammlung

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooooooooo oooooooo ooooooo oooo oooo oooo ooo

Aufbau einer Sprachdatensammlung

- **Einzelsprechersysteme**
 ≥ 10 , besser 50 oder 100 Aussprachebeispiele / Wortform
- **Sprecherunabhängige Systeme**
ausgewogene Population von ≥ 100 SprecherInnen
Geschlecht · Anatomie · Dialekt · Ideolekt · Soziolekt
- **Wohldefinierte & kontrollierte Sprachqualität**
Mikrofon/Telefon · Bandbreite · Störfaktoren · Dynamik
- **Äußerungseinheiten**
Einzelwörter · Kommandos/Sätze · Dialogturns · \geq Szenarien
- **Akquisition**
Spontansprachdaten · Textreproduktionen · „Wizard of Oz“
- **Etikettierung**
(deskriptiv/normativ)
Text · Sätze · Wörter · Phoneme · Allophone
Intonation · Pausen · nonverbale Phänomene · Überlappung

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○●○○○ ○○○○○○ ○○○ ○○○ ○○○ ○○○

Die Erlanger Bahnauskunft-Stichprobe

1. ich muesste morgen nach Falkenberg fahren , so _ dass ich ab zwei Uhr ankomme.
2. wann kommt der spaeteste ICE , der uebermorgen in Sonnenhofen abfahrt, in Nordhausen an ?
3. hat der Zug um acht Uhr in Bad Kleinen Anschluss nach Weida ?
4. um wieviel Uhr geht der spaeteste Intercity-Zug nach Schaffhausen Faehre ?
5. faehrt der erste Zug um fuenf Uhr sechs nach Heiligenstadt auch werktags ?
6. wann faehrt morgen der fruehesten Zug nach Kitzingen ?
7. guten Morgen, gibt es einen Intercity , der an einem Wochentag nachmittags direkt nach Flensburg faehrt?
8. wir muessen nach Neustadt fahren , und zwar uebermorgen .
9. ich haette gerne einen Intercity-Zug nach Eisenhuettenstadt .
10. heute Nacht musste ich mit dem Zug von Pasewalk nach Neuhaus fahren .
11. ich suche einen Zug von Kempten nach Stralsund mit Ankunftszeit ab sechs Uhr .
12. wann muss ich abfahren , damit ich moeglichst frueh in Leipzig Bayerischer Bahnhof bin ?
13. ich will fragen, wann uebermorgen Nachmittag ein Zug von Andernach nach Rosenheim geht ?
14. wann kann ich am naechsten Wochenende mittags nach Wasserbillig fahren ?
15. wann kommt der fruehesten IC , der um halb fuenf in Bad Bentheim abfahrt, in Berlin-Schoeneweide an ?
16. ich will morgen in Kassel-Wilhelmshoehe sein .
17. ist es moeglich, an einem Wochentag von Emden ueber Rendsburg nach Berlin zu fahren ?
18. hat der Zug um zehn Uhr in Dortmund Anschluss nach Worms ?
19. ich moechte direkt mit dem Intercity-Zug an Pfingsten nach Immenstadt fahren .
20. damit ich um viertel vor sieben in Hanau ankomme , wann muss ich in Neustadt losfahren ?
21. ich wollte fragen, ob man morgen auch nach Biberach fahren kann ?
22. welche Moeglichkeiten gibt es , zwischen fuenf und ein Uhr nach Rotenburg zu kommen ?
23. wuerden Sie mir am neunundzwanzigsten neunten die kuerzeste Verbindung nach Allensbach angeben ?
24. gibt es einen Zug , der uebermorgen nach Guestrow faehrt?
25. gibt es eine Moeglichkeit, spaetestens heute von Amstetten nach Dillenburg zu kommen ?
26. gibt es eine stuednliche Direktverbindung zwischen Dortmund und Forchheim ?
27. guten Morgen, faehrt am Heiligabend ein Intercity nach Gehlberg ?
28. wir moechten zwischen vier und fuenf Uhr in Celle sein .
29. gibt es eine IC-Verbindung am Mittwoch zwischen Schleswig und Fulda ?
30. faehrt der naechste Zug um ein Uhr von Mittenwald nach Bad Brambach auch in zwei Tagen ?
31. wir moechten einen Zug von Traunstein nach Goettingen
32. faehrt am kommenden Wochenende ein ICE nach Oberstdorf ?
33. was ist die spaeteste Moeglichkeit, um zwölf Uhr ueber Gelsenkirchen nach Saarburg zu fahren ?
34. gibt es einen Zug , der am Montag in Weimar ist ?
35. gibt es einen ICE , der nach viertel vor sechs von Hamburg-Altona nach Flensburg faehrt?
36. ich moechte wissen, ob heute um drei Uhr auch ein IC nach Buende faehrt ?
37.

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○●○ ○○○○○○ ○○○ ○○○ ○○○ ○○○

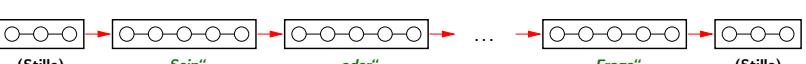
Wörter sind keine geeigneten Modellierungseinheiten !

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○●○○○ ○○○○○○ ○○○ ○○○ ○○○ ○○○

Eingeckettetes Lernen

• Einzelwortprobe

Wortrealisierungen liegen in Sprechpausen eingebettet vor



• Verbundwortprobe

Wortrealisierungen liegen in komplette Sätze eingebettet vor
Satzrealisierungen liegen in Sprechpausen eingebettet vor

• Diskontinuierliche Verbundwortprobe

Wortrealisierungen sind u.U. durch Stillebereiche unterbrochen
Stillebereiche sind *nicht* Bestandteil der Etikettierung

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○○●○ ○○○○○○ ○○○ ○○○ ○○○ ○○○

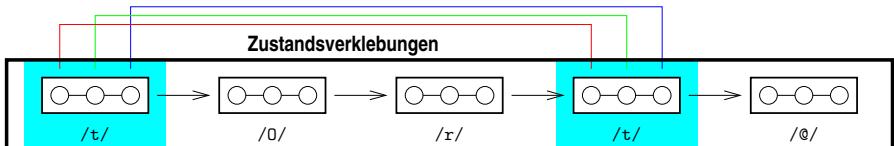
Analyse durch Synthese

Wort-HMMs werden aus Phonem-HMMs verkettet.

Beispiel

Das HMM für das Wort „Torte“ besitzt die Struktur

$$\lambda(/t0rt@/) = \lambda(/t/) \circ \lambda(/0/) \circ \lambda(/r/) \circ \lambda(/t/) \circ \lambda(/@/)$$

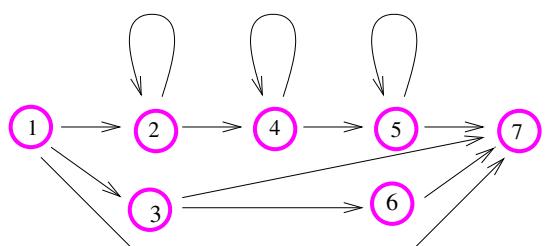


Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooo ● ooooooo ooooooo oooo ooooo ooo

HMM-Topologien für Einzellaute

- Linear- oder Bakis-HMM mit $N = 3$ Zuständen
- Spezialmodelle mit $N > 3$ für Diphthonge & Affrikate
- Kai-Fu Lees „Dreimaster“ mit $N = 7$ Zuständen:

$$A = \begin{pmatrix} 0 & a_{12} & a_{13} & 0 & 0 & 0 & a_{17} \\ 0 & a_{22} & 0 & a_{24} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{36} & a_{37} \\ 0 & 0 & 0 & a_{44} & a_{45} & 0 & 0 \\ 0 & 0 & 0 & 0 & a_{55} & 0 & a_{57} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$



Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooo ● ooooooo ooooooo oooo ooooo ooo

Gütekriterien

Wortuntereinheiten zur akustischen Modellierung

- **Präzision**
die WUE ist hochspezialisiert und folglich **trennscharf**
- **Robustheit**
großer Trainingsmaterialvorrat & wirksame Glättungsmaßnahmen
~~ **gute Schätzwerte**
- **Modularität**
fixes **Inventar moderaten Umfangs** für alle potentiellen Sprechakte
- **Transfer**
Synthese neuer Wortmodelle aus vorhandenen WUE nach orthografischer/phonematischer Umschrift
- **maschinelle Segmentierbarkeit ?**
heutzutage irrelevant wg. Analysis-by-Synthesis Strategie

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooo ooooooo ooooooo oooo ooooo ooo

Motivation

Wortbezogene Hidden Markov Modelle

Modellierungseinheiten unterhalb der Wortebene

Entwurfskriterien · Phonologische Einheiten · Akustische Einheiten

Kontextabhängige Phone

Subphonematische Modellierung

Modellierung phonetischer Effekte an den Wortgrenzen

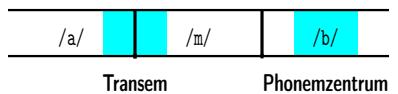
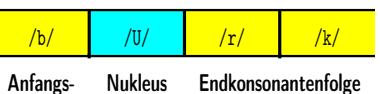
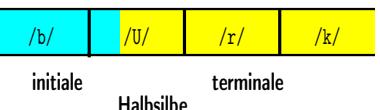
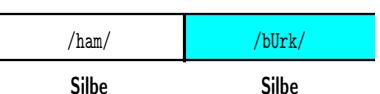
Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachenvarianten

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ooooooo ● ooooooo ooooooo oooo ooooo ooo

Phonologisch orientierte Wortuntereinheiten I

am Beispiel „Hamburg“

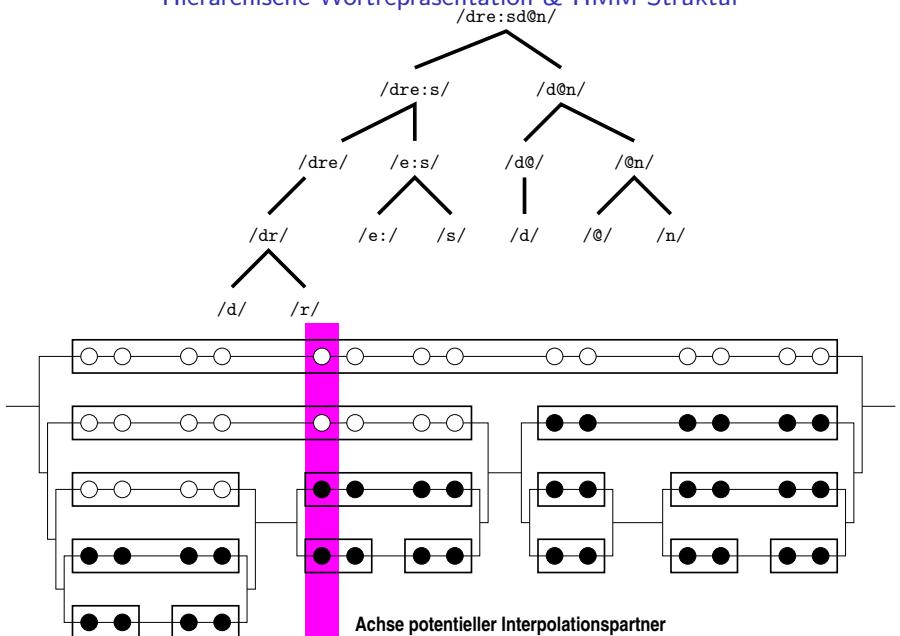


Phonologisch orientierte Wortuntereinheiten II

- **Phone** +modular, –präzis
je nach Differenzierungsgrad 40–200 universelle Einheiten
- **Phoneme** +modular, –präzis
je nach Sprache 20–60
- **Silben** +präzis, –modular
20 000 (engl.), 100 (japan.); Koartikulation primär innerhalb
- **Halbsilben** +trennscharf, ±modular
800/2560 initiale/terminale im Deutschen
- **Sylparts** guter Kompromiß
47 AKF, 20 Nuklei, 159 EKF im Deutschen
- **Diphone** besser: Transeme
1000–1500 Einheiten (engl./ital.), ungünstige Nahtstellen
- **Doppelhalbsilbe** ++trennscharf, --modular, --robust
Silbenkern–Silbenkern, 2 Mill. im Deutschen

CFU — „context-freezing units“

Hierarchische Wortrepräsentation & HMM-Struktur



HMM für längerdauernde Wortuntereinheiten

PRO

Aussprachevariabilität wird in ihrer lautlichen Umgebung eingefroren

KONTRA

verminderte Robustheit & mangelhafte Modularität

LÖSUNG

- hierarchische Zerlegung der Wortaussprache
- hierarchische Modellstruktur durch parallel verdrahtete HMMs
- HMMs ausschließlich für häufig auftretende Spracheinheiten
- Interpolation konkurrierender HMMs
- ⇒ 'context-freezing units' (CFU) sind Kompromiß zwischen stabil & unspezifisch versus labil & trennscharf

Motivation	Ganzwort-HMM	Wortuntereinheiten	CD-PLUS	Subphone	Wortgrenzen	Neue Wörter	A-Varianten	Σ
oooooooooo	oooo●ooo	ooooooo	oooooo	oooo	ooooo	ooooo	ooo	

Akustisch (statt phonetisch) orientierte Wortuntereinheiten

(Algorithmus)

- 1 **BOOTSTRAP-ERKENNER**
Aufbau eines initialen ASE-Systems
- 2 **FENONISCHE GRUNDFORM**
Analyse eines oder mehrerer Aussprachebeispiele je Wortform

$$W_\ell \xrightarrow{\quad} \mathfrak{F}(W_\ell) = f_1 \dots f_m$$

- 3 **WORTMODELL-KONFIGURATION**

$$\lambda(W_\ell) \stackrel{\text{def}}{=} \lambda(f_1) \circ \lambda(f_2) \circ \lambda(f_3) \circ \dots \circ \lambda(f_m)$$

- 4 **PARAMETER INITIALISIEREN**
- 5 **BAUM-WELCH TRAININGSALGORITHMUS**

(zum dritten Teil)

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooooooo ooooooo ● oooooo oooo oooo oooo ooo

Erzeugen einer fenonischen Wortumschrift

1. Methode:

Verschmelzung identischer sukzessiver VQ-Indizes

$$x_1 \dots x_T \rightsquigarrow o_1 \dots o_T \rightsquigarrow o_{t(1)} \dots o_{t(m)} \stackrel{\text{def}}{=} f_1 \dots f_m$$

2. Methode:

VQ-Gewinnerzellen der Zustände eines LR Ganzwort-HMM

$$f_j = \underset{k \in \mathcal{K}}{\operatorname{argmax}} P(o_t = k \mid q_t = j, \lambda(W)) = \underset{k \in \mathcal{K}}{\operatorname{argmax}} b_{jk}^{(W)}$$

3. Methode:

Bestparkettierung der Wortaussprache(n) mit Basis-HMMs

$$\prod_{m=1}^M P(\mathbf{X}_m \mid \lambda(\mathfrak{F})) = \prod_{m=1}^M P(\mathbf{X}_m \mid \lambda(f_1) \circ \dots \circ \lambda(f_m)) \stackrel{!}{\rightarrow} \text{MAX}$$

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooooooo ooooooo ● oooooo oooo oooo oooo ooo

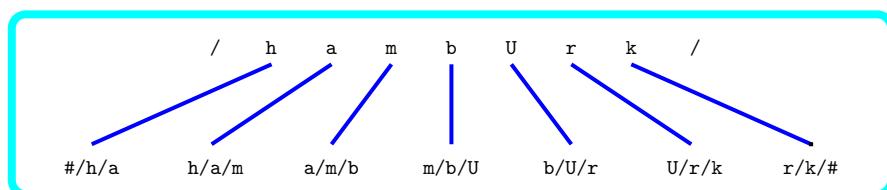
Phoneme im Kontext (Allophone)

- Segmentelle Basisspracheinheit kurzer Dauer (\rightsquigarrow Phonem)
- Modellierung durch HMM ist kontextabhängig

Beispiel

Das Phonem *r* im Wort Hamburg *hambUrk*,
als **Triphon** oder rechtes/*linkes Biphon* oder **Monophon**:

$$r \rightarrow U/r/k, \quad r \rightarrow /r/k, \quad r \rightarrow U/r/, \quad r \rightarrow /r/$$



Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooooooo ooooooo ● oooooo oooo oooo oooo ooo

Motivation

Wortbezogene Hidden Markov Modelle

Modellierungseinheiten unterhalb der Wortebene

Kontextabhängige Phone

Allophone · Triphone · Generalisierung · Polyphone

Subphonemische Modellierung

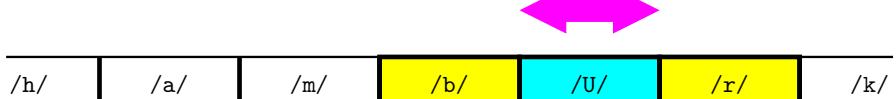
Modellierung phonetischer Effekte an den Wortgrenzen

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooooooo ooooooo ● oooooo oooo oooo oooo ooo

Trainieren von Triphon-HMMs materielles Korrelat



kontextueller Einzugsbereich

(Algorithmus)

- 1 Erfassung der phonematischen Wortumschrift
- 2 Konfiguration gewöhnlicher *Monophonmodelle*
Initialisierung & Optimierung der Parameter
- 3 Konfiguration linker/rechter *Biphonmodelle*
Initialisierung & Optimierung der Parameter
- 4 Konfiguration der *Triphonmodelle*
Initialisierung & Optimierung der Parameter
- 5 Synthese der Erkennungswortschatzmodelle (Rückgriff/Interpolation)



(zumdinogIA)

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○○○○○ ○○○○○○ ○○○ ○○○○ ○○○ ○○○

Verallgemeinerte Triphone

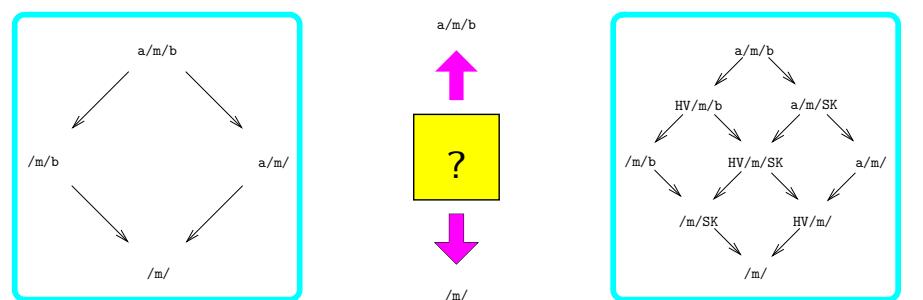
Problem

Seltene Triphone \rightsquigarrow labile HMM-Parameter

Fragmentierung der Trainingsdaten

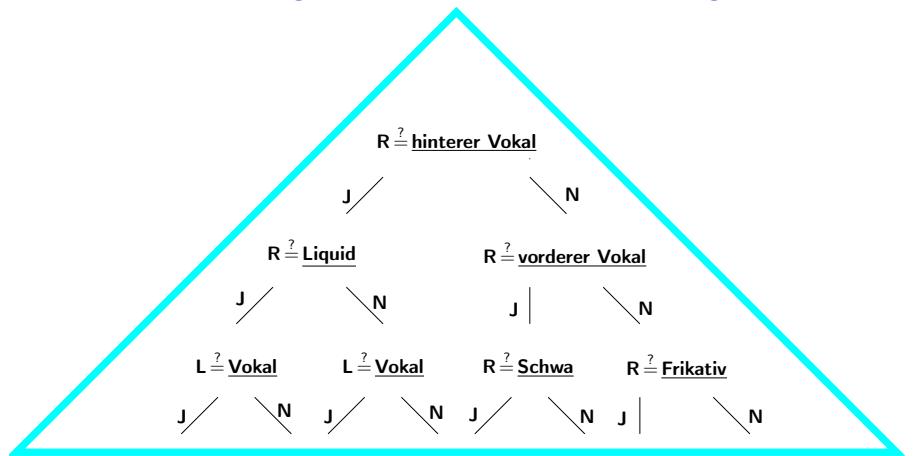
Lösung

Bündeln geeigneter Gruppen **kerngleicher** Triphone mit verwandten akustischen Eigenschaften



Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○○○○○ ○○○○○○ ○○○ ○○○○ ○○○ ○○○

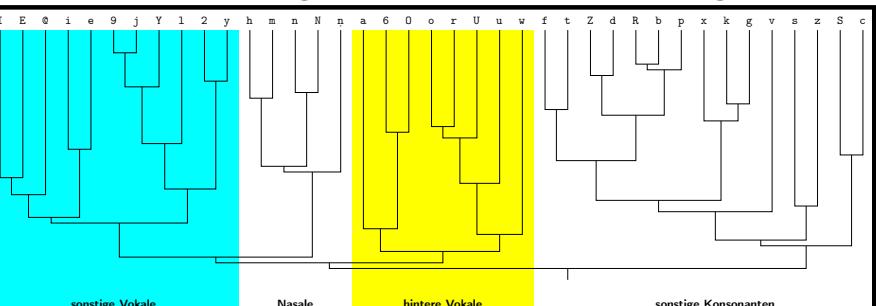
Datengetriebene Generalisierung



- Induktive Erzeugung eines binären Entscheidungsbaumes
- Fragen** = diskriminative Phonemkontakte
- Resultat auch für „ungesehene“ Kontexte

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○○○○○ ○○○○○○ ○○○ ○○○○ ○○○ ○○○

Umschriftgetriebene Generalisierung



- Phonemmodellierung $f_j(\mathbf{x}) = \mathcal{N}(\mathbf{x} | \mu_j, \mathbf{S}_j)$
- Ähnlichkeit als Transformation $\text{sim}_{ij} = \mathcal{I}(f_i, f_j)$
- Agglomerative Gruppierung \rightsquigarrow Phonetisches Dendrogramm
- Phonetisches Oberklassensystem ($C = 4$):**

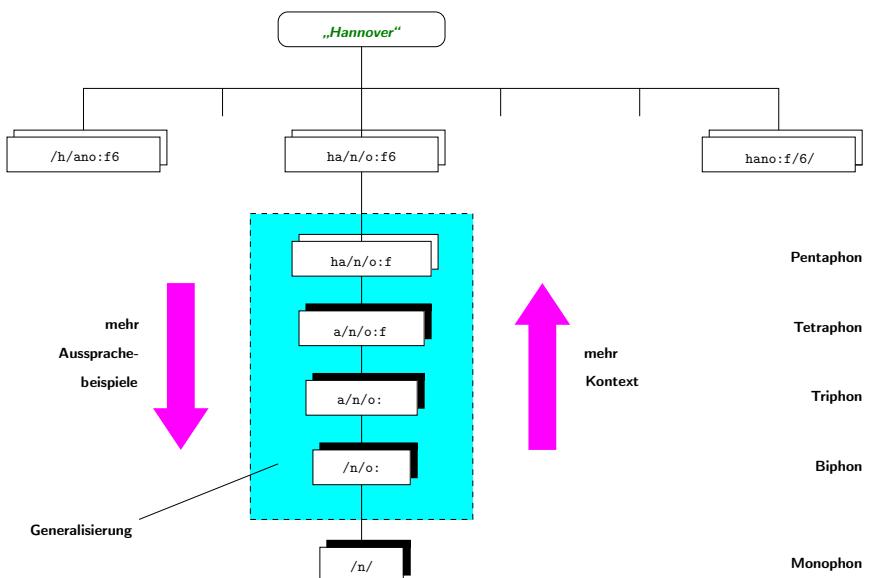
NA Nasale	HV hintere Vokale	SK sonstige Konsonanten	SV sonstige Vokale
--------------	----------------------	----------------------------	-----------------------

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 ○○○○○○○○ ○○○○○○ ○○○ ○○○○ ○○○ ○○○

Polyphone — Phoneme in beliebig breitem Kontext

- häufige Spracheinheiten sind modellierungs**fähig** robuste Schätzwerte
- häufige Spracheinheiten sind modellierungs**bedürftig** Redundanz & Verschleifung
- Phoneme in **beliebig breitem** rechten/linken Kontext indirekte Koartikulationseffekte
- Vergrößerung durch **balanciertes Abschälen** skalierte Kontextabhängigkeit
- Inkorporation von **Akzentzeichen** Modellierung betonter & unbetonter Silben
- Inkorporation von **Grenzmarkierungen** Phrasengrenze · Wortgrenze · Morphemgrenze · Silbengrenze

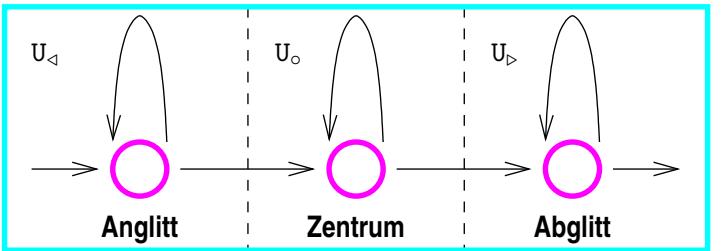
Polyphone — Beispielwort „*Hannover*“



Subphonemische Modellstruktur

Jedes **Phonem** gliedert sich segmentell in drei **Semiphone**:

1. **Anglitt** — abhangig vom Vorgngerphonem
 2. **Zentrum** — phonkontextunabhangig
 3. **Abglitt** — abhangig vom Nachfolgerphonem



Motivation

Wortbezogene Hidden Markov Modelle

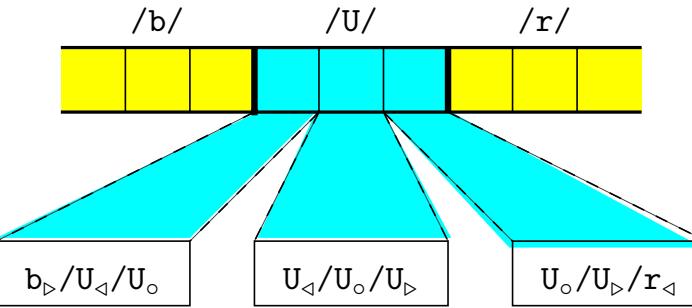
Kontextabhängige Phone

Subphonemische Modellierung

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

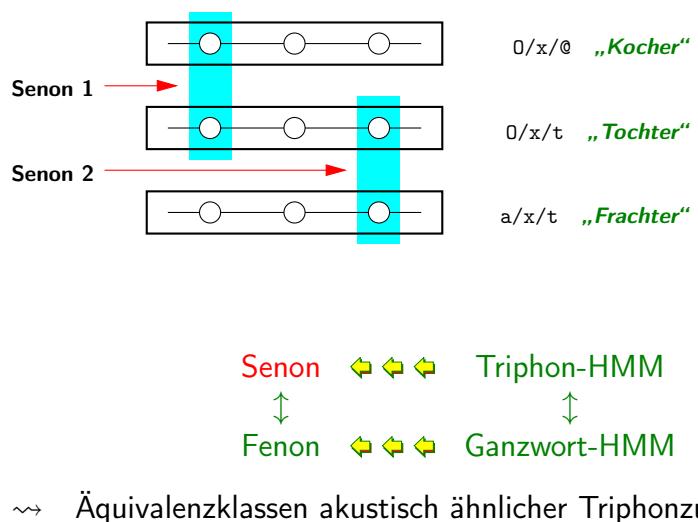
Semiphondarstellung



- für Semiphone werden einfache Links/Rechtskontexte postuliert
 - resultierende Spracheinheiten:
 - unabhängige **Phonemzentren**
 - viele kombinatorische **Transemhälften**
 - einige unabhängige **Anglitte & Abglitte**

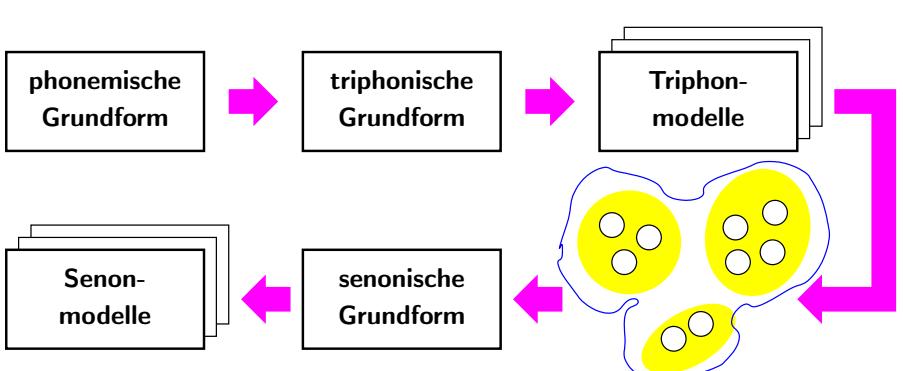
Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooo oooooooo ooooo ooo● oooo ooooo oooo

Senonische Spracheinheiten



Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone Wortgrenzen Neue Wörter A-Varianten Σ
 oooooooo oooooooo ooooo ooo● oooo ooooo oooo

Lernen senonischer Grundformen



Separate Clusteranalyse aller Triphonzustände eines **Kernphonems**

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone **Wortgrenzen** Neue Wörter A-Varianten Σ
 oooooooo oooooooo ooooo oooo oooo ooooo oooo

Motivation

Wortbezogene Hidden Markov Modelle

Modellierungseinheiten unterhalb der Wortebene

Kontextabhängige Phone

Subphonematische Modellierung

Modellierung phonetischer Effekte an den Wortgrenzen

Wortübergreifende Verschleifung · Lernphase · Erkennungsphase

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

Motivation Ganzwort-HMM Wortuntereinheiten CD-PLUS Subphone **Wortgrenzen** Neue Wörter A-Varianten Σ
 oooooooo oooooooo ooooo oooo ●ooo ooooo oooo

Verschleifungseffekte über die Wortfugen hinweg

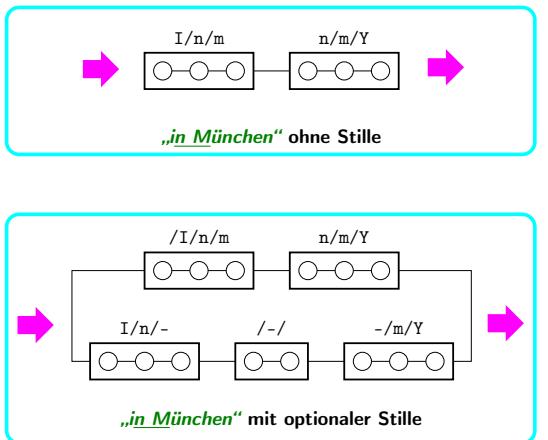
Wortübergreifende Koartikulation
 „in München“

/In/ + /mYnc@n/ /ImYnc@n/

Verstümmelung unbetonter Funktionswörter
 „Roß und Reiter“

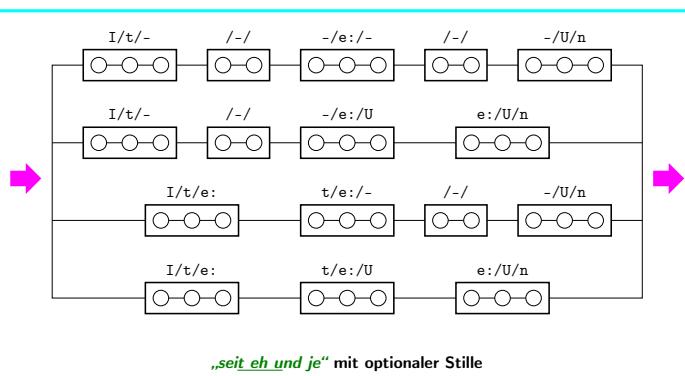
/r0s/ + /Unt/ + /raIt6/ /r0snraIt6/

Wortgrenzenübergreifende Triphone in der Lernphase



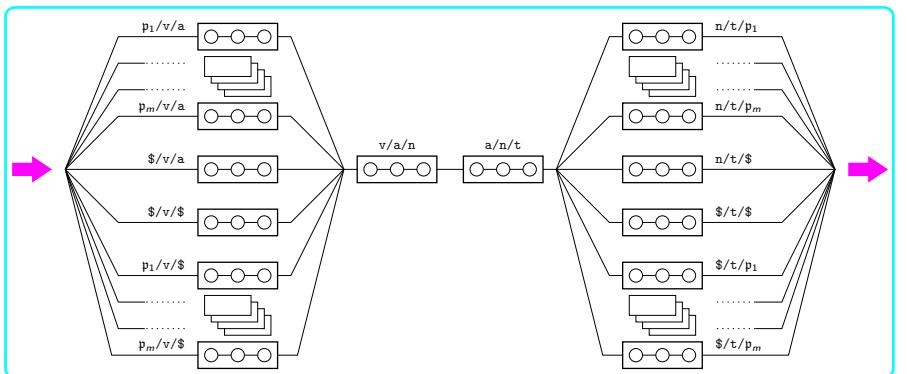
Koartikulation überbrückt keine Stillebereiche \Rightarrow $\left\{ \begin{array}{l} \text{mit } \left\{ \begin{array}{l} \text{ohne Stille} \\ \text{annotierter} \\ \text{latenter} \end{array} \right\} \text{ Stille} \end{array} \right\}$

Wortgrenzenübergreifende Triphone in der Lernphase



Kombinatorische Verwicklung bei einphonemigen Wörtern

Wortgrenzenübergreifende Triphone in der Erkennungsphase



Ungebremste Kombinatorik der kontextbedingt verästelten initialen/finalen Triphonmodelle des HMMs für /vant/ („Wand“)

#/v/a v/a/n a/n/t n/t/#

Wortbezogene Hidden Markov Modelle

Ad hoc Modellierung unbekannter Wörter

Nichtwörter · Detektion · Modellierung · Wiedererkennung

Modellierung von Aussprachevarianten

Performanzlücke der ASE bei Spontansprache

Unbekannte Wörter

- Wörter außerhalb des Erkennungswortschatzes

Außerlexikalische Einheiten

- Ungefüllte Pausen („Häsitationen“)
- Gefüllte Pausen („äh“, „mmh“)

Nichtverbale Realisierungen

- Räuspern, Husten, Lachen
- Atemgeräusche, Schmatzlaute

Nichtartikulatorische Störproduktionen

- Türenschlagen, Rascheln, Klopfen, ...

→ Detektion → ad hoc Modelle → Wiedererkennung

Modelle für „neue“ Wörter I

Phonematische Modellsynthese

- ↳ eine Phonemumschrift liegt i.a. nicht vor

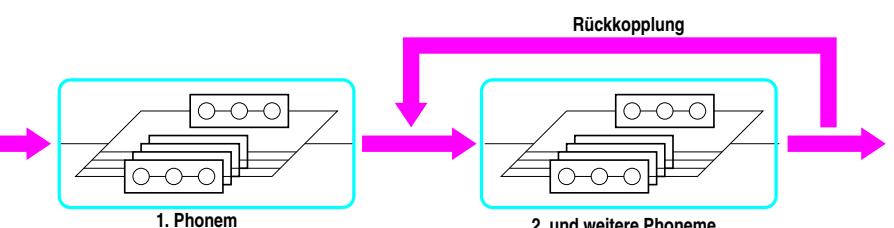
Akustische Modellsynthese

- **Viterbi-Transkription**
Bauplan = a posteriori wahrscheinlichste Phonemfolge
- **Fenon-Transkription**
Bauplan = a posteriori wahrscheinlichste Fenonfolge
- **Senon-Transkription**
Bauplan = wahrscheinlichste Senonfolge eines intermediären Ganzwort-HMMs

OOV-Detektion mit Rückweisungsmodellen

Füllmuster-HMM λ_\emptyset mit diffuser Wahrscheinlichkeitsverteilung:

$$\begin{aligned} P(\mathbf{X} | \lambda_\emptyset) &\leq P(\mathbf{X} | \lambda(w)) && \text{für das korrekte Wort } w \text{ und} \\ P(\mathbf{X} | \lambda_\emptyset) &\geq P(\mathbf{X} | \lambda(v)) && \text{für alle anderen Wörter } v \neq w \end{aligned}$$



- **SYNTHEZISIEREN:**
 λ_\emptyset als repetitive Verkettung von Lautalternativeblöcken
- **TRAINIEREN:**
Ganzwort-HMM λ_\emptyset , trainiert mit allen Sprachproben

Modelle für „neue“ Wörter II

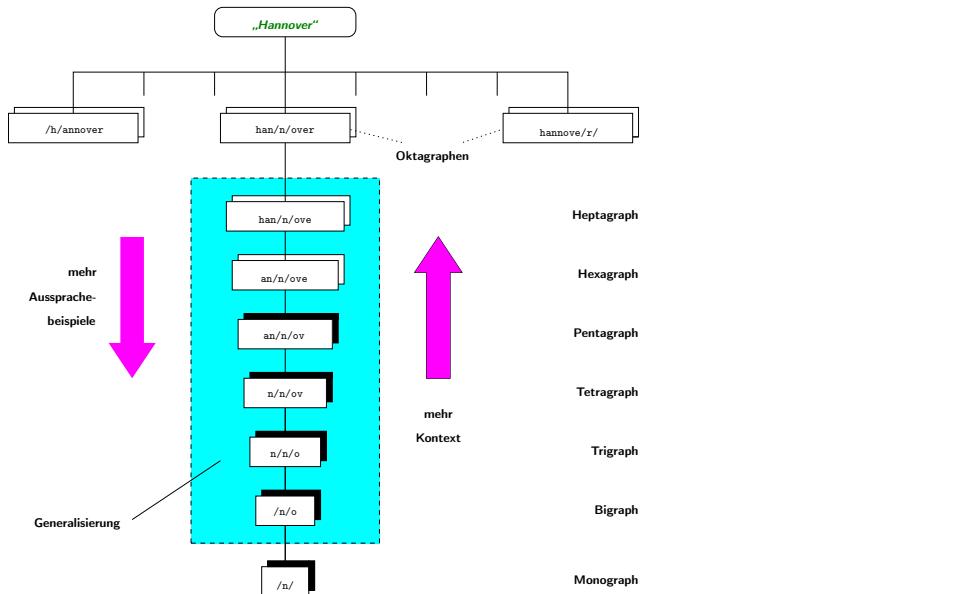
Buchstabiermodus

- ↳ „Hamburg“
- ≈ „ha–ah–em–beh–uh–er–geh“
- ≈ „Heinrich“–„Anton“–„Martha“–„Berta“–„Ulrich“–„Rudolf“–„Gustav“

Graphematische Modellsynthese

- **Regelbasiert**
Erzeugung einer phonemischen Wortumschrift aus der Orthographie
 - ↳ Beispiel NetTalk:
Künstliches Neuronales Netz (KNN) mit $\left\{ \begin{array}{l} 7 \times 29 \text{ Eingabe-,} \\ 80 \text{ Zwischen- und} \\ 26 \text{ Ausgabeneuronen} \end{array} \right\}$
- **Polygraphen**
Basis-HMMs für „kontextabhängige Buchstaben“

Polygraphen — kontextabhängige Buchstaben



Phonematische Standardumschrift

- Wörterbucheintrag der Form „*haben*“ \rightsquigarrow /ha:bən/
 - ein Wort · eine Umschrift · ein Wortmodell

→ *ungenaues Modell & streuende Parameter*

Ausgewählte Ausprachevarianten

- zum Beispiel „*zwei*“ \rightsquigarrow 1. /tsval/ und 2. /tsvo:/
 - ein Wort · mehrere Umschriften · konkurrierende Wortmodelle

☞ *hoher Dekodieraufwand & Datenfragmentierung*

Motivation

Wortbezogene Hidden Markov Modelle

Kontextabhängige Phone

Subphonemische Modellierung

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

Standardumschrift · Varianten · Expansion · Graphen

Maschinelle Erzeugung alternativer Ausspracheumschriften

Expansion durch phonetische Verschleifungsregeln

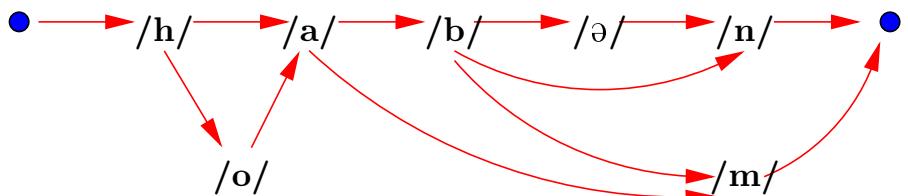
- Nichtdeterminiertes Textersetzungssystem:
 - EINGABE = Standardumschrift
 - REGELN = Assimilation, Elision, schwache Formen
 - AUSGABE = (große) Menge von Aussprachevarianten
 - *kombinatorische Explosion bei langen Wörtern*
Gefahr der Übergeneralisierung
Akquisition des Regelinventars ?!

Beispiel

Für das Wort „*haben*“ ergeben sich die Varianten

/ha:bən/ ; /ha:bn/ ; /ha:bm/ ; /ha:m/ ; /həam/ ;

Phonetische Wortrepräsentation durch Aussprachegraphen



Alle Aussprachevarianten eines Wortes werden in einen zyklusfreien, gerichteten Graphen eingebettet.

- **Paßfähigkeit eines Aussprachegraphen**

- $\mathcal{V} = \mathcal{P}$ exakte Ausschöpfung (hohe Knotenzahl)
- $\mathcal{V} \supset \mathcal{P}$ Übergeneralisierung (fehlerhafte Annahme)
- $\mathcal{V} \subset \mathcal{P}$ Überspezialisierung (fehlerhafte Ablehnung)

- **Variantenwahrscheinlichkeiten** \leftarrow EM-Algorithmus

- | | |
|--|-----------------------|
| eine Wahrscheinlichkeit je Variante | $(\sim L \cdot 100)$ |
| eine Wahrscheinlichkeit je Graphkante | $(\sim L \cdot 10)$ |
| eine Wahrscheinlichkeit je Verschleifungsregel | $(\sim 1 \cdot 1000)$ |

An Stelle einer Zusammenfassung

EXEMPLARISCHE BERECHNUNGSFOLGE ZUM WORTMODELLAUFBAU

Senon-gestützter HMM-Worterkenner

- 1 **Anlegen einer Sprachdatensammlung**
Entwurf — Aufnahme — Diskretisierung
- 2 **Erstellung eines Aussprachelexikons**
Phonemische Umschriften aller Wörter der Lernstichprobe
- 3 **Merkmalberechnung & Vektorquantisierung**
... für die gesamte Lernstichprobe; siehe (3), (4)
- 4 **Lernen der Monophon-HMM's**
Uniforme Initialisierung, Baum-Welch-Training
- 5 **Lernen der Triphon-HMM's**
Initialisierung mit den Monophon-HMM's, Baum-Welch-Training
- 6 **Clustern der Triphon-HMM-Zustände in Senonklassen**
Partitionieren der Mischungskoeffizientenvektoren mit LBG
- 7 **Lernen der Senon-HMM's**
Initialisierung mit den Triphon-HMM's, Baum-Welch-Training
- 8 **Rotationsmatrix aus Senon-LDA etc. ...**
Alle Zeitscheiben werden senonisch klassifiziert \rightsquigarrow LDA

Motivation

Wortbezogene Hidden Markov Modelle

Modellierungseinheiten unterhalb der Wortebene

Kontextabhängige Phone

Subphonemische Modellierung

Modellierung phonetischer Effekte an den Wortgrenzen

Ad hoc Modellierung unbekannter Wörter

Modellierung von Aussprachevarianten

Beispielaufbau

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil VII

Satzmodelle

Motivation Sprachmodelle n-Gramme Parameterglättung Generierung Wortkategorien Wortgruppierung Langzeitkontext

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Motivation Sprachmodelle n-Gramme Parameterglättung Generierung Wortkategorien Wortgruppierung Langzeitkontext

Modellierung der Wortfolgenwahrscheinlichkeit

Bayesregel zur Wortkettenerkennung

$$w^* = \underset{w \in \mathcal{V}^*}{\operatorname{argmax}} P(w|X) = \underset{w \in \mathcal{V}^*}{\operatorname{argmax}} \frac{\overbrace{P(w)}^{LSM} \cdot \overbrace{P(X|w)}^{ASM}}{P(X)}$$

GEGEBEN: ein Textkorpus

$$\mathcal{U} = \{u_1, \dots, u_M\} \subset \mathcal{V}^*$$

GESUCHT: eine Satzwahrscheinlichkeitsverteilung

$$P_{LM} : \begin{cases} \mathcal{V}^* & \rightarrow [0, 1] \\ w = w_1 \dots w_T & \mapsto P_{LM}(w) \end{cases}$$

Motivation

Grammatikkomponenten in der Spracherkennung

Fehlerkorrektur · Phänomene · Formalismus · A/L-Interaktion

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

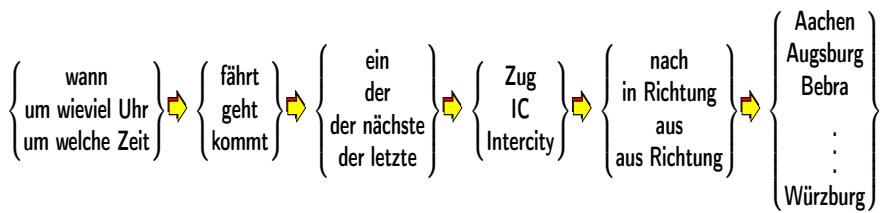
Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Das „Sentence Switchboard“



Beispiel

- Anfragesprache mit $L = 174$ verschiedenen Wörtern
 - 20 allgemeinsprachliche Wörter
 - zzgl. $L_{IC} = 154$ InterCity-Haltepunkte
- **ohne Grammatik:** $174^6 \approx 27.8 \cdot 10^{12}$ Wortketten der Länge 6
- **mit Grammatik:** $3 \cdot 3 \cdot 4 \cdot 3 \cdot 4 \cdot L_{IC} = 66.528$ Wortkombinationen
- ⇒ fast 9 Größenordnungen Unterschied !!

Typische Fehlentscheidungen bei der automatischen Spracherkennung

- (1) „den nächsten Zug“ → „den nächste Zug“
- (2) „um sechs in Bonn“ → „um sechs den Bonn“
- (3) „von Essen nach Kiel“ → „von ist in nach Kiel“
- (4) „nach Köln fahren“ → „nach Ulm fahren“

Phänomene

1. Kongruenzfehler
2. Vertauschung, Auslassung oder Einfügung von Funktionswörtern
3. Parkettierung von Inhaltswörtern durch Funktionswörter
4. grammatisch unauffällige Vertauschungen (gefährlich!)

Sprachverständnis und Sprachverwendung

Sprachkompetenz

“[...] das sprachliche Vermögen eines idealen Sprecher-Hörers, der in einer völlig homogenen Sprachgemeinschaft lebt, seine Sprache ausgezeichnet kennt und bei der Anwendung seiner Sprachkenntnis in der aktuellen Rede von solchen grammatisch irrelevanten Bedingungen wie

- begrenztes Gedächtnis, Zerstreutheit und Verwirrung
- Verschiebung in der Aufmerksamkeit und im Interesse
- Fehler (zufällige oder typische) nicht affiziert wird.”

Chomsky 1973, A.d.SynTh

Sprachperformanz

“[...] spontansprachliche Äußerungen einer auskunftssuchenden oder diktierenden Person [...] im Spannungsfeld von Sparsamkeit und Verschwendungen.”

Schachtl/Block 1991

Spontansprachliche Phänomene im Dialog I

Satzeinleitende Wörter und Floskeln

1. „also“, „ja“, ...
2. „und zwar“, „sozusagen“, ...

Kongruenzfehler und Neuansätze

3. „Ich wollte mit den nächsten Zug nach Konstanz.“
4. „Ihre Augen sind, äh, was machen Sie heute abend?“

Anakoluthe

5. „Das riecht ja scheußlich riecht das ja!“

Spontansprachliche Phänomene im Dialog II

Anaphern und Ellipsen

6. „Ist der neu?“
7. „Nein, [...] mit Perwoll gewaschen.“

Formstufen sprachlicher Ausprägung

(Bsp. Besitzanzeige)

8. „die Kraniche des Ibikus“ (genitivisch)
9. „der BMW vom Chef“ (präpositional)
10. „dem Montague seiner Theorie ihre Schwachstellen“
(pronominal)

Sind natürliche Sprachen regulär ?

NEIN!

Es treten **geschachtelte Abhängigkeiten** auf:

Die Katze, die den Hund, der die Ratte biß, jagte, starb.

Die formalen Sprachen

$$\begin{aligned}\mathcal{L}_1 &= \{a^n b^n \mid a, b \in \mathcal{V}\} \\ \mathcal{L}_2 &= \{a_1 \dots a_n a_n \dots a_1 \mid a_i \in \mathcal{V}\} \\ \mathcal{L}_2 &= \{a_1 \dots a_n b_n \dots b_1 \mid a_i, b_j \in \mathcal{V}\}\end{aligned}$$

sind nicht regulär!

Sind natürliche Sprachen kontextfrei ?

NEIN!

Es treten **verschränkte Abhängigkeiten** auf:

Jan säät das mer d'chind em Hans es huus lönd hälfe aastriche.
Jan sagt daß wir die Kinder dem Hans das Haus lassen helfen anstreichen

Die formalen Sprachen

$$\begin{aligned}\mathcal{L}_1 &= \{a^n b^n c^n \mid a, b, c \in \mathcal{V}\} \\ \mathcal{L}_2 &= \{a_1 \dots a_n a_1 \dots a_n \mid a_i \in \mathcal{V}\} \\ \mathcal{L}_2 &= \{a_1 \dots a_n b_1 \dots b_n \mid a_i, b_j \in \mathcal{V}\}\end{aligned}$$

sind nicht kontextfrei!

Sind natürliche Sprachen deshalb kontextsensitiv?

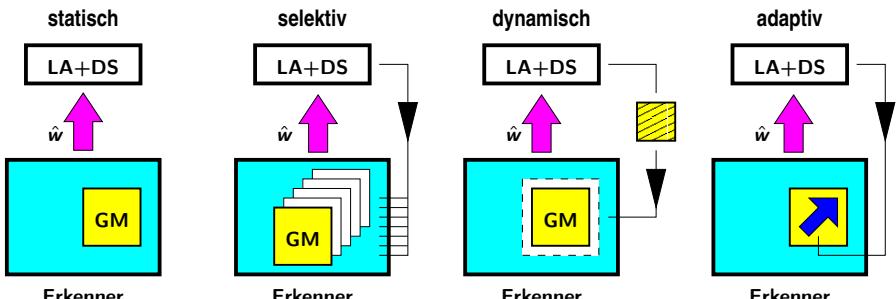
NICHT UNBEDINGT!

- **Flache Schachtelungen** sind problemlos regulär modellierbar.
Hierarchie statt Rekursion \rightsquigarrow endliche Sprache
- **Tiefe Schachtelungen** sind vielleicht grammatisch korrekt,
aber von geringer Akzeptabilität

MODELLIERUNG DER SPRACHPERFORMANZ

- **Strukturelle Mächtigkeit** eines Grammatikformalismus
- **Robuste Akquirierbarkeit** seiner Regeln und Parameter
- Übergeneralisierung \leftrightarrow Überanpassung

Integration der Sprachmodellkomponente in das ASE-System



Eine einzige **fixe** Grammatik wird „lebenslang“ verwendet.

Aus mehreren alternativen G-Modellen wird **dialogschrittabhängig** ein Passendes ausgewählt.

Linguistische Analyse (LA) und Dialogsteuerung (DS) produzieren **dynamisch** das aktuell passende G-Modell.

Das G-Modell wird durch Rückmeldung von LA & DS laufend **adaptiert**.

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Kettenregel · begrenztes Gedächtnis

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Kettenregel für die stochastische Wortsequenzerzeugung

DEFINITION BEDINGTER WAHRSCHEINLICHKEITEN & Kettenregel

$$P(B|A) = P(A, B) / P(A) \quad \Rightarrow \quad P(A, B) = P(A) \cdot P(B|A)$$

Satzproduktionsmodell:

$$\begin{aligned} P(w_1 w_2 \dots w_T) &= P(w_1) \\ &\cdot P(w_2 | w_1) \\ &\cdot P(w_3 | w_1 w_2) \\ &\cdot P(w_4 | w_1 w_2 w_3) \\ &\cdot \dots \dots \dots \\ &\cdot P(w_T | w_1 \dots w_{T-1}) \\ &= \prod_{t=1}^T P(w_t | w_1 \dots w_{t-1}) \end{aligned}$$

Bedingte Wortverteilung mit beschränktem Kontext

$$P(w_t | w_1 \dots w_{t-1}) \approx P(w_t | \underbrace{w_{t-n+1} \dots w_{t-1}}_{(n-1)\text{-Gramm-Kontext}})$$

Unigramm-, Bigramm- und Trigramm-Modelle

$$P_{1g}(w) = \prod_{t=1}^T P(w_t)$$

$$P_{2g}(w) = P(w_1) \cdot \prod_{t=2}^T P(w_t | w_{t-1})$$

$$P_{3g}(w) = P(w_1) \cdot P(w_2 | w_1) \cdot \prod_{t=3}^T P(w_t | w_{t-2} w_{t-1})$$

n-Gramm Sprachmodelle

$$P_{ng}(w) = \prod_{t=1}^T P(w_t | w_{t-n+1}^{t-1})$$

Bemerkungen

Das **Unigramm**-Modell besitzt keinerlei Gedächtnis und $L - 1$ freie Parameter.

Das **Bigramm**-Modell ist äquivalent zu einer Markovquelle (π, A) erster Ordnung und besitzt $L^2 - 1$ freie Parameter.

Das **Trigramm**-Modell entspricht einer Markovquelle mit $L^2 + L$ Zuständen.

Das **n-Gramm**-Modell besitzt ca. L^n freie Parameter:

$$\sum_{i=0}^{n-1} L^i \cdot (L - 1) = \frac{L^n - 1}{L - 1} \cdot (L - 1) = L^n - 1$$

Partiell inhomogene Wortwahrscheinlichkeiten, z.B.:

$$P_{2g}(w) = P(w_1 | w_\alpha) \cdot P(w_2 | w_1) \cdot P(w_3 | w_2) \cdot P(w_4 | w_3) \cdot P(w_\omega | w_4)$$

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

ML-Schätzung · lineare Interpolation · EM-Algorithmus

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Maximum-Likelihood-Schätzformeln

Definition

Es sei $\mathbf{u} \in \mathcal{V}^N$ eine Textdatenprobe. Das **n-Gramm-Sprachmodell** $P_{ng}(\cdot)$ nimmt auf \mathbf{u} die maximale Wahrscheinlichkeit an, wenn seine Parameter gemäß der Formel

$$\hat{P}(w|v) = \frac{\#\mathbf{u}(vw)}{\#\mathbf{u}(v)} = \frac{\text{„Anzahl der } n\text{-Gramme } vw“}}{\text{„Anzahl der } (n-1)\text{-Gramme } v“}}$$

relativer Häufigkeiten geschätzt werden.

Bemerkung

Genaugenommen steht der obige Nenner $\#\mathbf{u}(v)$ für die Summe

$$\sum_{w' \in \mathcal{V}} \#\mathbf{u}(vw') .$$

Beide Ausdrücke sind nur dann gleich, wenn die Sätze des Korpus \mathbf{u} durch Anfangs- und Endmarkierungen voneinander separiert sind.

Das Problem beschränkten Lerndatenumfangs

- Wenn der Zähler $\#_u(vw) = 0$ ist (unbeobachtetes n -Gramm), so ist auch $\hat{P}(w|v)$ **gleich Null**.
- Wenn der Nenner $\#_u(v) = 0$ ist (unbeobachtetes $(n-1)$ -Gramm), so ist $\hat{P}(w|v)$ sogar **undefiniert**.
- Ist $P(v)$ die **wahre** Auftretenswahrscheinlichkeit einer Wortsequenz, so ist $N \cdot P(v)$ der Erwartungswert der absoluten Auftretenshäufigkeit $\#_u(v)$, $u \in \mathcal{V}^N$.

Definition

Ein Worteintrag $w \in \mathcal{V}$, der im Lerndatenkorpus u nicht auftritt ($\#_u(w) = 0$), heißt **ungesehenes Wort** ('out of vocabulary', OOV).

Folgerung

Für ungesuchte Wörter w verschwindet die (ML-geschätzte) Unigrammwahrscheinlichkeit $\hat{P}(w)$.

Alle bedingten Wortwahrscheinlichkeiten $\hat{P}(v | \dots w \dots)$, $v \in \mathcal{V}$ sind **undefiniert**.

Glättung durch Verfärbung der Zählfunktion $\#(\cdot)$

$$\hat{p}_\ell^{\text{ML}} = \frac{N_\ell}{N} \quad \Rightarrow \quad \tilde{p}_\ell = N_\ell^* \Bigg/ \sum_{k=1}^L N_k^*$$

- Laplace-Glättung** (uniforme Bayesschätzung)

$$N_\ell^* \stackrel{\text{def}}{=} N_\ell + 1 \quad \rightsquigarrow \quad \tilde{p}_\ell = \frac{N_\ell + 1}{N + L}$$

- Jeffrey-Glättung**

$$N_\ell^* \stackrel{\text{def}}{=} N_\ell + \frac{1}{2} \quad \rightsquigarrow \quad \tilde{p}_\ell = \frac{N_\ell + \frac{1}{2}}{N + \frac{L}{2}}$$

- Quadratmittel-Glättung**

$$N_\ell^* \stackrel{\text{def}}{=} N_\ell + \frac{1}{2}\sqrt{N}$$

- Licklider-Glättung** ('relative discounting')

$$N_\ell^* \stackrel{\text{def}}{=} N_\ell + \varrho \quad \rightsquigarrow \quad \tilde{p}_\ell = \frac{N_\ell + \varrho}{N + \varrho \cdot L}$$

Lineare Interpolation von n -Gramm-Prädiktoren

Bilaterale Interpolation (n -Gramm & Zerogramm)

$$\tilde{P}(w|\mathbf{v}) \stackrel{\text{def}}{=} (1 - \lambda_0) \cdot \hat{P}_n(w|\mathbf{v}) + \lambda_0 \cdot 1 / L$$

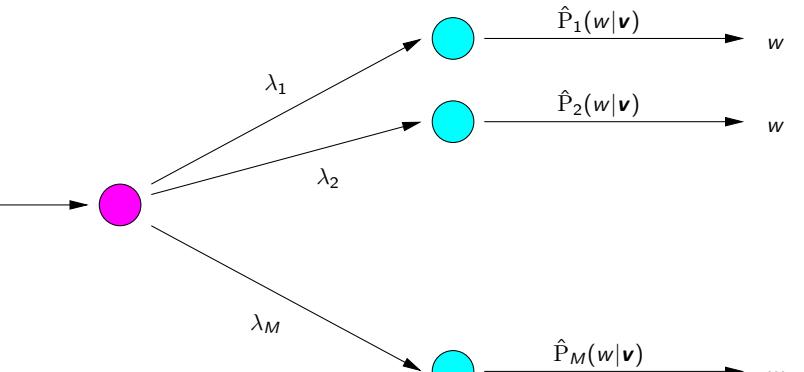
Optimales Gewicht $\lambda_0 \approx \eta_1/N$ mit $\eta_1 := \#\{w \mid N_w = 1\}$

Multilaterale Interpolation (alle ν -Gramme von 0 bis n)

$$\begin{aligned} \tilde{P}^{(n)}(w_t \mid w_1 \dots w_{t-1}) &= \lambda_0^{(n)} \cdot \frac{1}{L} + \lambda_1^{(n)} \cdot \hat{P}(w_t) \\ &\quad + \lambda_2^{(n)} \cdot \hat{P}(w_t \mid w_{t-1}) \\ &\quad + \lambda_3^{(n)} \cdot \hat{P}(w_t \mid w_{t-2} w_{t-1}) \\ &\quad + \dots \dots \dots \\ &\quad + \lambda_n^{(n)} \cdot \hat{P}(w_t \mid w_{t-n+1} \dots w_{t-1}) \end{aligned}$$

Welches sind die optimalen Interpolationsgewichte $\sum_\nu \lambda_\nu^{(n)} = 1$?

Zweistufiger Zufallsprozeß



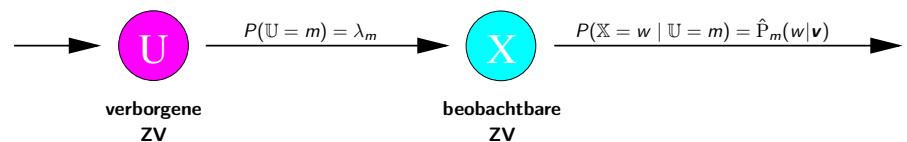
1. Prozeß

Bei Wortposition t wird mit Wahrscheinlichkeit λ_ν der Wortprädiktor \hat{P}_ν ausgewählt.

2. Prozeß

Das aktuelle Wortvorkommen w_t wird aus dem Kontext $w_{t-\nu+1} \dots w_{t-1}$ der Breite $(\nu - 1)$ vorhergesagt.

Expectation-Maximization Prinzip



Aufgabenstellung

Schätzung der ML-optimalen Parameter

$$\{\lambda_\nu^*\} = \underset{\{\lambda_\nu\}}{\operatorname{argmax}} \prod_t P(\mathbb{X} = w_t | \{\lambda_\nu\})$$

zur Randverteilung

$$\underbrace{P(\mathbb{X} = w | \{\lambda_\nu\})}_{\text{Randverteilung}} = \sum_\nu \underbrace{P(\mathbb{X} = w, \mathbb{U} = m | \{\lambda_\nu\})}_{\text{Verbundverteilung}}$$

eines Zufallsprozesses mit latenten Variablen.

Expectation-Maximization Algorithmus

1 DATENPARTITIONIERUNG

- Zerlegung der Textdatensammlung in zwei Teile
- Schätzung der Prädiktoren $\hat{P}_\nu(\cdot)$, $\nu = 0..n$ aus der ersten Probe

2 INITIALISIERUNG

Als Startparameter werden $\lambda_0 = \lambda_1 = \dots = \lambda_n = \frac{1}{(n+1)}$ verwendet.

3 EXPECTATION-SCHRITT

Berechnung der a posteriori Auswahlwahrscheinlichkeiten

$$\gamma_t(\nu) = P(\mathbb{U}_t = \nu | \mathbf{w}) \propto \lambda_\nu \cdot \hat{P}_\nu(w_t | \mathbf{w}_{t-\nu+1}^{t-1})$$

4 MAXIMIZATION-SCHRITT

Neuberechnung der Interpolationsgewichte

$$\lambda'_\nu = \sum_{t=1}^T \gamma_t(\nu) / T, \quad \nu = 0, 1, 2, \dots, n$$

5 TERMINIERUNG

Ende oder weiter bei Schritt (3)

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Monte-Carlo Simulation · Beispieldaten

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

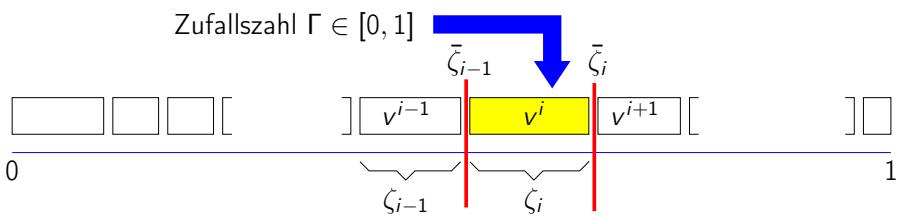
Modellgetriebene stochastische Texterzeugung

Gegeben:

Ein n -Gramm-Sprachmodell $\tilde{P}(\mathbf{w}) = \prod_t \tilde{P}(w_t | \mathbf{w}_1^{t-1})$

Gesucht:

Eine Folge $u_1 u_2 u_3 \dots u_t u_{t+1} \dots$ von Wörtern mit asymptotischer Wahrscheinlichkeitsverteilung $\tilde{P}(\cdot)$



N-Gramme auf Buchstabenebene

- **Trainingsmaterial**

ich möchte gerne wissen ob der Supersparpreis auch an Ostern gilt
 kann ich am Samstag Abend nach zweiundzwanzig Uhr noch von Frankfurt nach München kommen
 ich möchte am ersten Feiertag nach Koblenz fahren
 ich möchte nächsten Montag nach Paris

- **Polygramm 1. Ordnung (Unigramm)**

naM
 dM
 o hcittleG
 èVcdi ot egno n reulm v mimejcnbnifhg vkaneC t ohgh nrod AcwFlel
 uodheatrdmnas e
 ljngAnnncrnnfecarhc erZnh

- **Polygramm 2. Ordnung (Bigramm)**

n Häg
 Audaharten
 äg
 ien ns de ngüchespabugenden n anchm Ullen Febugenah mmiechau ienür
 heittamich e
 kona min

N-Gramme auf Buchstabenebene

Fortsetzung: $N \in \{3, 4\}$

- **Trainingsmaterial**

ich möchte gerne wissen ob der Supersparpreis auch an Ostern gilt
 kann ich am Samstag Abend nach zweiundzwanzig Uhr noch von Frankfurt nach München kommen
 ich möchte am ersten Feiertag nach Koblenz fahren
 ich möchte nächsten Montag nach Paris

- **Polygramm 3. Ordnung (Trigramm)**

nach Güber späch ben fürg garlie naupt wann Zugs esteleiunch Febzein einen
 Mür hen
 wein Ulm
 neit fünft
 nen
 indund drei Ulm egen Ulmittag

- **Polygramm 4. Ordnung (Tetragramm)**

nach Flughaffentag
 ich in ob derhonnerst wann fahren ich Morgen
 ach möchte alich möcht ja
 von heute
 kann im Geschen Vormatione Verbinund welchen

N-Gramme auf Buchstabenebene

Fortsetzung: $N \in \{5, 6, 7\}$

- **Polygramm 5. Ordnung (Pentagramm)**

nach Bamber muß Dienstag von Freitag von nach Köln
 nein Zug nach nach München zehn Uhr dreiundzwanz
 kann
 nein liebsten Samstag
 gibt etwa um sechs Uhr

- **Polygramm 6. Ordnung (Hexagramm)**

nach Bamberg von Abend ich am Donnen sieber ersten Samstag von Münster
 die nächt
 kann ich es noche ab dreinen Zug nach München
 gibt es geht möchte vom Freitag zum Tragflächen
 ich benruh
 ich muß schneller

- **Polygramm 7. Ordnung (Heptagramm)**

nach Bamberg sein
 Grüß in Saarbrücken auf Wiederhören
 etwa um drei Uhr in Hamburg aus in es möglichkeinen Fall ich nicht anke
 kann wäre die
 ich Morgens in der erste Zug nach Münchetzt

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Wortkategorien · nicht/deterministische Systeme · Gruppierung

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Die 44 Grundkategorien des Pedro Bermudo (1610–1648)

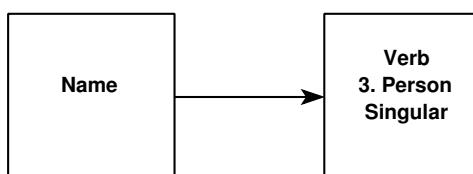
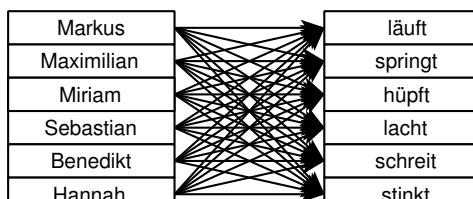
Umberto Eco, 1994:

- „1. Elemente. 2. Himmlische Größen. 3. Geistige Größen. 4. Weltliche Größen. 5. Kirchliche Größen. 6. Kunstgriffe. 7. Instrumente. 8. Affekte. 9. Religion. 10. Sakramentale Konfession. 11. Gericht. 12. Armee. 13. Medizin. 14. Häßliche Tiere. 15. Vögel. 16. Reptilien und Fische. 18. Gerätschaften. 19. Speisen. 20. Getränke und andere Flüssigkeiten. 21. Kleider. 22. Seidengewebe. 23. Wollstoffe. 24. Segeltücher und andere Textilien. 25. Nautica und Aromen. 26. Metalle und Münzen. 27. Diverse Artefakte. 28. Steine. 29. Juwelen. 30. Bäume und Früchte. 31. Öffentliche Orte. 32. Maße und Gewichte. 33. Zahlen. 34. Zeit. 35-42. Nomina, Adjektive, Adverbien und so weiter. 43. Personen. 44. Wanderschaft.“

Wortkategorien („parts-of-speech“)

Die Gesamtheit aller Wörter einer Sprache zerfällt (?) in Kategorien (?) ähnlicher (?) Wörter.

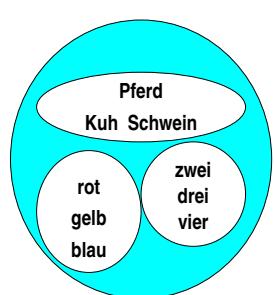
Wörter mit ähnlichen statistischen Verteilungseigenschaften



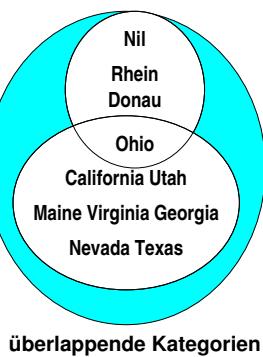
Funktionale Äquivalenz

Wörter sind einander **ähnlich**, falls sie sprachlich in gleicher Weise verwendbar sind („Austauschbarkeit“).

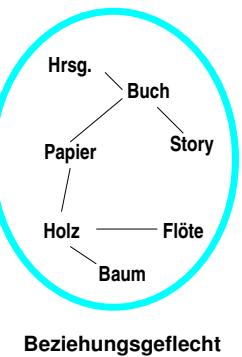
Repräsentation von Wortassoziationen



disjunkte Kategorien



überlappende Kategorien



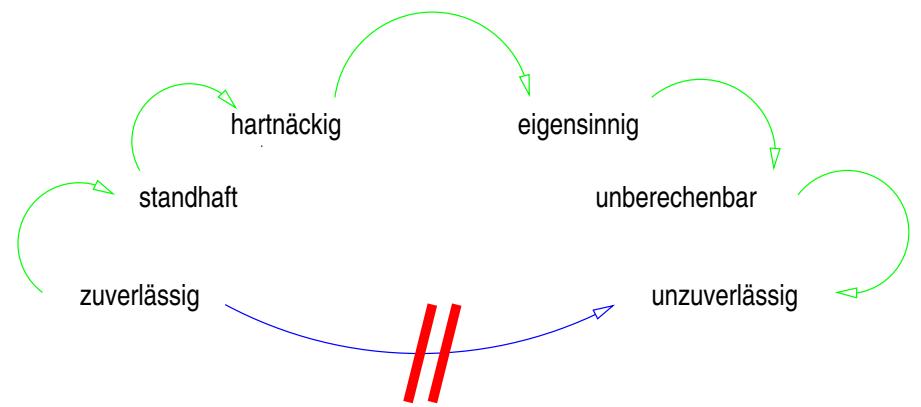
Beziehungsgeflecht

Repräsentation von Wortassoziationen

Es gibt Ähnlichkeitsbeziehungen („Assoziationen“) zwischen Wörtern.

- ⇒ Resultiert daraus eine *Gruppenbildung* ?
- ⇒ Sind die Gruppen paarweise *disjunkt* ?

Unschärfe — nichttransitive Synonymie



Fakt

Es ist davon auszugehen, daß jegliche Strukturierung des Wortschatzes zu 90% im Auge des Betrachters liegt!

Deterministische Kategoriensysteme I

$$\mathcal{C} = \{C^1, C^2, \dots, C^K\}$$

- Eindeutige Wort-Kategorie-Abbildung

$$\hat{\kappa} : \begin{cases} \mathcal{V} & \rightarrow \mathcal{C} \\ w & \mapsto \hat{\kappa}(w) = \text{eindeutige Wortkategorie von } w \end{cases}$$

- Kategorien als Partition (disjunkte Zerlegung)

$$\mathcal{V} = C^1 \cup C^2 \cup \dots \cup C^K \quad \text{mit } C^i \cap C^j = \emptyset \text{ für alle } i \neq j$$

- Eindeutige kategoriale Annotation von Sätzen

$$\hat{\kappa}^{(T)} : \begin{cases} \mathcal{V}^T & \rightarrow \mathcal{C}^T \\ w_1 \dots w_T & \mapsto c_1 \dots c_T = \hat{\kappa}(w_1)\hat{\kappa}(w_2) \dots \hat{\kappa}(w_T) \end{cases}$$

Deterministische Kategoriensysteme II

Kategoriebezogene bedingte Wortwahrscheinlichkeiten

$$\begin{aligned} P(w_t | w_1 \dots w_{t-1}) &= P(w_t, c_t | w_1 \dots w_{t-1}) \\ &= P(w_t | c_t, w_1 \dots w_{t-1}) \cdot P(c_t | w_1 \dots w_{t-1}) \\ &\approx P(w_t | c_t) \cdot P(c_t | c_1 \dots c_{t-1}) \end{aligned}$$

Kategoriebezogene Satzwahrscheinlichkeiten

$$\begin{aligned} P(w_1 \dots w_T) &= \prod_{t=1}^T P(w_t | w_1 \dots w_{t-1}) \\ &= \underbrace{\prod_{t=1}^T P(w_t | c_t)}_{P(\mathbf{w} | \mathbf{c})} \cdot \underbrace{\prod_{t=1}^T P(c_t | c_1 \dots c_{t-1})}_{P(\mathbf{c})} \end{aligned}$$

Deterministische Kategoriensysteme III

Kategoriebezogene Zählfunktion

$$\#(c_1 \dots c_n) \stackrel{\text{def}}{=} \sum_{w_1 \in c_1} \dots \sum_{w_n \in c_n} \#(w_1 \dots w_n)$$

ML-Schätzung der bedingten Wortwahrscheinlichkeit

$$\hat{P}(w | c) = \#(w) / \#(c)$$

ML-Schätzung der Kategorie- n -Gramm-Wahrscheinlichkeit

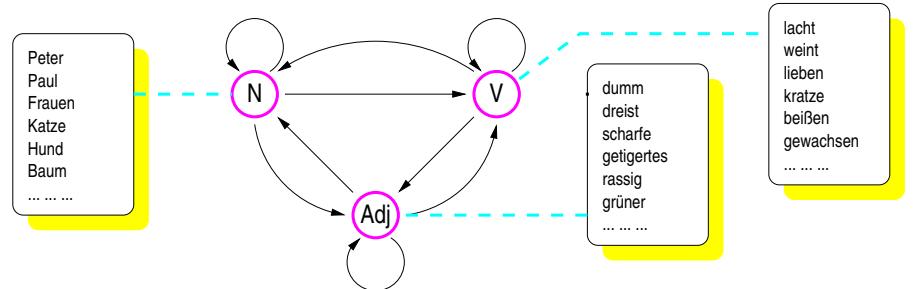
$$\hat{P}(c | b_1 \dots b_{n-1}) = \#(b_1 \dots b_{n-1} c) / \#(b_1 \dots b_{n-1})$$

Beispiel

Das kategoriebezogene n -Gramm-Modell besitzt $(K^n - 1) + (L - K)$ freie Wahrscheinlichkeitsparameter, zum Beispiel $(10^6 - 1 + 900)$ statt $(10^9 - 1)$ im Falle $n = 3$, $L = 10^3$, $K = 10^2$.

Nichtdeterministische Kategoriensysteme

Beobachtete Wörter $\mathbf{w} = w_1 \dots w_T$ · verborgene Kategorien $\mathbf{c} = c_1 \dots c_T$



Kategoriales n -Gramm-Modell als Randverteilung

$$P(\mathbf{w}) = \sum_{\mathbf{c} \in \mathcal{C}^T} \left[\prod_{t=1}^T \{P(w_t | c_t) \cdot P(c_t | c_{t-n+1} \dots c_{t-1})\} \right]$$

Das nichtdeterministische, kategoriale Bigramm-Modell ist ein **Hidden-Markov-Modell**.

Motivation

Grammatikkomponenten in der Spracherkennung

n -Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

ML-Kriterium · Kombinatorische Suche · Globale Suchalgorithmen

Optimale disjunkte Wortkategoriensysteme

Gegeben:

- Wortschatz $\mathcal{V} = \{W^1, W^2, W^3, \dots, W^L\}$
- Kategoriealphabet $\mathcal{C} = \{C^1, C^2, C^3, \dots, C^K\}$
- Textkorpus $w = w_1 \dots w_T \in \mathcal{V}^T$

Gesucht:

- Partition $\mathfrak{K} : \mathcal{V} \rightarrow \mathcal{C}$ des Wortschatzes mit

$$P_{\mathfrak{K}}(w) = P(w | \mathfrak{K}(w)) \cdot P(\mathfrak{K}(w)) \quad \xrightarrow{\text{MAX}}$$

Problem:

Diese kombinatorische Suchaufgabe ist von exponentieller Komplexität, denn es gibt $L^K / K!$ wesentlich verschiedene Partitionen von \mathcal{V} mit der Kardinalität $K = |\mathcal{C}|$.

Spezialfall: Kategoriebigramme mit ML-Parametern

$$P_{\mathfrak{K}}(w) = \prod_{t=1}^T \underbrace{\left(P(w_t | \mathfrak{K}(w_t)) \cdot P(\mathfrak{K}(w_t) | \mathfrak{K}(w_{t-1})) \right)}_{q(w_t | c_{t-1})}$$

Maximum-Likelihood-Schätzwerte der Parameter

$$q(w|c) = \hat{P}(\mathbb{W}_t = w | \mathfrak{K}(\mathbb{W}_{t-1}) = c) = \frac{\#(cw)}{\#(c)} = \frac{\sum_{\mathfrak{K}(v)=c} \#(vw)}{\sum_{\mathfrak{K}(v)=c} \#(v)}$$

Zielfunktion der Kategorieoptimierung

$$\log \hat{P}_{\mathfrak{K}}(w) = \sum_{b,c \in \mathcal{C}} \#(bc) \log \#(bc) - 2 \cdot \sum_{c \in \mathcal{C}} \#(c) \log \#(c) + \text{konst}$$

Hill climbing — Verfahren des steilsten Anstiegs

-
- 1 Wähle zufälligen **Startwert** $\mathfrak{K}^* \in \mathcal{C}^{\mathcal{V}}$ und setze $\phi^* = f(\mathfrak{K}^*)$.

- 2 Bestimme die Menge **benachbarter** Lösungskandidaten:

$$\mathcal{U} = \mathcal{U}(\mathfrak{K}^*) = \{ \mathfrak{K} \in \mathcal{C}^{\mathcal{V}} \mid \mathfrak{K} \text{ und } \mathfrak{K}^* \text{ unterscheiden sich in einem Wert} \}$$

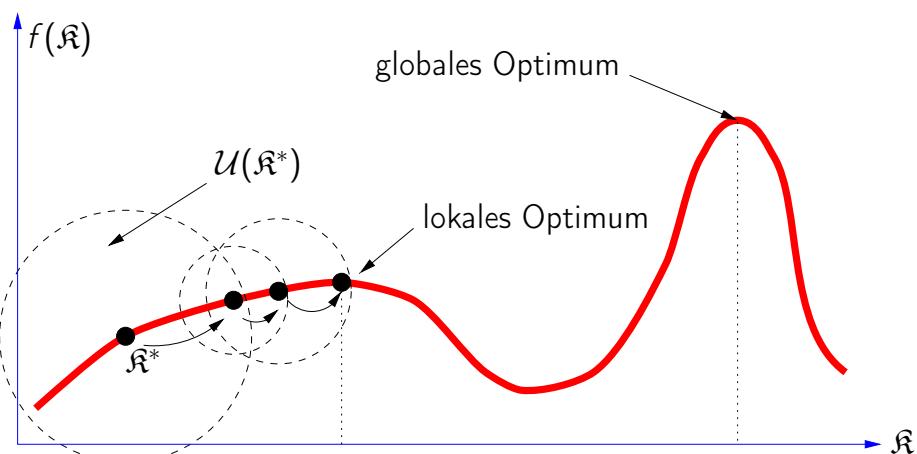
- 3 Bestimme den **lokalen Sieger**

$$\tilde{\mathfrak{K}} = \underset{\mathfrak{K} \in \mathcal{U}}{\operatorname{argmax}} f(\mathfrak{K})$$

und seine Bewertung $\tilde{\phi} = f(\tilde{\mathfrak{K}})$.

- 4 Falls $\tilde{\phi} < \phi^*$, war \mathfrak{K}^* schon lokal optimal; \rightsquigarrow ENDE.
Sonst ersetze $\mathfrak{K}^* = \tilde{\mathfrak{K}}$, $\phi^* = \tilde{\phi}$ und marschiere zurück $\rightsquigarrow [1]$.

Hill climbing (HC)

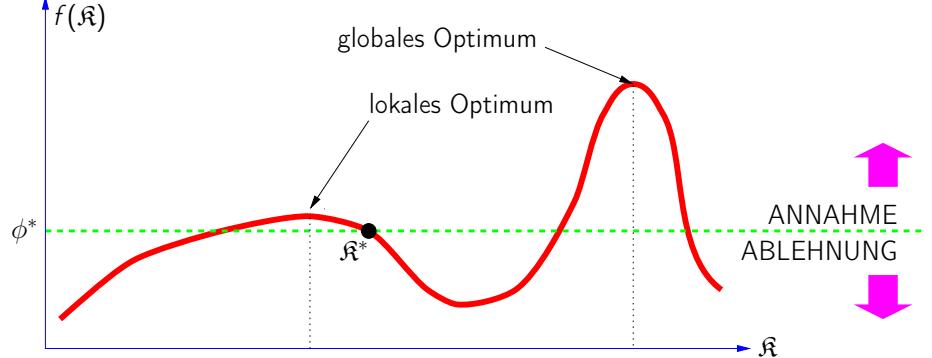


Stochastische Relaxation — randomisierte Kandidatenwahl

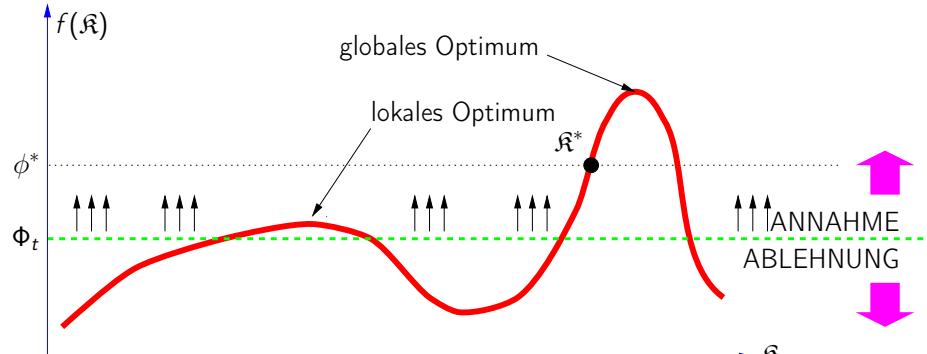
$$\mathfrak{T} : \begin{cases} \mathcal{C}^V & \rightarrow \mathcal{C}^V \\ \kappa & \mapsto \kappa' \text{ gemäß einer vorgegebenen Verteilung } P_{\mathbb{K}'|\mathbb{K}}(\cdot|\kappa) \end{cases}$$

-
- (Algorithmus)**
- 1 Initialisiere per Zufall $\kappa^* \in \mathcal{C}^V, \phi^* = f(\kappa^*)$
 - 2 Würfe einen Nachfolger aus $\kappa = \mathfrak{T}(\kappa^*)$
 - 3 Bewerte den Nachfolger $\phi = f(\kappa)$
 - 4 Prüfe die Annahmebedingung $\phi > \phi^*$
 - 5 Ersetze im Erfolgsfall das Zwischenergebnis $\kappa^* = \kappa, \phi^* = \phi$
 - 6 Prüfe die Abbruchbedingung $\rightsquigarrow [2] \text{ oder } \rightsquigarrow \text{ENDE}$
-
- (zum Vergleich)**

Stochastische Relaxation (SR)



Sintflutalgorithmus (GD: great deluge)



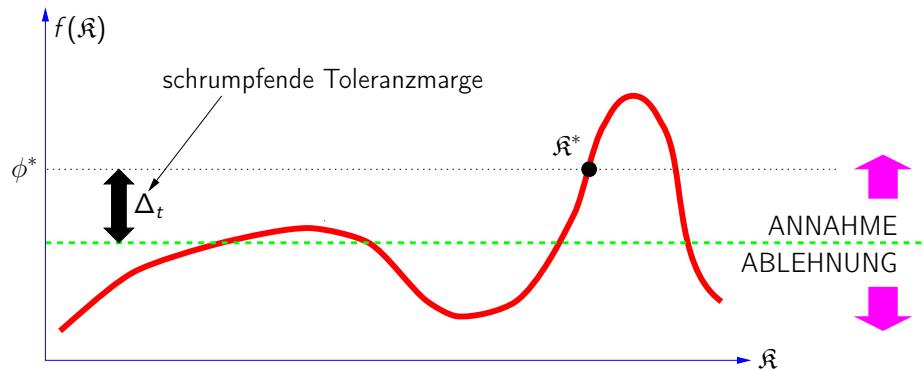
Führe einen zeitabhängigen **Wasserstand** ein

mit einer geeigneten **Flutungsstrategie** (monoton steigend)

$$\Phi_{t+1} = \Phi_t + \delta$$

und überprüfe die Annahmebedingung $\phi > \Phi_t$.

Schwellwertannahme (TA: threshold acceptance)



Führe eine zeitabhängige **Toleranzschwelle** $\Delta : \mathbb{N} \rightarrow \mathbb{R}$ ein, z.B.:

$$\Delta_{t+1} = \Delta_t \cdot (1 - \delta)$$

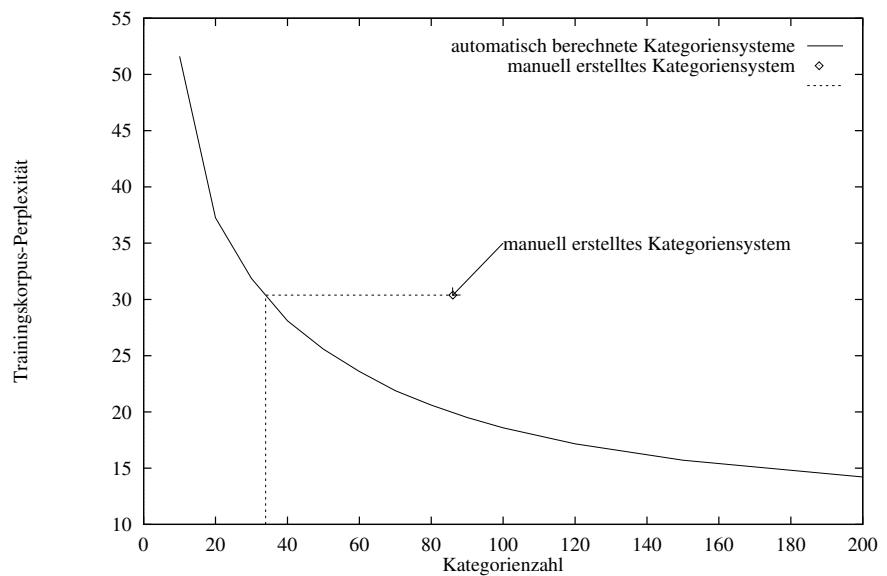
und überprüfe die Annahmebedingung $\phi > \phi^* - \Delta_t$.

Automatisch optimierte Kategoriensysteme

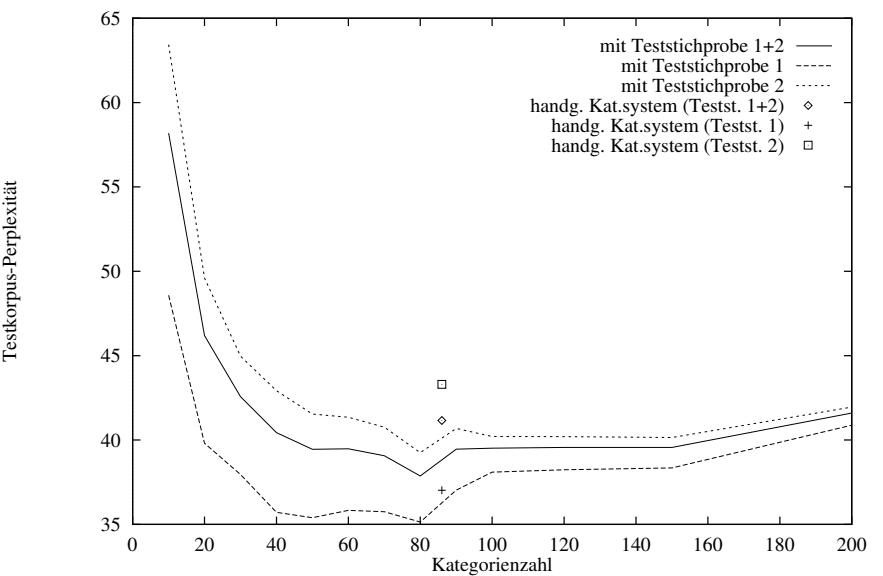
Beispiel

1. wo wann ob mitnehmen*
2. Vormittag Nachmittag Mittag Abend März Zeiten*
3. einundzwanzigsten zehnten einunddreißigsten* vierten sieben elften Weihnachtsfeiertag zweiten fünften ersten vierundzwanzigsten sechsten* siebzehnten zweiundzwanzigsten* neunzehnten* dreieinundzwanzigsten siebenundzwanzigsten* dritten zehnter*
4. Koblenz Hof* Dortmund Saarbrücken Osnabrück* Ulm Augsburg Frankfurt Paris Nürnberg Göttingen* Köln Bebra Weihnachten Heidelberg Würzburg Bonn
5. Ochtrup Mannheim Bamberg* Hamburg Athen Düsseldorf Graz* Berlin Abensberg* Solingen* Kiel* Oberstaufen* Utting* London Aachen Bremen Regensburg Wien Hause* Münster Stuttgart Rom Ansbach* Offenburg* Wuppertal* Hannover Karlsruhe Amsterdam*
6. Februar April Mai Juni Juli August September* Oktober Dezember* zweiundneunzig Vormittags einundneunzig Feiertag* neunzehnhunderteinundneunzig* Weihnachtstag
7. vierzehn wieviel neunzehn einundzwanzig fünfzehn zehn zwei vierundzwanzig dreizehn zweiundzwanzig zwanzig dreiundzwanzig achtzehn
8. Feiertagen Fahrrad* Gültigkeit* sechzehnten* Abfahrtszeit Woche S-Bahn

Bigrammperplexität optimierter Kategoriensysteme auf dem Lerndatenkorpus



Bigrammperplexität optimierter Kategoriensysteme auf dem Testdatenkorpus



Weitgespannte Wortabhängigkeiten

Motivation · Positionelle Bigramme · Lückenhafte n -Gramme

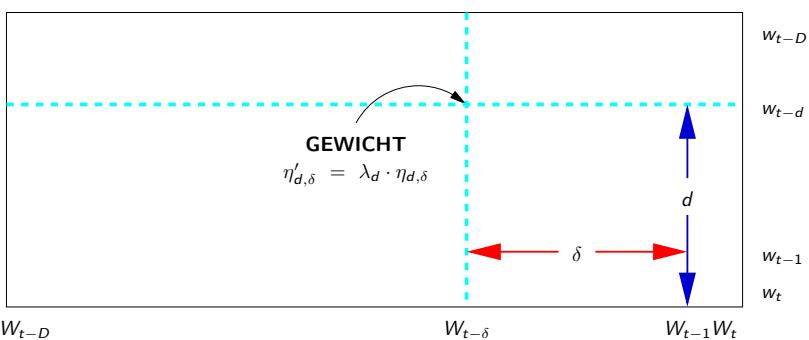
Positionelle Bigramme

Positionelle Verzerrung

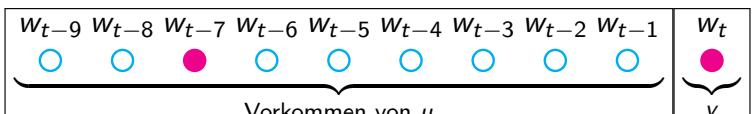
$$\tilde{P}_d(w|v) = \sum_{\delta=1}^D \eta_{d,\delta} \cdot \hat{P}_{\delta}(w|v) , \quad \sum_{\delta} \eta_{d,\delta} = 1$$

Doppelt konvexes Bigramm-Sprachmodell

$$\tilde{P}(w_t \mid \mathbf{w}_1^{t-1}) = \sum_{d=1}^D \sum_{\delta=1}^D \underbrace{\lambda_d \cdot \eta_{d,\delta}}_{\eta'_{d,\delta}} \cdot \hat{P}_{\delta}(w_t \mid w_{t-d})$$



Positionelle Bigramme



- Positionell gespreizte Bigrammwahrscheinlichkeiten

$$P_d(w|v) \stackrel{\text{def}}{=} P(\mathbb{W}_t \equiv w \mid \mathbb{W}_{t-d} \equiv v)$$

- Lineare Interpolation der Bigrammprädiktoren

$$\tilde{P}_d(w_t | \mathbf{w}_1^{t-1}) = \sum_{d=1}^D \lambda_d \cdot \hat{P}_d(w_t | w_{t-d})$$

- Maximum-Likelihood-Schätzer für positionelle Bigramme

$$\hat{P}_d(w|v) = \#(v \underbrace{\dots}_{(d-1)} w) / \#(v)$$

Lückenhafte n -Gramme

Lineare Interpolation von

- gewöhnlichen ν -Grammen $\nu = 0, \dots, n$
 - positionellen δ -Bigrammen $\delta = 0, \dots, n$
 - positionellen (δ, τ) -Trigrammen $\delta = 0, \dots, n$

	$\nu = ?$ 2 3 4 5	$\delta = ?$ 1 2 3 4	$\frac{\delta}{\tau} = ?$ $\frac{1}{1}$ $\frac{1}{2}$ $\frac{1}{3}$ $\frac{2}{1}$ $\frac{2}{2}$ $\frac{3}{1}$
w_{t-1}	● ● ● ●	● ○ ○ ○	● ● ● ○ ○ ○
w_{t-2}	○ ○ ○ ○	○ ● ○ ○	● ○ ○ ○ ○ ○
w_{t-3}	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○ ○ ○
w_{t-4}	○ ○ ○ ○	○ ○ ○ ○	○ ○ ○ ○ ○ ○
	ν -grams	δ -bigrams	δ, τ -trigrams

Motivation

Grammatikkomponenten in der Spracherkennung

n-Gramm Wahrscheinlichkeitsverteilungen

Schätzung bedingter Wortwahrscheinlichkeiten

Zufällige Erzeugung von Wortfolgen

Wörter und ihre Part-of-Speech Kategorien

Optimale disjunkte Wortkategoriensysteme

Weitgespannte Wortabhängigkeiten

Beispielaufbau

An Stelle einer Zusammenfassung

ERSTELLUNG EINER WORTKLASSENBEZOGENEN *n*-GRAMM-GRAMMATIK

Stochastische Satzgrammatik

- 1 **Erfassung des Anwendungsszenarios**
Handentwurf — Allgemeinkorpora — Mensch-Mensch-Dialoge - WOZ
- 2 **Erstellung eines Textdatenkorpus**
Kunstsyntax + Monte Carlo; reale Daten + Auswahl
Markierungen für Satzanfang, Satzende, OOV-Wörter
- 3 **Wortbezogenes Bigramm-Sprachmodell**
Häufigkeitsstatistik, Laplace-Glättung der ML-Schätzwerte
- 4 **Konstruktion von 100 Wortkategorien**
Clustern mittels sintflutgesteuerten Austauschverfahrens
- 5 **Kategoriebezogene *n*-Gramm-Sprachmodelle**
Interpolationskoeffizienten mit EM aus Validierungsdaten
- 6 **Auswahl des Tetragramm-Modells**
Perplexitätsvergleich für $n \in \{1, 2, 3, 4, 5, 6\}$ mit Entwicklungsdatensatz

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil VIII

Dekodierung mit HMMs

Motivation	HMM-Netzwerke	Wortsegmentierung	Synchrone Suche	Asynchrone Suche	Lexikon	Mehrphasendekodierung	Σ
oooooooo	oooooo	ooooo	ooooo	ooooooo	oooo	ooooooo	

Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Suche in Wortfolgenrichtung

Wortschatzorganisation

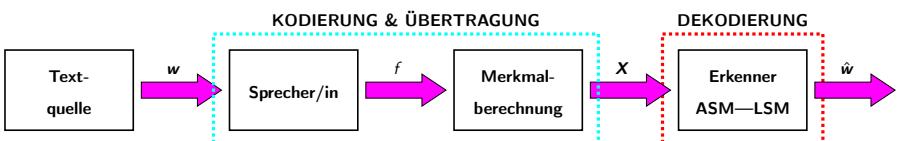
Mehrphasendekodierung

Beispielaufbau

Motivation	HMM-Netzwerke	Wortsegmentierung	Synchrone Suche	Asynchrone Suche	Lexikon	Mehrphasendekodierung
oooooooo	oooooo	ooooo	ooooo	ooooooo	oooo	ooooooo

Dekodierung \hat{w} \triangleq Maximierungsaufgabe

$$\hat{w}^* = \underset{w}{\operatorname{argmax}} \{ P_{AM}(X | w) \cdot P_{LM}(w) \}$$



Massives Resourcenproblem

1. viele Wörter, Modelle, Verteilungen
2. zerklüfteter Suchraum aufgrund mächtiger Grammatikmodelle
3. Kombinatorik unbekannter Wortgrenzen in kontinuierlicher Sprache

Lösungsansätze

Angriffsflächen

- **Rekombination** von Teillösungen
- **Beschneidung** des Suchraums ('pruning')
- **Sequentielle Dekomposition** der Analyse

Programmtechnisches Vorgehen

- Zeitliche Überlagerung bei der Speicherverwaltung
- Impliziter Suchraumaufbau
- Datenflußkontrolle: Dichteberechnungen, Cache

Risiken und Nebenwirkungen ?

Modellierungsfehler:

w^* ≠ gesprochene Wortfolge

Dekodierungsfehler:

gefundene Wortfolge ≠ w^*

Kompilierte Netzwerke aus HMMs

- **Bigramm**-Grammatik oder weniger
- Jedes Wortmodell $\lambda(W)$ besitzt je einen **E/A-Zustand**
- **Vernetzung** der Wort-HMMs im Sinne der Grammatik
- **Dekodierung** durch Viterbi-Algorithmus auf dem Netzwerk
- Optimale Zustandsfolge ↵ Lösung w^*



Sychrone Suche

Strikte Verarbeitung des Eingabesignals in Zeitrichtung
(„von links nach rechts“)

Motivation

Kompilierte HMM-Netzwerkstrukturen

Sychrone Suche · Einzelwort · Verbundwort · Bigramm-Modell

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

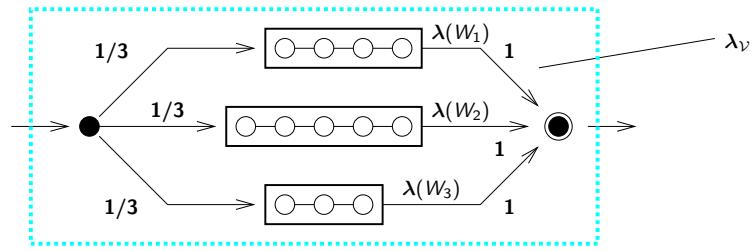
Suche in Wortfolgenrichtung

Wortschatzorganisation

Mehrphasendekodierung

Beispielaufbau

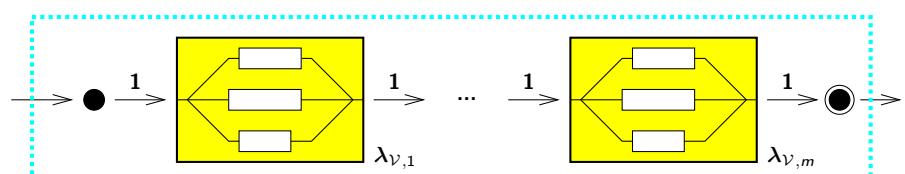
Einzelworterkennung



Die Modelle aller Wortschatzeinträge werden **parallel** geschaltet.

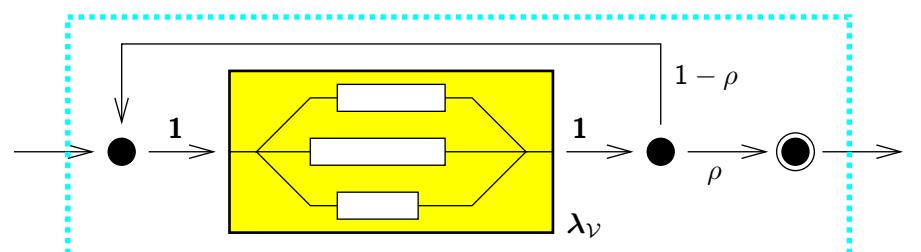
Es können **Unigrammwahrscheinlichkeiten** eingebracht werden.

Verbundworterkennung mit bekannter Satzlänge



Es werden m Wortmodellbündel **in Serie** geschaltet.

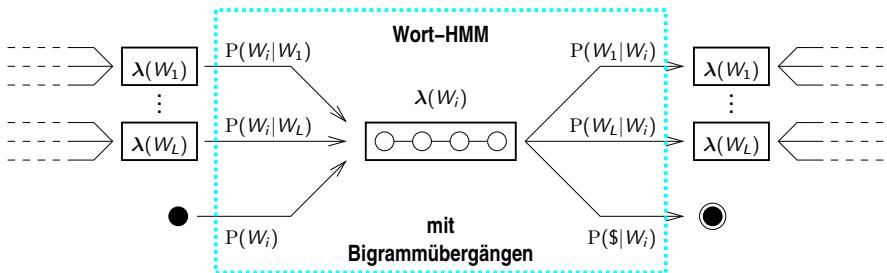
Verbundworterkennung mit unbekannter Satzlänge



Ein Wortmodellbündel wird zu einer **Schleife** verschaltet.

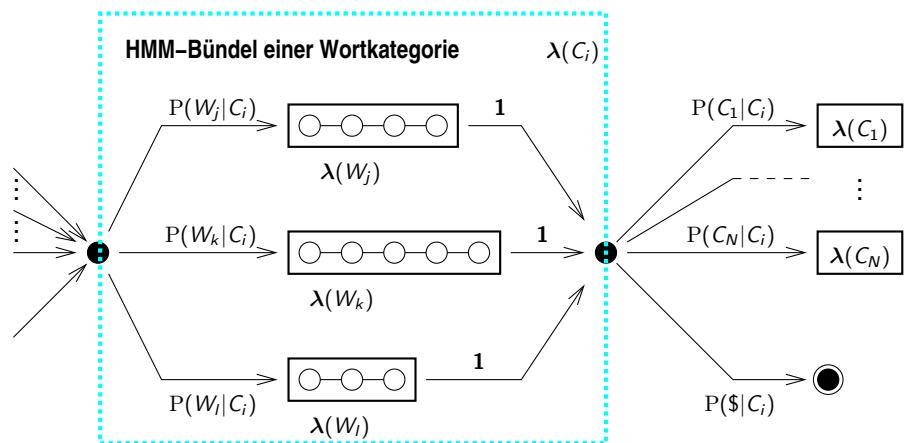
Eine **Fluchtwahrscheinlichkeit ρ** regelt die (mittlere) Wortanzahl.

Verbundworterkennung mit wortbezogenen Bigrammen



L Wortmodelle und L^2 Übergangskanten mit Bigramm-W'keiten

Verbundworterkennung mit kategoriebezogenen Bigrammen



L Wortmodelle im Falle **disjunkter** Wortkategorien

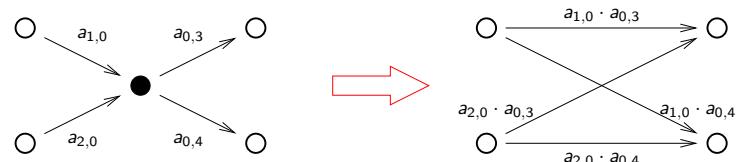
Wann „realisiert“ ein HMM-Netzwerk eine Grammatik?

Für alle Wortfolgen $w \in \mathcal{V}^*$ muß gelten:

$$P(X | \lambda(w)) \cdot P(w) \stackrel{?}{=} P(X, w | \lambda) \stackrel{\text{def}}{=} \sum_{q \in S^T|w} P(X, q | \lambda)$$

(es bezeichnet $S^T|w$ die Menge aller Zustandsfolgen der Dauer T , welche die Kette w traversieren)

Expansion **konfluenter** Zustände:



Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

One-Stage/Level-Building · Vorwärtsdekodierung · PTB

Suche in Zeitrichtung

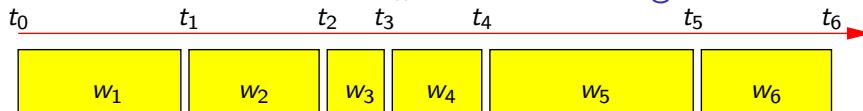
Suche in Wortfolgenrichtung

Wortschatzorganisation

Mehrphasendekodierung

Beispielaufbau

Welches ist die „beste“ Wortfolge?



- **Viterbi-Wortfolge**

$$w_{VA}^* = w(q^*) , \quad q^* = \underset{q \in S^T}{\operatorname{argmax}} \{P(w) \cdot P(X, q | \lambda(w))\}$$

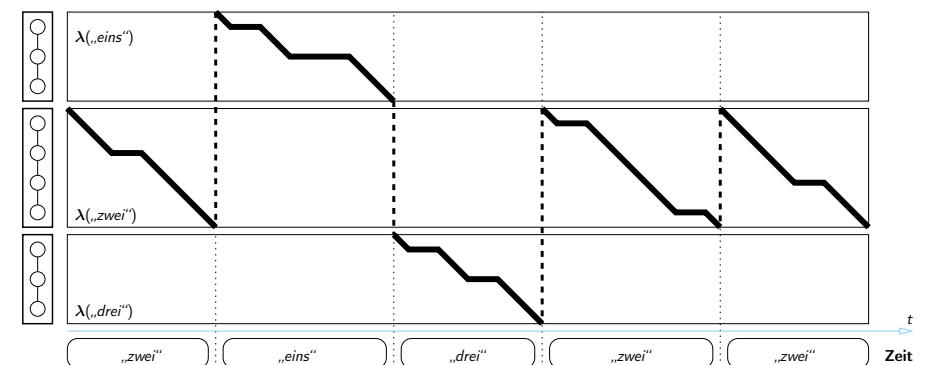
- **Optimale Wortsegmentierung**

$$(t^*, w^*) = \underset{t, w}{\operatorname{argmax}} P(t, w | X) = \underset{t, w}{\operatorname{argmax}} \{P(X, t | w) \cdot P(w)\}$$

- **Maximum a posteriori-Wortfolge**

$$w_{MAP}^* = \underset{w \in \mathcal{V}^*}{\operatorname{argmax}} \{P(w) \cdot P(X | \lambda(w))\}$$

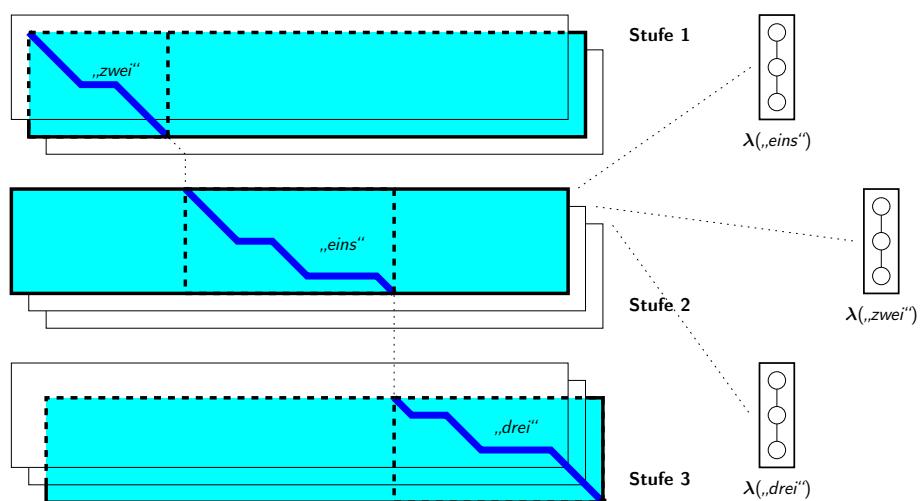
Einstufige Verbundwortdekodierung



One-stage Algorithmus

(Vintsyuk '71, Bridle '82)

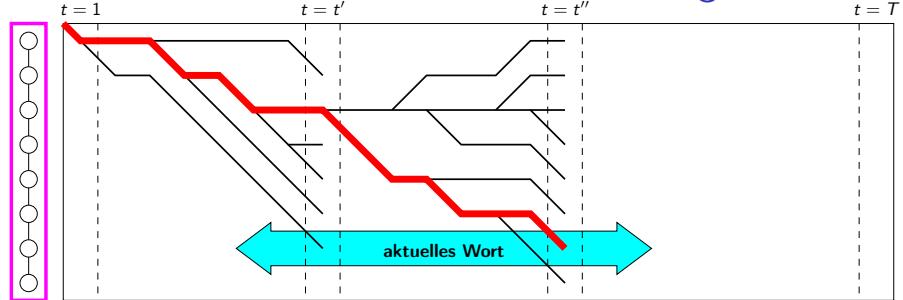
Mehrstufige Verbundwortdekodierung



Level-building Algorithmus

(Myers '81, Rabiner '85)

Schrifthaltende Teildekodierung



- Anfangspartien von w^* bereits eher als in $t = T$ berechnen !
~~ Worterkennung vor dem Wortende !
- $q(j, t)$ = wahrscheinlichste Folge, die in t den Zustand s_j erreicht
- Menge aller Zustände, die zum Zeitpunkt t' eingenommen wurden und auf einer optimalen, in t'' endenden Folge liegen:
$$Q_{t't''} = \{q_t(j, t'') \mid j = 1, \dots, N\}, \quad 1 \leq t' \leq t''$$
- Zwischenbilanz für Zeitpunkt t' , sobald $Q_{t't''}$ einelementig ist

Vorwärtsdekomposition

(Algorithmus)

- INITIALISIERUNG. Setze für alle $j = 1, \dots, N$

$$\vartheta_1(j) = \pi_j b_j(x_1) \quad \text{und} \quad \psi_1(j) = 0$$

- REKURSION.

Für alle $j = 1, \dots, N$ setze $\psi_t(j) = \operatorname{argmax}_i \vartheta_{t-1}(i) a_{ij}$ sowie

$$\vartheta_t(j) = \begin{cases} \max_i (\vartheta_{t-1}(i) a_{ij}) \cdot b_j(x_t) & \text{falls } s_j \text{ Wortanfangszustand ist} \\ \sum_i (\vartheta_{t-1}(i) a_{ij}) \cdot b_j(x_t) & \text{für alle sonstigen } s_j \end{cases}$$

- TERMINIERUNG. Setze

$$P^*(X \mid \lambda) = \vartheta_T(N) \quad \text{und} \quad q_T^* = \vartheta_T(N)$$

- RÜCKVERFOLGUNG. Für $t = t-1, \dots, 1$ setze $q_t^* = \psi_{t+1}(q_{t+1}^*)$

- LÖSUNGSWORTKETTE. Setze $w^* = w(q^*)$.

(zum Beispiel)

Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Strahlsuche · Vorwärts-Rückwärts-Suche

Suche in Wortfolgenrichtung

Wortschatzorganisation

Mehrphasendekodierung

Beispielaufbau

Viterbi-Algorithmus — vorwärts schauend

(Algorithmus)

1 INITIALISIEREN

Für alle $j \in \{1, \dots, N\}$ setze $t \leftarrow 1$ und $\vartheta_t(j) \leftarrow \pi_j \cdot b_j(\mathbf{x}_t)$.

2 VORBESETZEN

$$\vartheta_{t+1}(j) \leftarrow 0 \quad (\forall j)$$

3 VORWÄRTS FEUERN

$$\vartheta_{t+1}(j) \leftarrow \max \left\{ \frac{\vartheta_t(i) \cdot a_{ij}}{\vartheta_{t+1}(i)} \right\} \quad (\forall i, j)$$

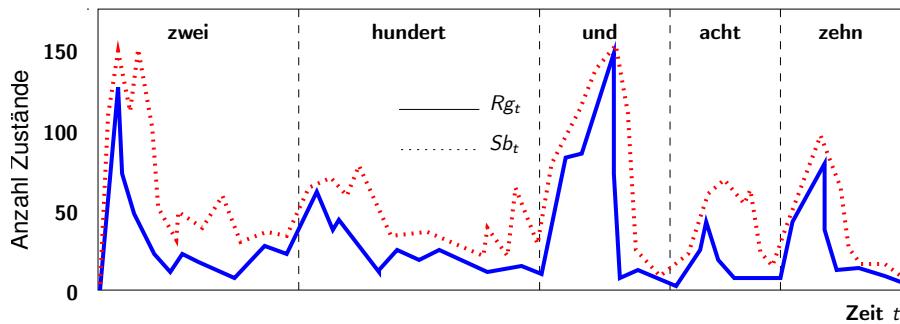
4 ABSCHLIESSEN

$$\vartheta_{t+1}(j) \leftarrow \vartheta_{t+1}(j) \cdot b_j(\mathbf{x}_{t+1}) \quad (\forall j)$$

5 WEITERSCHALTEN Setze $t \leftarrow t + 1$ oder \rightsquigarrow ENDE.

(zum Diagramm)

Strahlbreite und Hypothesenrang



- R_{gt} = lokaler Wahrscheinlichkeitsrang der global besten Wortkette

- S_{bt} = Anzahl konkurrierender Kandidatenzustände zum Zeittakt t

→ typische Aufwandsreduktion:
Faktor 10–20 bei $\leq 1\%$ erhöhter Fehlerrate

Strahlsuchverfahren

Obsolete Maximumoperationen

- falls $a_{ij} = 0$ oder
- falls $\vartheta_t(i) = 0$

Aktive & passive Zustände

$$\mathcal{O}_t \stackrel{\text{def}}{=} \{i \mid \vartheta_t(i) \neq 0\}$$

Passive Zustände müssen nicht mehr feuern!

Beschneidungsstrategie

$$\mathcal{O}_t^{B_0} \stackrel{\text{def}}{=} \{i \mid \vartheta_t(i) \geq B_0 \cdot \Lambda_t\} \quad \text{mit } \Lambda_t = \max_j \vartheta_t(j)$$

verfolgt nur eine kleine Schar *wahrscheinlichster* aktueller Zustände

- die Anzahl der Kandidaten („Strahlbreite“) ist adaptiv

$$B_0 = 10^{-2} \dots 10^{-4} \dots 0$$

Vorwärts-Rückwärts-Suche I

Problem

- Immer noch hohe Kandidatenanzahl an den Wortübergängen!
- Wörter von Strahlsuche aktiviert & gleich wieder deaktiviert.

Lösungsansätze

- getrennte Kandidatenlisten & Strahlkonstanten
- schnelle Vorauswahl mutmaßlicher Fortsetzungswörter
- Reduktion der Menge \mathcal{O}_t aktiver Zustände durch 'look-ahead':

1. Viterbi-Algorithmus vorwärts

mit einfachen akustischen und grammatischen Modellen

2. Speichern der „aktiven“ Wahrscheinlichkeitsbewertungen

$$\{\vartheta_t(i) \mid i \in \mathcal{O}_t, 1 \leq t \leq T\}$$

3. Viterbi-Algorithmus rückwärts

mit komplexeren akustischen und grammatischen Modellen zur Berechnung der zeitinversen Bewertungen

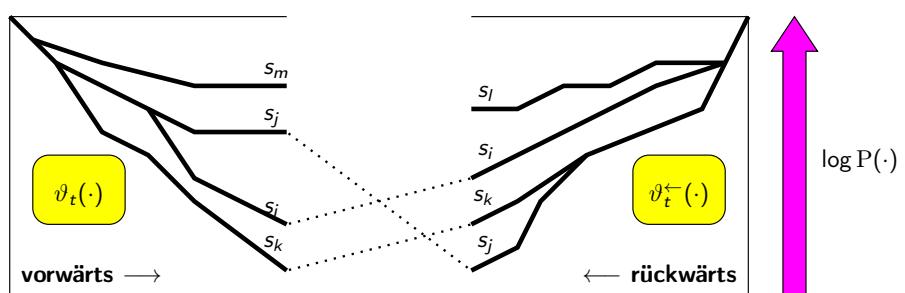
$$\vartheta_t^{\leftarrow}(i) = \max_j a_{ij} b_j(\mathbf{x}_{t+1}) \vartheta_{t+1}^{\leftarrow}(j)$$

Vorwärts-Rückwärts-Suche II

Beschleunigung bei den Rückwärtstransitionen $s_j \rightarrow s_i$:

- Wenn $i \notin \mathcal{O}_t$, so $\vartheta_t(i)$ außerhalb des Suchstrahls $\{\vartheta \mid \vartheta \geq B_0 \Gamma_t\}$.
 - Wenn $i \in \mathcal{O}_t$, so steht $\vartheta_t(i)$ zur Verfügung.
- Überprüfe die Ungleichung

$$\underbrace{\vartheta_t(i) \cdot a_{ij} \cdot b_j(x_{t+1}) \cdot \vartheta_{t+1}^\leftarrow(j)}_{P^*(X, q_t=s_i, q_{t+1}=s_j)} \geq B_0 \cdot \Gamma_T$$



Motivation

Kompilierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Suche in Wortfolgenrichtung
Graphsuche · Kellersuche

Wortschatzorganisation

Mehrphasendekodierung

Beispielaufbau

Graphsuche

Aufgabenstellung

Suche bestbewerteten Zielknoten eines gerichteten Graphen

Bewerteter gerichteter Graph $(\mathcal{K}, \mathcal{E}, d)$

- Knotenmenge $\mathcal{K} = \{k_1, k_2, \dots\}$
- Kantenmenge $\mathcal{E} \subseteq \mathcal{K} \times \mathcal{K}$
- Nichtnegative Kostenfunktion $d : \mathcal{E} \rightarrow \mathbb{R}_0^+$

Pfade, Lösungen und ihre Kosten

- Gerichteter Pfad $k = (k_1, \dots, k_m)$ falls alle $(k_i, k_{i+1}) \in \mathcal{E}$
- k Lösungspfad falls $k_1 \in \mathcal{K}_\alpha$ und $k_m \in \mathcal{K}_\omega$
- Kumulative Kosten

$$D(k) \stackrel{\text{def}}{=} \sum_{i=1}^{m-1} d(k_i, k_{i+1})$$

Heuristisch informierte geordnete Suche

(Algorithmus)

- 1 INITIALISIERUNG
Setze $\mathcal{O} = \mathcal{K}_\alpha$
- 2 AUSWAHL
Ermittle besten Knoten $k = \operatorname{argmin}_{\ell \in \mathcal{O}} \hat{f}(\ell)$
- 3 TERMINIERUNG
Wenn $k \in \mathcal{K}_\omega$ dann \rightsquigarrow ENDE
- 4 EXPANSION
Berechne $\hat{f}(k')$ für alle $(k, k') \in \mathcal{E}$
Sortiere die $\hat{f}(k')$ in die Schlange \mathcal{O} ein
- 5 ITERATION
Gehe \rightsquigarrow [2]

(zum DiagnosA)
Die „heuristische Funktion“ $\hat{f}(\cdot)$ schätzt die Erfolgschance der Expansion

Spezialfall A*-Algorithmus

Eine zulässige & effiziente heurist. inform. Graphsuche

1. Wahre Zielfunktion

$$f(k) \stackrel{\text{def}}{=} \operatorname{argmin} \{D(k) \mid k \in \mathcal{K}, k \text{ Lösung}\}$$

2. Additive Zerlegung

Weg vom Start nach k — Weg von k ins Ziel

$$f(k) = g(k) + h(k)$$

3. Dynamische Programmierung

$$\hat{g}(k) \stackrel{\text{def}}{=} \text{bislang günstigster Pfad von } \mathcal{K}_\alpha \text{ nach } k$$

4. Optimistische Restschätzung

$$\hat{h}(k) \leq h(k) \quad (\forall k \in \mathcal{K})$$

Zulässigkeit & Optimalität des A*-Algorithmus

Satz

Die geordnete Suche mit

$$\hat{f} = \hat{g} + \hat{h}$$

heißt **A*-Algorithmus** und besitzt die folgenden Eigenschaften:

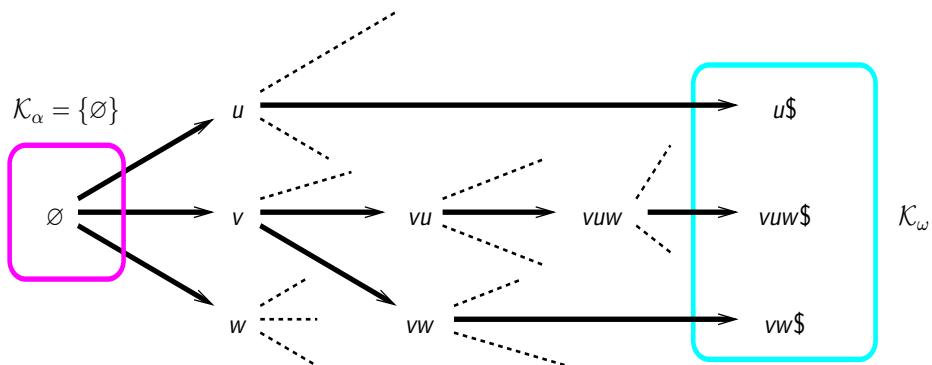
1. **Terminierung** — Algorithmus endet
2. **Monotonie** — nach Expansion von $k \in \mathcal{K}$ gilt $\hat{g}(k) = g(k)$
3. **Zulässigkeit** — die erste expandierte Lösung ist die beste
4. **Optimale Effizienz** — min. Anzahl expand. Knoten bzgl. $\hat{h}(\cdot)$
5. **Anordnung** — liefert ggf. die n besten Lösungen in Folge

Kellersuche ('stack decoding')

$$\mathcal{K} = (\mathcal{V} \cup \{\$\})^*$$

$$\mathcal{K}_\alpha = \{\emptyset\}$$

$$\mathcal{K}_\omega = \{w\$ \mid w \in \mathcal{V}^*\}$$



Kellersuche und Kostenfunktionen

- Kumulative Kosten

$$g_t(\mathbf{w}) \stackrel{\text{def}}{=} P(\mathbf{w}, \mathbf{x}_1 \dots \mathbf{x}_t) = P(\mathbf{w}) \cdot P(\mathbf{x}_1 \dots \mathbf{x}_t \mid \mathbf{w})$$

Suchgraph $\hat{=}$ Suchbaum $\Rightarrow \hat{g}_t \equiv g_t$

- Restwahrscheinlichkeit

$$h_t(\mathbf{w}) \stackrel{\text{def}}{=} \max_{\mathbf{u} \in \mathcal{V}^*} P(\mathbf{u} \mid \mathbf{w}) \cdot P(\mathbf{x}_{t+1} \dots \mathbf{x}_T \mid \mathbf{u})$$

- Lokales Bewertungsprofil

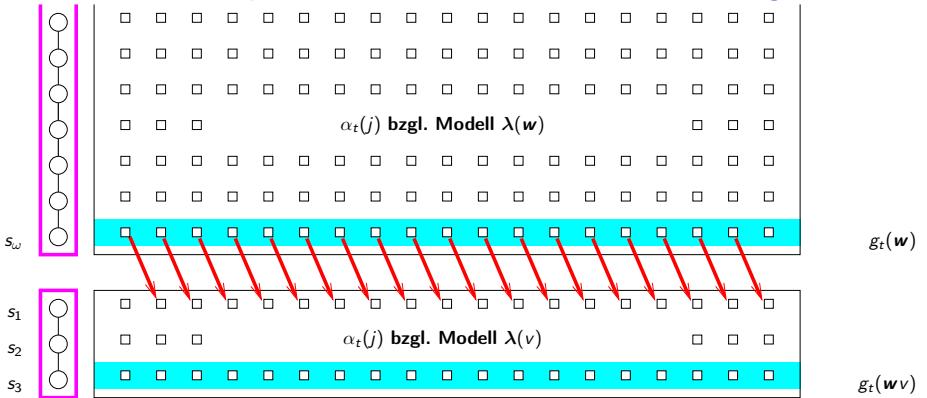
$$f_t(\mathbf{w}) = g_t(\mathbf{w}) + h_t(\mathbf{w})$$

Für $f = \max_t f_t$ gilt: $f(\mathbf{w}) = \begin{cases} \max_u P(\mathbf{w}u, \mathbf{X}) & \mathbf{w} \notin \mathcal{K}_\omega \text{ partiell} \\ P(\mathbf{w}, \mathbf{X}) & \mathbf{w} \in \mathcal{K}_\omega \text{ vollständig} \end{cases}$

- Restschätzung ('shortfall')

$$\hat{h}_t(\mathbf{w}) = \hat{h}_t = \prod_{s=t+1}^T \max_{j=1..N} b_j(\mathbf{x}_s)$$

Gestapelte Vorwärtsmatrixberechnung



$$g_t(\mathbf{w}) \stackrel{\text{def}}{=} P(\mathbf{x}_1 \dots \mathbf{x}_t | \mathbf{w})$$

$$g_t(\mathbf{wv}) \stackrel{\text{def}}{=} P(\mathbf{x}_1 \dots \mathbf{x}_t | \mathbf{wv})$$

$$\alpha_t^{v|\mathbf{w}}(1) = \begin{cases} b_1(\mathbf{x}_1) & t = 1 \\ b_1(\mathbf{x}_t) \cdot (\alpha_{t-1}^{v|\mathbf{w}}(1) \cdot a_{11} + g_{t-1}(\mathbf{w}) \cdot a_{01}) & t > 1 \end{cases}$$

Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Suche in Wortfolgenrichtung

Wortschatzorganisation

Suffixäquivalenz · Phonetischer Baum · Dendrophone

Mehrphasendekodierung

Beispielaufbau

Ökonomische Wortschatzorganisation

Aufgabenstellung

Komprimierung des HMM-Wortmodellnetzwerks
→ Reduktion des Speicher- und Berechnungsaufwandes

Vorgehensweise

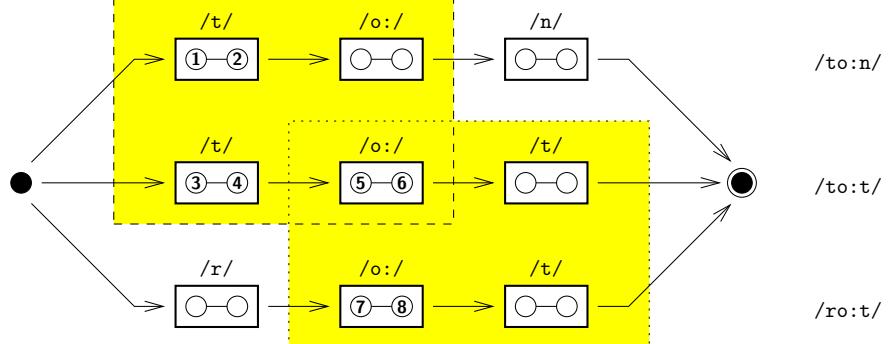
- Angriffsfläche: die *Phonmodellkopien* ($\sim 90\%$ Reduktion)
- Präfixäquivalenz**
identische α - oder ϑ -Wahrscheinlichkeiten

$$\left\{ \begin{array}{l} \text{auswerten} \\ \text{auswertet} \\ \text{ausfahre} \\ \text{ablehne} \end{array} \right\} \rightarrow a \left\{ \begin{array}{l} \text{us} \\ \left\{ \begin{array}{l} \text{werte} \\ \text{fahre} \\ \text{blehne} \end{array} \right\} \\ \left\{ \begin{array}{l} n \\ t \end{array} \right\} \end{array} \right\}$$

- Postfixäquivalenz**
vorweggenommene Siegerwortentscheidungen

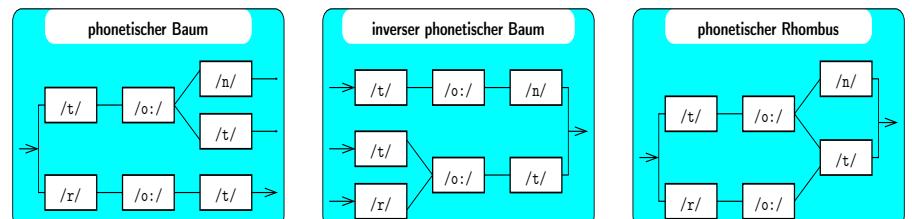
$$\left\{ \begin{array}{l} \text{abkaufe} \\ \text{einkaufe} \\ \text{Kernseife} \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \left\{ \begin{array}{l} \text{ein} \\ \text{ver} \end{array} \right\} \\ \text{kau} \\ \text{Kernsei} \end{array} \right\} fe$$

Präfixäquivalenz & Postfixäquivalenz



$$\begin{aligned} P(\mathbf{X}, \mathbf{q}^*) &= \max \left\{ \max_{\mathbf{q}} P(\mathbf{X}, \mathbf{q} | /t/), \max_{\mathbf{q}'} P(\mathbf{X}, \mathbf{q}' | /o:/) \right\} \\ &= \max_t \left(\max \left\{ P^*(\mathbf{x}_1 \dots \mathbf{x}_t | /t/), P^*(\mathbf{x}_1 \dots \mathbf{x}_t | /r/) \right\} \cdot P^*(\mathbf{x}_{t+1} \dots \mathbf{x}_T | /o:/) \right) \end{aligned}$$

Phonetischer Lexikonbaum & CD-PLUs



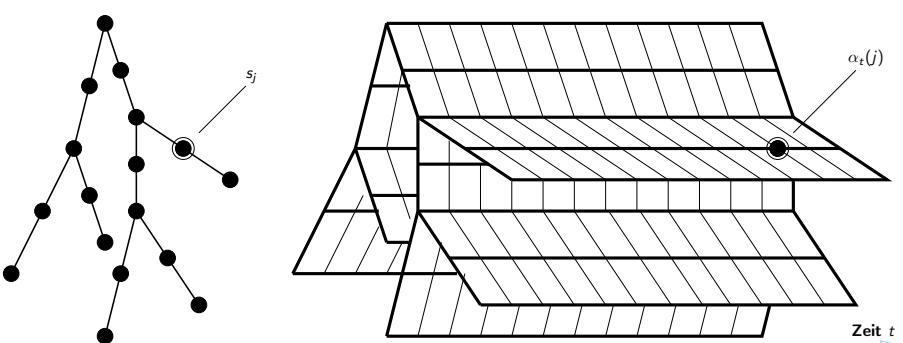
Triphone

„Ton“ \rightsquigarrow /t/o: t/o:/n o:/n/
 „tot“ \rightsquigarrow /t/o: t/o:/t o:/t/
 „rot“ \rightsquigarrow /r/o: r/o:/t o:/t/

Dendophone

„Ton“ \rightsquigarrow /t/ t/o:/ to:/n/
 „tot“ \rightsquigarrow /t/ t/o:/ to:/t/
 „rot“ \rightsquigarrow /r/ r/o:/ ro:/t/

Baumstrukturierter Viterbi-Suchraum



Speicherplatztopologie der Vorwärtswahrscheinlichkeiten eines phonetischen Lexikonbaums

Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Suche in Wortfolgenrichtung

Wortschatzorganisation

Mehrphasendekodierung

Schrittweise Verfeinerung $\cdot n$ beste Wortketten

Beispielaufbau

Mehrphasendekodierung

Problem

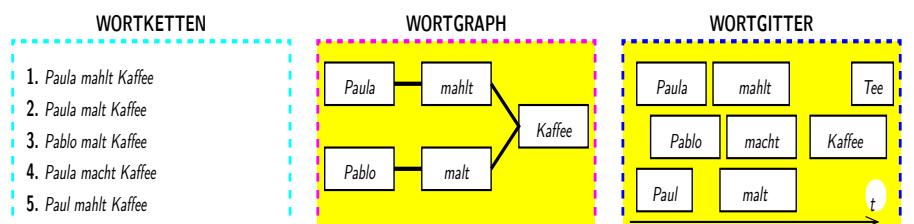
Nicht alle Grammatikformalismen sind HMM-Netzwerk-kompatibel:

- Pentagramm-Sprachmodelle $|\text{Zustandsraum}| = L^4$
- Kreditkartennummer/Kontrollbedingung „127 teilt k“
- Spielekommandos „Springer schlägt Dame auf c3“

Lösung = schrittweise Verfeinerung

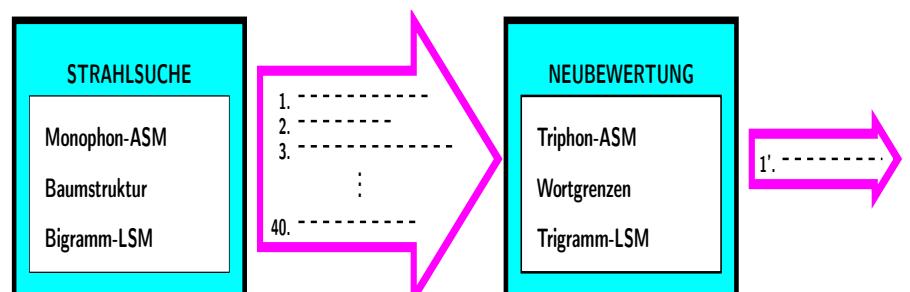
- 1 **Schnelle Suche**
zur Erzeugung konkurrierender Zwischenlösungen
- 2 **Sequentielles Ausfiltern**
vermöge akustischer & grammatischer Neubewertung

Zwischenlösungsrepräsentationen für die Satzerkennung



- **Wortketten**
aufzählende Wortfolgeinformation · hochredundant
- **Wortgraphen**
explizite Wortfolgeinformation · hochökonomisch
- **Wortgitter**
implizite Wortfolgeinformation durch Zeitstempel · übergeneralisierend

Die Systemarchitektur BYBLOS



Die Neubewertung von Wortketten unterliegt keinerlei Einschränkungen hinsichtlich der Struktur akustischer & grammatischer Modelle!

n-best Algorithmen

Näherungsweise Berechnung der n besten Wortketten mit Varianten des Viterbi-Algorithmus

- **Zustandsbezogener NBVA** (Bayer '86)
hält in jedem Gitterpunkt (t, j) die n besten Kandidaten in bewertungssortierter Liste $\mathcal{D}_t(j)$ und berechnet

$$\vartheta_t^{(k)}(j) = \max^{(k)} \left\{ \vartheta_{t-1}^{(l)}(i) \cdot a_{ij} \cdot b_j(x_t) \mid 1 \leq i \leq N, 1 \leq l \leq n \right\}$$

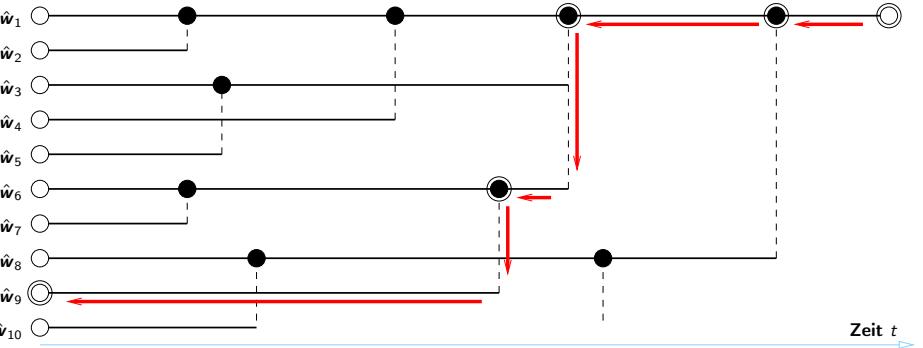
$\mathcal{D}_t(j)$

- **Satzbezogener NBVA** (Steinbiss '89)
rekombiniert konkurrierende Kandidaten für gleiche Wortfolgen

- **Gitterbezogener NBVA** (Marino&Monte '89)
keine Listen im Wortinneren, nur das **dichte Wortgitter**

$$\{(w, \vartheta_t(w), \tau_t(w)) \mid t = 1, \dots, T, w \in \mathcal{W}\}$$

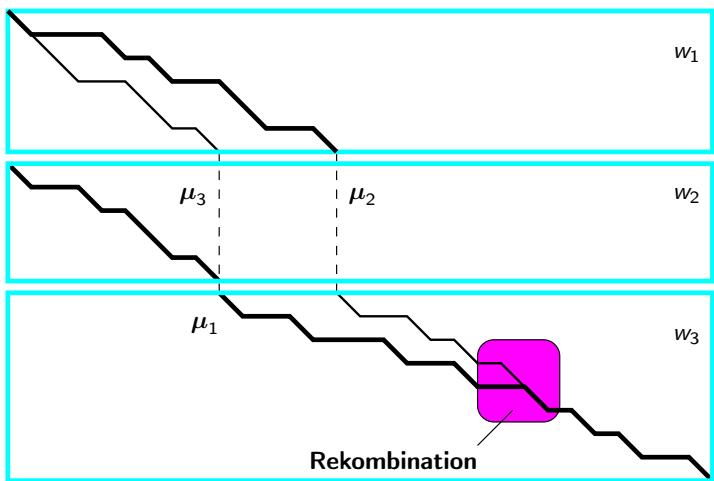
Kettenrekonstruktion aus dem Wortgitter



Rekonstruktion der 10 bestbewerteten Wortketten aus dem *dichten Wortgitter* mit der Rekursion

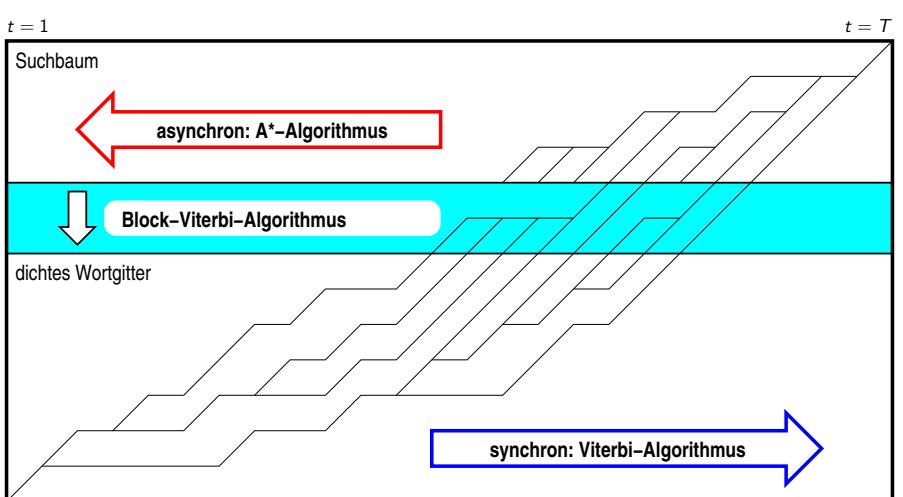
$$P^*(t_0 \dots t_i, w_1 \dots w_i) = P^*(t_0 \dots t_{i-1}, w_1 \dots w_{i-1}) \cdot \frac{\vartheta_{t_i}(w_i)}{\max_v \vartheta_{t_{i-1}}(v)}$$

Der gitterbezogene n -best-Viterbi ist suboptimal!



Im Innern des Wortes w_3 werden die Zustandsfolgen μ_1 und μ_2 rekombiniert.
Die Wortfolge $w_1 w_3$ wird fortan durch Pfad μ_3 vertreten, auch wenn
 $P(X, \mu_2) > P(X, \mu_3)$ gelten sollte!

Der Tree-Trellis-Algorithmus (Soong)



Motivation

Komplizierte HMM-Netzwerkstrukturen

Die wahrscheinlichste Wortsegmentierung

Suche in Zeitrichtung

Suche in Wortfolgenrichtung

Wortschatzorganisation

Mehrphasendekodierung

Beispielaufbau

An Stelle einer Zusammenfassung

EXEMPLARISCHE BERECHNUNGSFOLGE ZUR DEKODIERUNG EINES GESPROCHENEN SATZES

Grammatikgesteuerte Spracherkennung

- 1 **Vorverarbeitung des Eingabesignals**
Diskretisierung – Merkmalberechnung – Vektorquantisierung
- 2 **Strahlgesteuerter Viterbi-Algorithmus vorwärts**
Phonemischer Baum, Monophone, Bigramm-Grammatik
- 3 **Wortgitterberechnung rückwärts**
Inverser Phonemischer Baum, Dendrophone, Bigramm-Grammatik
- 4 **Konstruktion der 100 besten Wortketten**
A*-Algorithmus oder Dynamische Programmierung
- 5 **Umbewerten & Umsortieren der Wortketten**
HMM's mit wortgrenzenübergreifenden Polyphonen, Polygramm-Grammatik

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil IX

Maschinelle Schrifterkennung

Schriftdaten
oooooooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Schriftdaten
●oooooooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Hardware zur digitalen Schriftdatenerfassung I

Schriftdaten

Maschinelle Schriftdatenerfassung

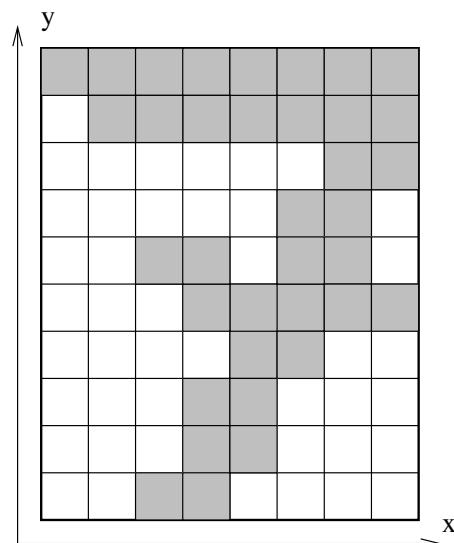
Schriftdatenvariabilität

Segmentierung von Wörtern in Zeichen

Anwendungsszenarien

Explizit segmentierende Systeme

Implizit segmentierende Systeme



Der SCANNER
wandelt eine Papiervorlage in eine
Bildmatrix

$$[f(x, y)]_{x=1..N; y=1..M}$$

Vorteile „off-line“
Verarbeitung bestehender
Dokumente
hohe Lesegeschwindigkeit

Schriftdaten
oooooooo●oooooooooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Erfassungsbedingte Variabilität

Störungen

Papierunreinheiten
Stempel
Hilfslinien
Zeichenboxen

Deformationen durch Scanner

Löcher
Unterbrechungen
isolierte Punkte

Deformationen durch Digitizer

Strichanfangdefekt (*pen-down*)
Häkchen am Strichende (*pen-up*)

Schriftdaten
oooooooo●oooooooooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Statische Variabilität

Geometrische Transformationen

Größe, Form, Neigung, Rotation, Liniendicke

Topologische Kategorien

Kontinuum *Spitze* → *Schlinge*



Segmentale Kategorien

Kontinuum *Spitze* → *Buckel*



Schriftdaten
oooooooo●oooooooooooo

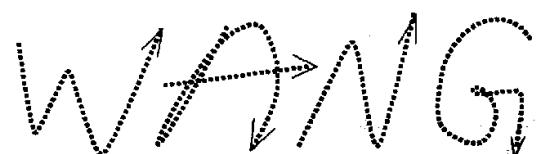
Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Dynamische Variabilität

- Strichfolge
- Strichzahl
- Schreibtempo
- Nachspurung



Fragestellung

In welchem Verarbeitungsschritt wird das jeweilige Phänomen behandelt?
(*Merkmalgewinnung, Segmentierung, Matching, Modellierung*)

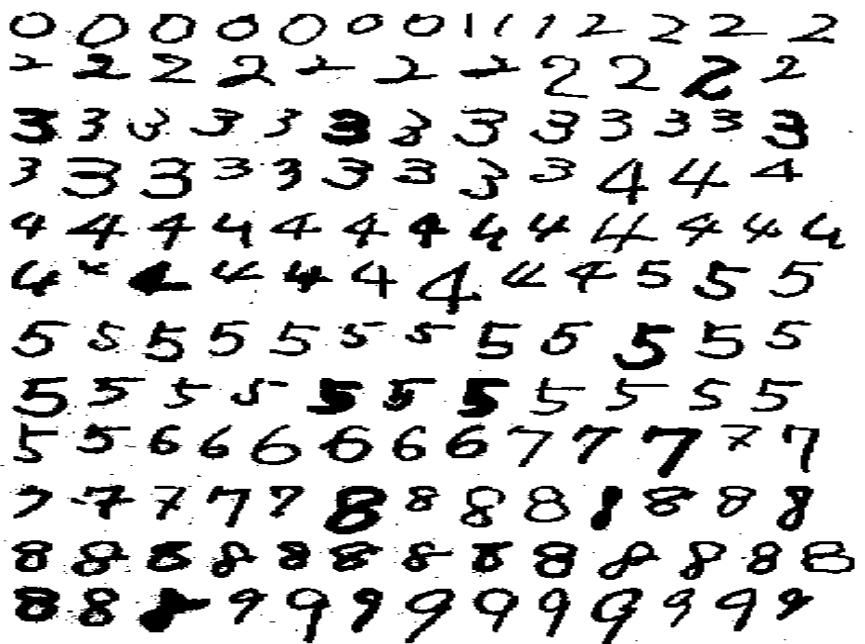
Schriftdaten
oooooooo●oooooooooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
oooooooooooo

Beispiel: Ziffernrealisierungen



Schriftdaten
oooooooooooo●ooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Schreiberverfassung

The party begins.

1 I can drive when I drink.
Two drinks later.

I can drive when I drink

After four drinks.

2 I can drive when I drink
After five drinks.

3 I can drive when I drink
Seven drinks in all.

I can drive when I drink

Schriftdaten
oooooooooooo●ooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Selbst Ziffern zu segmentieren ist aussichtslos!

65473 60198 68544
70065 70117 19032 98720
27260 618208 18559
74136 19133 63101
20878 60521 38004
48640-2398 20902 14882

Überschneidung von Ziffernbestandteilen
Überlappung von Ziffernaufenthaltsbereichen
Übergreifende Unterstreichungen & Dekorierungen

Schriftdaten
oooooooooooo●ooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Darbietungsform

Kursivschrift

jedes Einzelwort ist zusammenhängend

Handschrift

Blockschrift in Zeichenboxen

Handschrift

Blockschrift gesperrt

topologisch getrennt

Handschrift

Blockschrift verklebt

Berührung; Überlappung

Handschrift

Kursivschrift durchbrochen

Handschrift

Block- und Kursivschrift

gemischt

U.S. Mail

Schriftdaten
oooooooooooo●ooooo

Anwendungsszenarien
oooooooo

Explizit segmentierend
oooooooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Zeichensegmentierung

Die Segmentierung von 'on-line'-Schriftdaten ist einfacher?

Mehrstrichhaltige Zeichen

d = c + l

Verzögerte Ligaturen & Diakritika

tt vier

Geteilte Strichverantwortung

Ha

50

Schriftdaten
oooooooooooooooAnwendungsszenarien
oooooooExplizit segmentierend
ooooooooooooooooooooImplizit segmentierend
ooooooooooooooo

Arabische Schriftzeichen

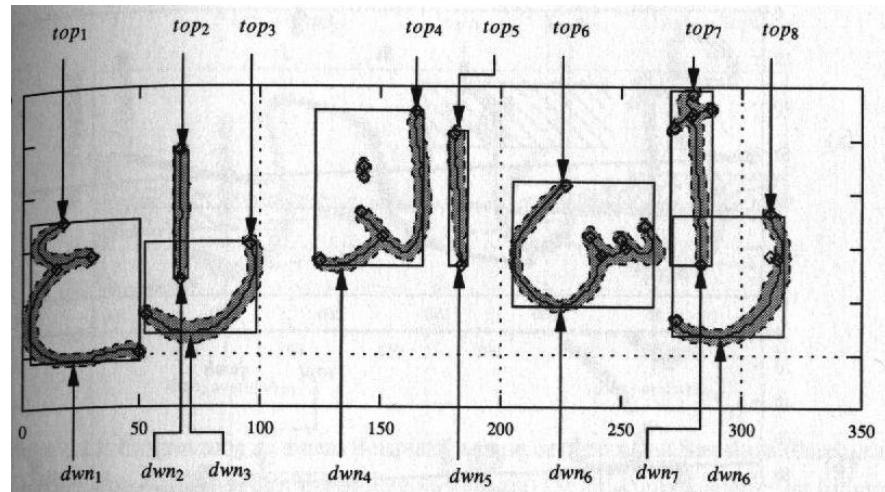
تونس القباشة الأصلية
تونس القباشة الأصلية
تونس القباشة الأصلية

Besonderheiten arabischer Schrift

- keine Majuskel/Minuskel-Unterscheidung
- Punkte oben/unten/mittig (≤ 3) als Unterscheidungsmerkmal
- Diatkritische Ergänzungen: *alif, madda, hamza, ta'marbuta*
- Schreibrichtung rechts→links
- PAW (*pieces of arabic words*) wegen Basisligatur & Ausnahmen
- 4 positionsabhängige Buchstabenausprägungen
- Überlappungen auf der Horizontalen
- Ungewöhnliche Größenunterschiede (Breite und Höhe)
- Vokale optional & als Diakritika realisiert
- Buchstabenverdopplung mittels *chadda*

Schriftdaten
oooooooooooooooAnwendungsszenarien
oooooooExplizit segmentierend
ooooooooooooooooooooImplizit segmentierend
ooooooooooooooo

Segmentierung, Umschreibung, Typographie ?

Schriftdaten
oooooooooooooooAnwendungsszenarien
oooooooExplizit segmentierend
ooooooooooooooooooooImplizit segmentierend
ooooooooooooooo

Arabisches Alphabet

Buchst.	Isoliert	Ende	Mitte	Beginn	Buchst.	Isoliert	Ende	Mitte	Beginn
Alif	ا	ل			Dhad	ض	ض	ض	ض
Ba	ب	ب	ب	ب	Taa	ط	ط	ط	ط
Ta	ت	ت	ت	ت	Dha	ظ	ظ	ظ	ظ
Tha	ث	ث	ث	ث	Ayn	ع	ع	ع	ع
Jim	ج	ج	ج	ج	Ghayn	غ	غ	غ	غ
Ha	ح	ح	ح	ح	Fa	ف	ف	ف	ف
Kha	خ	خ	خ	خ	Qaf	ق	ق	ق	ق
Dal	د	د			Kaf	ك	ك	ك	ك
The	ذ	ذ			Lam	ل	ل	ل	ل
Ra	ر	ر			Mim	م	م	م	م
Zai	ز	ز			Nun	ن	ن	ن	ن
Sin	س	س	س	س	He	ه	ه	ه	ه
Chin	ش	ش	ش	ش	Waw	و	و	و	و
Sad	ص	ص	ص	ص	Ya	ي	ي	ي	ي

Schriftdaten
ooooooooooooooooooooAnwendungsszenarien
oooooooExplizit segmentierend
ooooooooooooooooooooImplizit segmentierend
ooooooooooooooo

Schriftdaten

Anwendungsszenarien

- Handschrift
- Handschrift und Druckschrift
- Druckschrift

Explizit segmentierende Systeme

Implizit segmentierende Systeme

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
●○○○○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooo

Anwendungen optischer Schriftzeichenerfassung - OCR

Postanschriftenleser

Automatische Poststücksortierung
Adressfeld = PLZ, Ort, Straße, (Bundesstaat)

Formularleser

Bank-, Zoll- und Versicherungsformulare

Ziffern — Geldbetrag, Konto, BLZ, Datum, KFZ

Block — Vor/Zuname, Institut, Warenkennung, ...

Kursiv — Betragstext, Unterschrift

Texterfassungssysteme

文字言忍言能技术

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
○○○○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooo

Elektronische Tinte auf intelligentem Papier

Papier-und-Bleistift Kommunikation

- Datenerfassung in mobilen Einsatzbereichen
- wenn eine Tastatur zu groß/laut wäre
- zur hybrid textuell-graphischen Eingabe
- zur Textkorrektur und -annotation
- für die Zugangskontrolle

menschliche Schreibgeschwindigkeit

- 1.5–2.5 Zeichen/sek
(*alphanumerisch, Blockschrift*)
2.5–5.0 Zeichen/sek
(*alphanumerisch, Kursivschrift*)
0.2–2.0 Zeichen/sek
(*chinesisch*)

PEN-Computer, Notepad

Kleinrechner in Notizblockgröße
Dateneingabe per Stift auf PAD
(*pen-and-display*)
keine Tastatur

PenRight! / PalmPrint

multilingual ohne Wörterbuch
benutzerdefinierte Zeichen
Unterschrift statt Paßwort
Buchstaben, Ziffern und (*Zeige-)*Gesten
lernfähig, mit Gastmodus

Intelligente Wandtafeln

Anschrieb von CCD-Kamera erfaßt
Protokollierung von Vorträgen
Transliteration (*LATEX*)
On-line Grafikeinblendung
On-line Grafikerstellung
On-line Formelauswertung

Weltläufige Kameras

Liest und übersetzt Dokumente
Liest und verortet Straßenschilder

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
○○○○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooo

U.S. Postal Service

Kuvertaufdruck überschrieben

Allen Page
Route 1 Box 877
Columbia, KY 42728

Hilfslinien berührt & Bildrotation

1. Name + Vorname
PUBLICATION
PO BOX 55129
ADDRESS
Gould Co 80329
STATE
ZIP CODE

Umrandungsbox verlassen

JUICE ROSS + COMPANY
MR JONATHAN R. BOND
ME MARITIME PLAZA 15TH FLOOR
SAN FRANCISCO, CALIF 94111

Hilfslinien überschrieben

To: USA
M. VAZIRI 1015
MARYLANE OAKDALE
LA, 71463

Poststempel überlagert

M. Margaret
2120 E DuPont Dr
Bella WV 25401

Briefmarke & Poststempel

M. Wilma
PO. Box 757
Berkham, KY 40004

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
○○○●○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooo

Überweisungsbelege

Überweisung/Zahlschein

Sparkasse Bielefeld

(Name und Sitz des überweisenden Kreditinstituts)

4 8 0 5 0 1 6 1

Bankleitzahl

Den Vordruck bitte nicht beschädigen, knicken, bestempeln oder beschützen.

Begünstigter: Name, Vorname/Firma (max. 27 Stellen)

MARKOV, HIDDEN UND PARTNER

Konto-Nr. des Begünstigten

2718281828

Bankleitzahl

31415926

Kreditinstitut des Begünstigten

STATISTISCHE ZENTRALBANK

Kunden-Referenznummer - Verwendungszweck, gef. Name und Anschrift des Überweisenden - (nur für Begünstigten)

ASSORTIMENT NR. 4098 DIV.

Noch Verwendungszweck (max. 2 Zeilen à 27 Stellen)

DR. FINK, GERNOT A.

Kontoinhaber/Einzahler: Name, Vorname/Firma, Ort (max. 27 Stellen, keine Straßen- oder Postfachangaben)

DR. FINK, GERNOT A.

Konto-Nr. des Kontoinhabers

43656532

18

Schreibmaschine: normale Schreibweise
Handschreibmaschine: handschriftliche Schreibweise in GROSSEBUCHSTABEN.
Bitte Ziffern in Ziffern verwenden.

1.4.2002 Gernot Fink

Datum, Unterschrift

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
oooo●○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
oooooooooooo

Kontonummern auf Überweisungsbelegen

43584542	40311050	40470369	44036345
44500633-	40500476-	40501787	44036345
44036345-	44259149	49632424	(47234965)
44215277	43561936	36310215	46167276
46433120	40630748	49614568-	40590301
49240586	45078044-	45078044-	45078044-
41063682	49291025	42240356	42240356

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
oooo●○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
oooooooooooo

Verkehrsüberwachung — Gefahrgüter



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
oooo●○○

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
oooooooooooo

Kraftfahrzeug-Kennzeichen

APPG 51	LÜK 6862	VISC 907
MBHX 537	ZIAN 591	PLSAT 342
OSMR 6791	BGJ 9144	BTFG 818
FRND 269	CBPM 69	BEH 4707
SHKH 831	DDDN 5680	BNM 6144
PLVJ 25	BTF S 764	WEYM 38
BZPW 740	GRZCB 51	EF 34047
PF P 6110	APJB 47	KS HK 175
WMSZH 42	SHKH 600	HH PZ 1204
EFHT 152	APKR 13	33 JJ KS 1

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
oooooooooooo

Schriftdaten

Anwendungsszenarien

Explizit segmentierende Systeme

Wortvereinzelung
Segmentierung & Merkmalgewinnung
Matching & Fehlernachbearbeitung

Implizit segmentierende Systeme

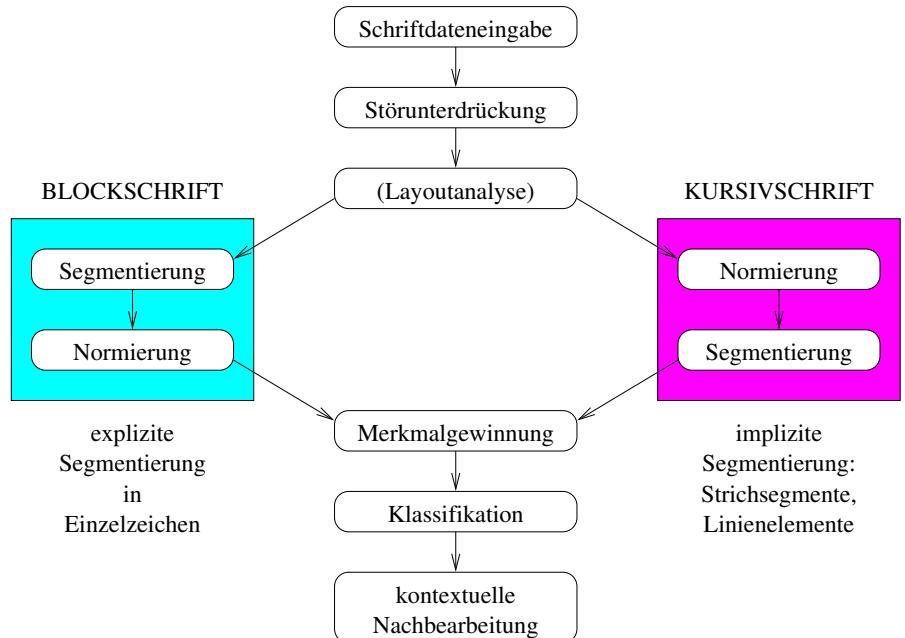
Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Sequentielle OCR-Architektur



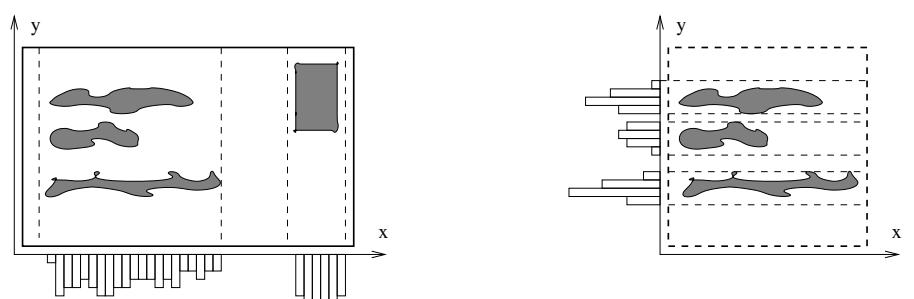
Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Wiederholte Bildprojektion



Iteriertes Auswerten vertikaler & horizontaler Bildprojektionen

- Spalten
- Textblöcke
- Zeilen
- Wörter

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

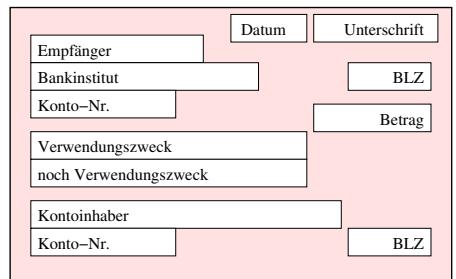
Explizit segmentierend
●ooooooooooooooo

Implizit segmentierend
ooooooooooooooo

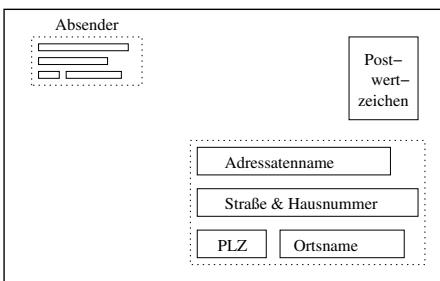
Layout-Analyse

Zerlegung

eines Dokuments in seine logischen Bestandteile



Überweisungsbeleg ↗



↳ Briefkuvert

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
○●ooooooooooooooo

Implizit segmentierend
ooooooooooooooo

Normierung

(Elimination von Variabilitäten im Vorfeld der Klassifikation)

Rotation

der Vorlage oder der Schriftzeile um einen Winkel α

twele

Neigung

der vertikalen Schriftkomponenten um einen Winkel α

twele

Größe

relative Ausdehnung r_x, r_y von Zeichen/Wörtern in x- und y-Richtung

twele

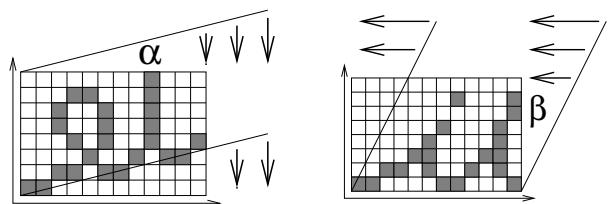
Liniendicke

des gescannten Schriftbildes übersteigt i.a. einen Pixel

twele

Koordinatentransformation

- Bestimmung der Normierungsfaktoren α, β, r_x, r_y
- Transformation der Bildkoordinaten $\mathfrak{T} : (x, y) \mapsto (x', y')$



Rotation

Vertikale Scherungsoperation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x & \\ y - x \cdot \tan \alpha & \end{bmatrix}$$

Neigung

Horizontale Scherungsoperation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x - y \cdot \cot \beta & \\ y & \end{bmatrix}$$

Größe

Anisotrope Skalierungsoperation

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} x/r_x \\ y/r_y \end{bmatrix}$$

Typographische Begrenzung

Vertikale Schriftbereiche:

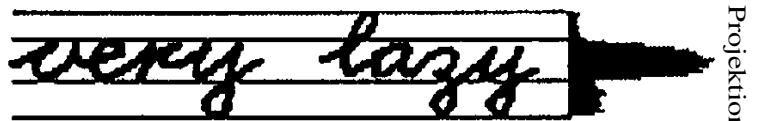
- Oberlängenbereich
- Schriftkorpus oder -basis
- Unterlängenbereich

Oberlinie

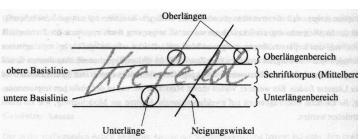
Mittellinie

Grundlinie

Unterlinie



Sind die vier Begrenzungslinien **parallele Geraden**, so genügt zur Detektion eine Vertikalprojektion.



Schriftaufrichtung

Originalbild mit geneigter Schrift

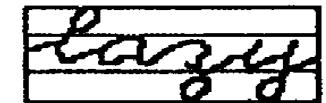
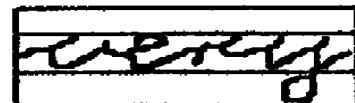


Akkumulatorebene mit den Punktdichten (x, α)

aufgerichtetes Schriftbild

Linienverdünnung

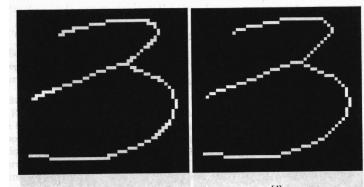
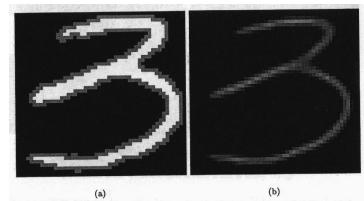
(auf die Breite eines Pixels)



Skelettierung mit Gaborfiltertechnik

Die Filterausgabe ergibt eine Bewertung aller Bildpunkte hinsichtlich ihrer **Mittelachseneigenschaft**.

Schwellwertbildung liefert schon fast ein Linienmuster.



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

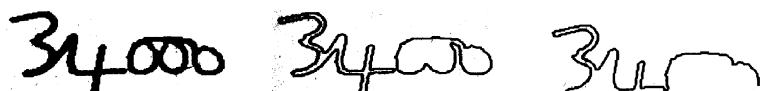
Explizit segmentierend
oooooooo●oooooooo

Implizit segmentierend
ooooooooooooooo

Segmentierung in Einzelzeichen — Blockschrift

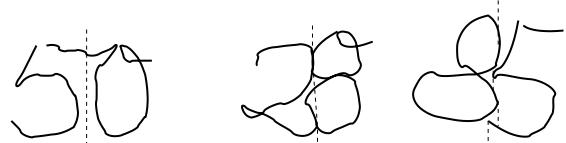
Blockschrift

- Nullstellen der vertikalen Bildprojektion
- Analyse zusammenhängender Gebiete



Berührende oder überlappende Blockschrift

- relative Minima der vertikalen Projektion
- Verbinden der lokalen Extrema der oberen & unteren Wortkontur
- objektrandgesteuerter Abstieg von lokalen Minima der oberen Wortkontur



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

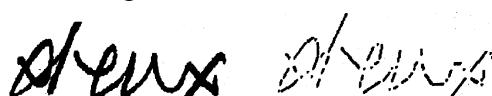
Explizit segmentierend
oooooooo●oooooooo

Implizit segmentierend
ooooooooooooooo

Segmentierungstechniken I

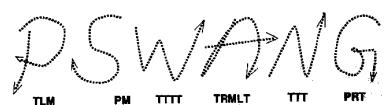
Geometrische Segmentierung in Striche

- lokale x- und y-Extrema
- Steigungsdiskontinuitäten
- Wendepunkte
- Krümmungsmaxima



Segmentierung in Formelemente

reguläre/singuläre Ereignisse · PDL (*picture description language*)



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

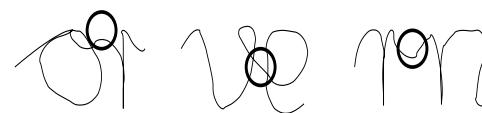
Explizit segmentierend
oooooooo●oooooooo

Implizit segmentierend
ooooooooooooooo

Segmentierung in Einzelzeichen — Kursivschrift

Probleme

- Untere Ligaturen sind zur Zeichensegmentierung unzureichend!
- Striche überspannen Buchstabenfolgen!
- Striche unterteilen Einzelzeichen!



'o', 'v' und 'r' bilden obere Ligaturen



'u' und 'w' beinhalten untere Ligaturen

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooo●oooooooo

Implizit segmentierend
ooooooooooooooo

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooo●oooooooo

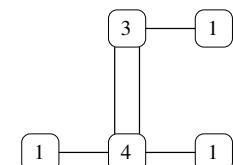
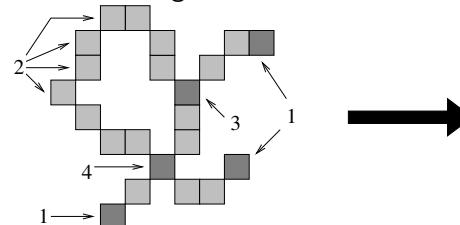
Implizit segmentierend
ooooooooooooooo

Segmentierungstechniken II

Topologisch orientierte Liniensegmentierung

Anzahl Nachbarpixel:

- 1 = Linienende
- 2 = Linieninneres
- 3 = Gabelung
- 4 = Kreuzung



Dynamische Strichsegmentierung

- kleinste motorische Einheiten
- Zyklus „Beschleunigung–Tempogipfel–Verlangsamung“

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooo●oooo

Implizit segmentierend
oooooooooooo

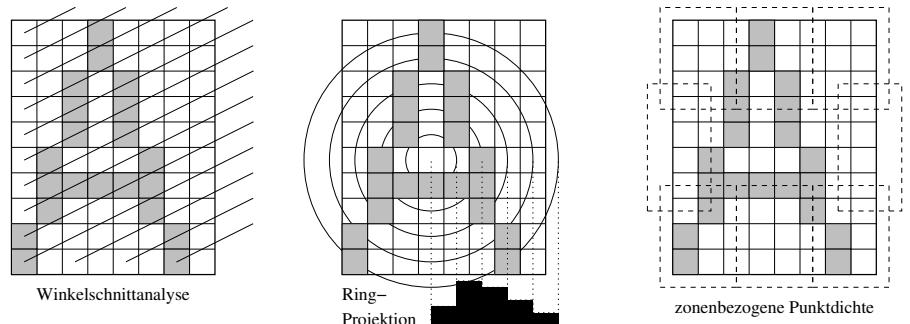
Einzelzeichenmerkmale

Repräsentation des Schriftzeichens durch einen Merkmalvektor

$$\mathbf{x} = (x_1, \dots, x_D)^\top \in \mathbb{R}^D$$

Beispiele für Merkmale:

$f(x, y)$ selbst · 2D-FFT · PCA · zentrale Momente · (siehe ⤵)



Schriftdaten
ooooooooooooooo

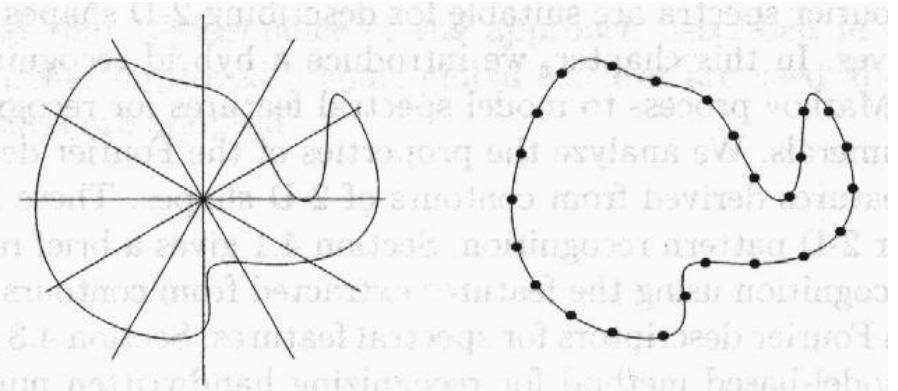
Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooo●oooo

Implizit segmentierend
oooooooooooo

Polare und spektrale Merkmale

- Betragsspektrum der Konturpunktfolge $\in \mathbb{C}^T$
- Polarkoordinaten der Konturpunkte bzgl. Objektschwerpunkt



Schriftdaten
ooooooooooooooo

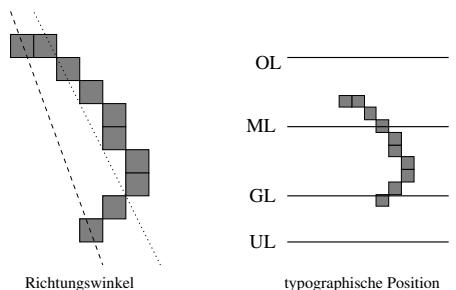
Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooo●oooo

Implizit segmentierend
oooooooooooo

Segmentbezogene Merkmale

- **Länge** — Anzahl der Segmentbildpunkte
- **Referenzwinkel** zwischen x-Achse und Sehne oder Regressionsgerade
- **Mittlere Krümmung** $|\kappa(t)|^2 = (d^2/dt^2)^2 + (d^2y/dt^2)^2$
- **Dynamische Merkmale** — mittleres Schreibtempo und max. Beschleunigung
- **Typographische Position** bzgl. Grund-, Mittel-, Unter- und Oberlinie
- **Kettencode** — grobe Quantisierung der möglichen Referenzwinkel



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

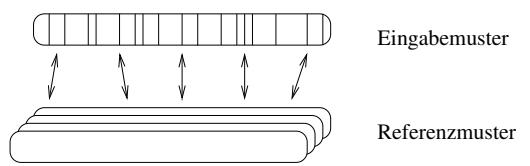
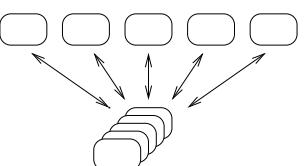
Explizit segmentierend
oooooooooooo●ooo

Implizit segmentierend
oooooooooooo

Zeichen- und Wortklassifikation

Analytische Methode

{Schriftzeichen **explizit** segmentiert
ein zeichenbezogener Merkmalvektor} → numerische Klassifikation



Syntaktische Methode

{Kette, Baum oder Graph von Segmenten
(Zeichenebene)} → Zeichenmatching
{Kette, Baum oder Graph von Segmenten
(Wortebene)} → Wortmatching

Schriftdaten
ooooooooooooooo

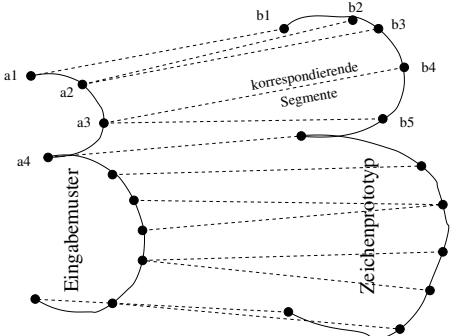
Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooooo●○

Implizit segmentierend
ooooooooooooo

Elastischer Mustervergleich

Vergleich einer parametrisierten Strichsegmentfolge mit einem Zeichen- oder Wortprototypen:



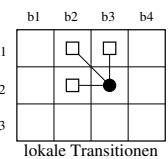
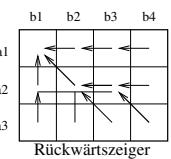
DTW-Algorithmus zur dynamischen Skalenverzerrung

	b1	b2	b3	b4
a1	1	4	5	8
a2	4	3	2	7
a3	7	4	9	0

lokale Distanzen

	b1	b2	b3	b4
a1	1	5	10	18
a2	5	4	6	13
a3	12	8	13	6

kumulative Distanzen



Schriftdaten

Anwendungsszenarien

Explizit segmentierende Systeme

Implizit segmentierende Systeme

Bayesregel & Systemarchitektur

Serialisierung in ein Longitudinalmuster

Hidden-Markov- und andere Wahrscheinlichkeitsmodelle

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
oooooooooooooo●●

Implizit segmentierend
ooooooooooooo

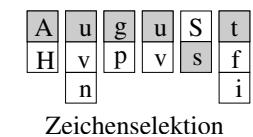
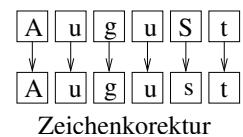
Kontextuelle Nachverarbeitung

Motivation

- Leicht verwechselbare Zeichen
- Distribution von Majuskeln/Minuskeln und Buchstaben/Ziffern
- Buchstabenfolgen ergeben korrekte Wörter !
- Wortfolgen ergeben domänenpezifisch sinnvolle Ausdrücke
- Testpassagen sind syntaktisch & semantisch wohlgeformt
- mittlere Worthäufigkeiten

Klasse AugSt Tel.: 85-78T3 Dienstag DM 2S0.-
97058 Erlangen

Nachverarbeitung



Schriftdaten
ooooooooooooooo

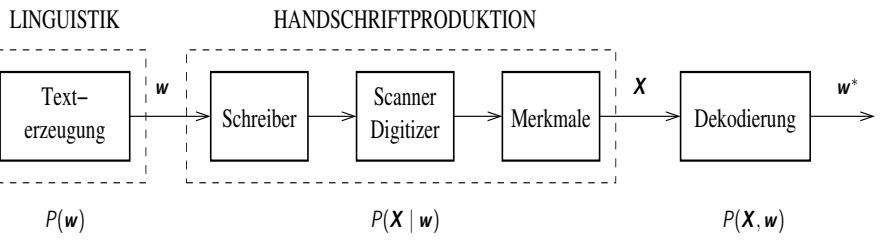
Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
●ooooooooooooo

Fundamentalformel der (Hand-)Schrifterkennung

LINGUISTIK



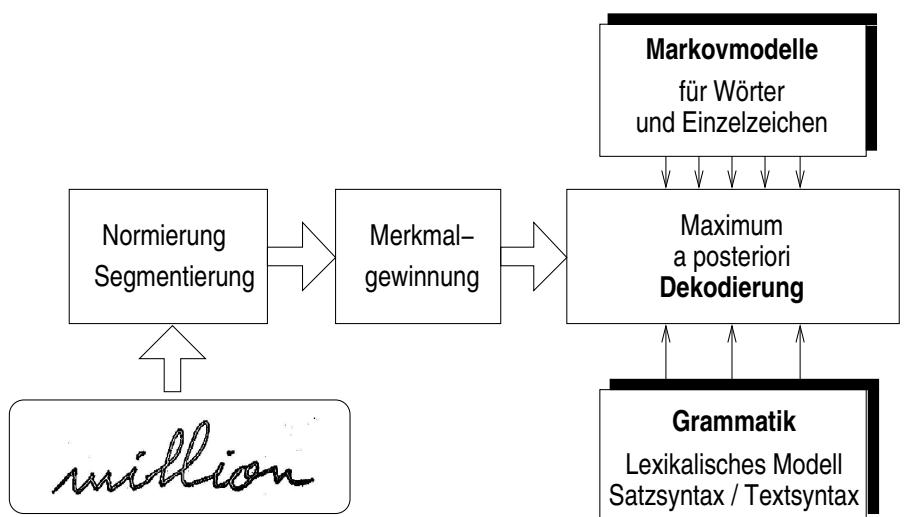
Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
o●oooooooooooo

Architektur eines HSE-Systems



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

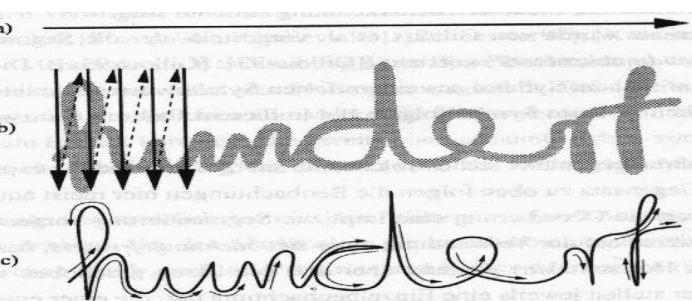
Implizit segmentierend
o●oooooooooooo

Serialisierung des Schriftbildes

Transformation eines Rasterbildes → Sequenz von Merkmalvektoren für das HMM

Repräsentation des Schriftzuges als Vektorsequenz

- Überlappende Zerlegung in schmale Bildspalten
- Mäandernde 2D-Traversierung des Schriftbildes
- Geometrie- oder produktionsorientierte Schriftkonturverfolgung



Schriftdaten
ooooooooooooooo

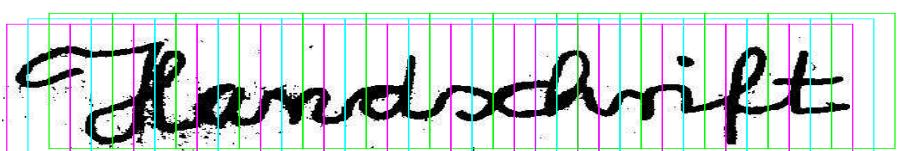
Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

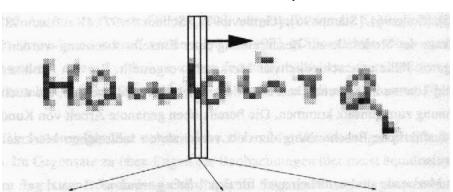
Implizit segmentierend
o●oooooooooooo

Überlappende Abtastung des Schriftzuges

Achtung! — Widerspruch zur zweiten Markoveigenschaft des HMM!



- Fensterbreite**
Größenordnung einer durchschnittlichen Zeichenbreite
- Fortschaltung**
etwa 2–5 Fenster/Zeichen



Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

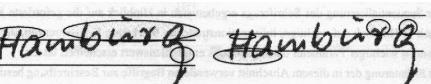
Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
o●oooooooooooo

Problematik der Fensterbildung

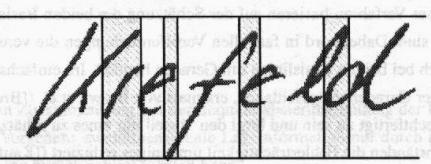
Vertikalausrichtung

Zeichenkorpus, Oberlänge oder Unterlänge befinden sich außerhalb ihrer korrekten typographischen Region



Horizontalausrichtung

Komponenten benachbarter Zeichen teilen sich dieselbe Bildspalte (Schriftneigung)



Schriftdaten
ooooooooooooooo

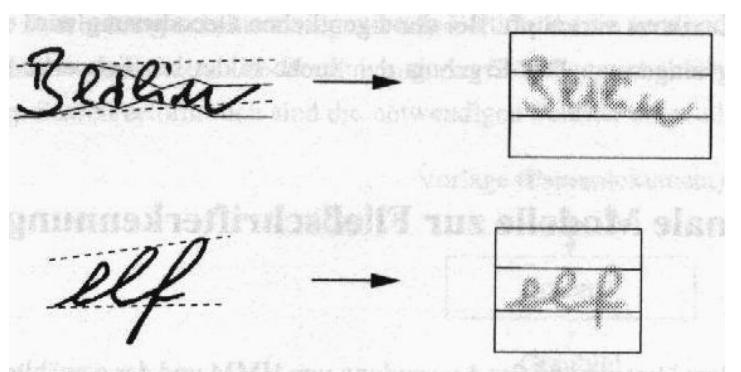
Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooo●ooooo

Nichtlineare Modelle der typographischen Struktur

Basislinie · Mittellinie · Oberlinie · Unterlinie



- Konische obere/untere Basis- und Begrenzungslinien
- Wellenförmige typographische Begrenzungen

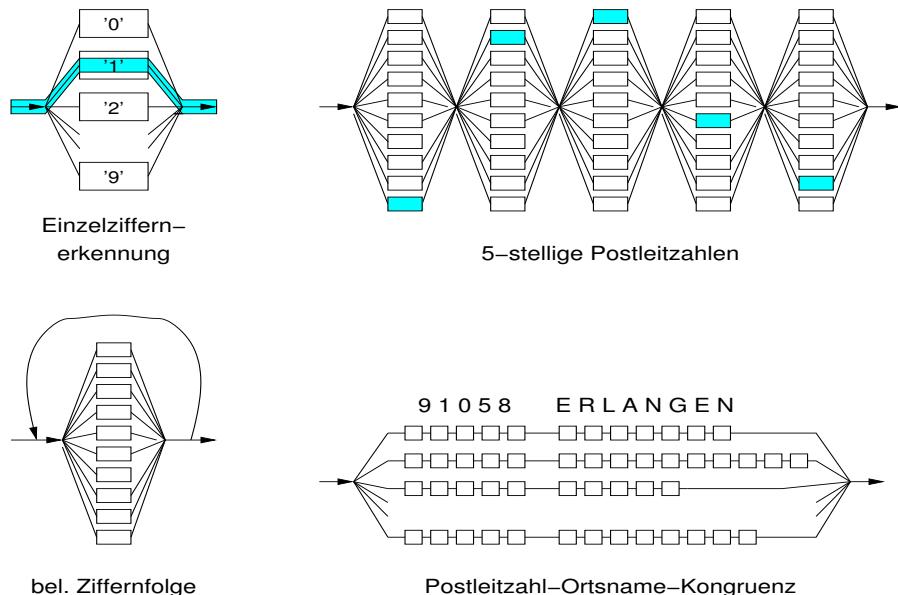
Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooo●ooooo

Dekodierung von Ort und Postleitzahl



Schriftdaten
ooooooooooooooo

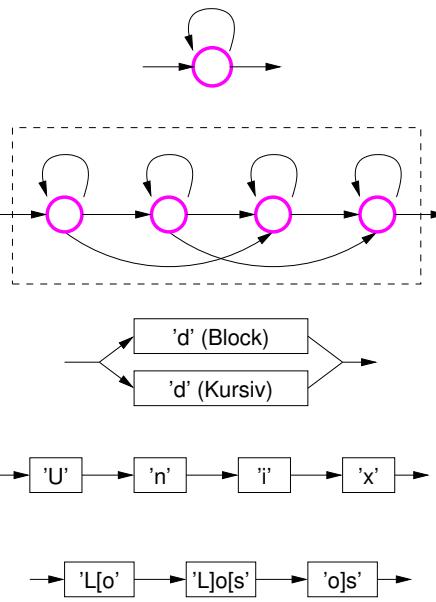
Anwendungsszenarien
ooooooo

Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooo●ooooo

Hidden Markov Modelle für Zeichen und Wörter

Verknüpfung von HMMs: seriell, parallel, Rückkopplung



elementarer HMM-Zustand für Liniensegmente

einfaches Zeichenmodell aus Segmentmodellen

komplexes Zeichenmodell

Wortmodell aus Zeichenmodellen

Wortmodell aus kontextabhängigen Zeichenmodellen

Schriftdaten
ooooooooooooooo

Anwendungsszenarien
ooooooo

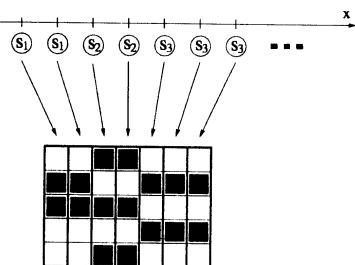
Explizit segmentierend
ooooooooooooooo

Implizit segmentierend
ooooo●ooooo

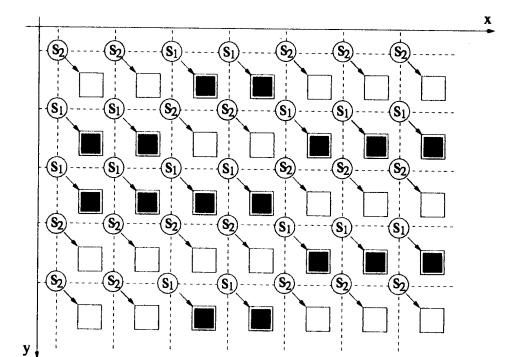
Eindimensionale & zweidimensionale Zufallsprozesse

1D-HMM

HMM → Zustandsfolge



Zustand → Bildspalte



2D-HMM

HMM → Zustandsmatrix

Zustand → Bildpunkt

2D Hidden Markov Modelle

Markvgitter (*Was ist das?*) & Pixelausgabeverteilung

$$P(\mathbf{X}) = \sum_{\mathbf{Q} \in \mathcal{S}^{N \times M}} P(\mathbf{Q}, \mathbf{X}), \quad \mathcal{S} = \{s_1, \dots, s_K\}$$

$$P(\mathbf{Q}, \mathbf{X}) = P(\mathbf{Q}) \cdot P(\mathbf{X}|\mathbf{Q})$$

$$P(\mathbf{X}|\mathbf{Q}) = \prod_{\xi=1}^N \prod_{\eta=1}^M b_{q_{\xi\eta}}(x_{\xi\eta})$$

Graphische Modelle

(Gibbspotenziale/Cliquen)

$$P(\mathbf{Q}) = \prod_{A \in \mathcal{C}} \phi_A(\mathbf{Q}_A)$$

Kausale Modelle

(Kettenregel/reduziert · Variablenfolge?)

$$P(\mathbf{Q}) = \prod_{\xi=1}^N \prod_{\eta=1}^M P(Q_{\xi\eta} = q_{\xi\eta} | \dots)$$

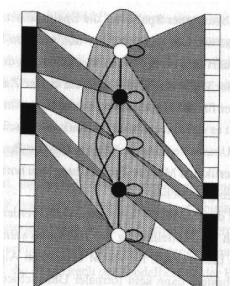
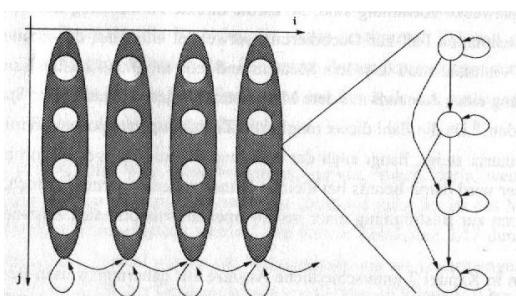
Pseudo 2D Hidden Markov Modelle

Geschachtelte elastische Membran

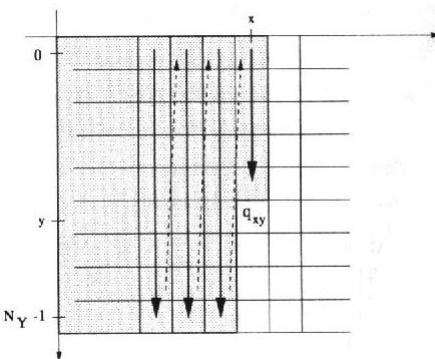
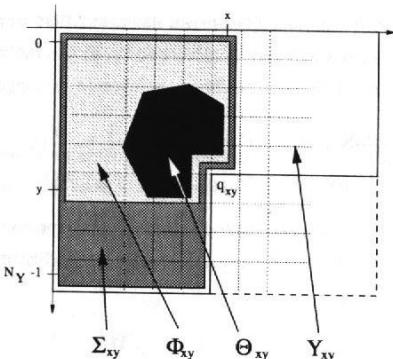
HMM Metazustandsfolge

Metazustand Zustandsfolge

Zustand Bildpunkt



MRF — Markov Random Fields



MRF Abhängigkeitstopologien

- **Kausales MRF**

$$P(Q_{xy} = s_k | Q_{\Phi_{xy}}) = P(Q_{xy} = s_k | Q_{\Theta_{xy}})$$

- **Markov Mesh**

$$P(Q_{xy} = s_k | Q_{\Gamma_{xy}}) = P(Q_{xy} = s_k | Q_{\Theta_{xy}})$$

- **Unilaterales MRF**

$$P(Q_{xy} = s_k | Q_{\Sigma_{xy}}) = P(Q_{xy} = s_k | Q_{\Theta_{xy}})$$

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil X

Rekursive Markovmodelle

Architektur oooo	Theorie RMM oooooooooooo	Zustandstypen oooooooooooo	Welche $b_j(x_t)$? ooo	E-Zustände oooooooo	Dekodierung oooooooo
---------------------	-----------------------------	-------------------------------	----------------------------	------------------------	-------------------------

Architektur eines hierarchischen HMM-Systems

Baupläne für HMMs · Kompilierte HMM-Netzwerke ·
HMM-Hierarchien

Mathematische Theorie des RMM

Spezialisierte Typen von RMM-Zuständen

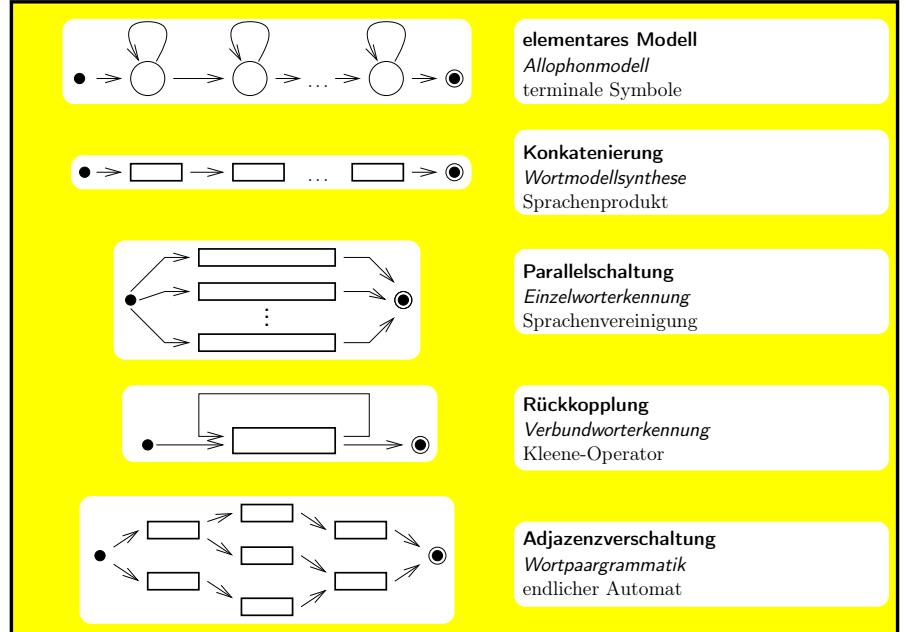
Ausgabeverteilungen der Zustände

Elementare Zustände des RMM

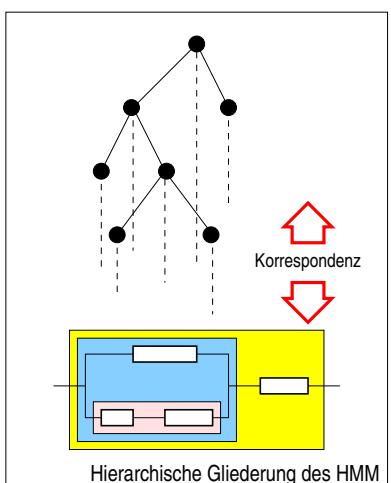
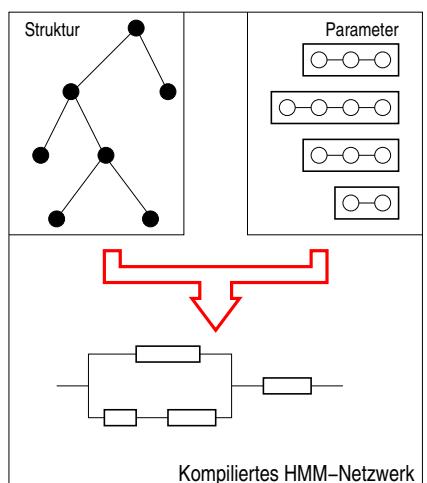
Geschachtelte symbolische Beschreibung

Architektur ●ooo	Theorie RMM oooooooooooo	Zustandstypen oooooooooooo	Welche $b_j(x_t)$? ooo	E-Zustände oooooooo	Dekodierung oooooooo
---------------------	-----------------------------	-------------------------------	----------------------------	------------------------	-------------------------

Finite-State Verschaltung von HMMs



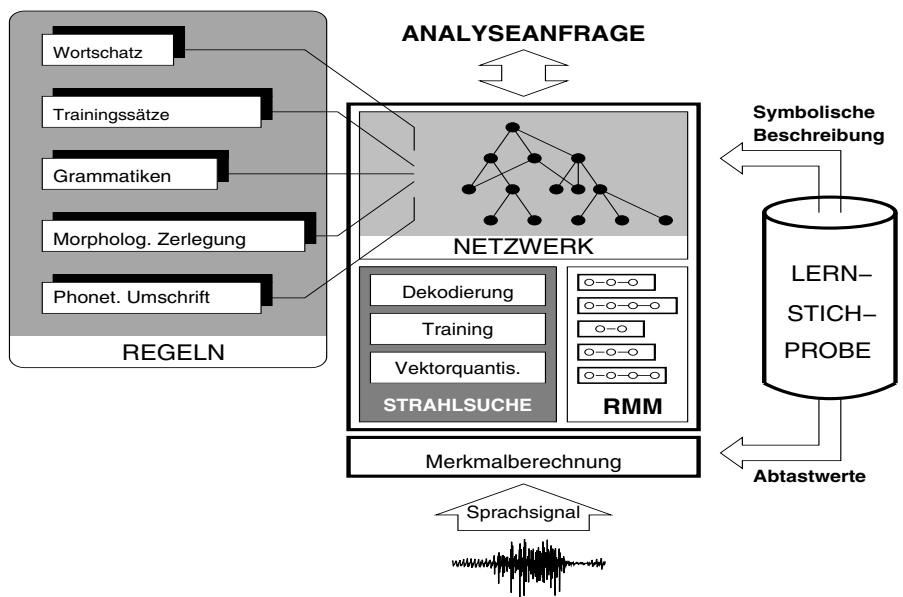
Kompilation versus Integration



Gesucht ist ein Modellierungsformalismus mit ...

- **expliziten Verknüpfungsoperationen** für Hidden Markov Modelle
- **strukturerhaltender Repräsentation** zur Erzeugung symbolischer Beschreibungen der Eingangsdaten
- **statistischem Vererbungsmechanismus** für die (Spezialisierungshierarchie der) Ausgabeverteilungen

ISADORA-System zur Spracherkennung



Architektur eines hierarchischen HMM-Systems

Mathematische Theorie des RMM

Elementare & komplexe Zustände · RFA/RBA/RVA/RBWA

Spezialisierte Typen von RMM-Zuständen

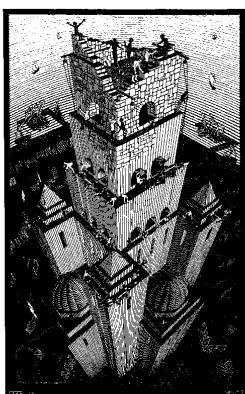
Ausgabeverteilungen der Zustände

Elementare Zustände des RMM

Geschachtelte symbolische Beschreibung

Johannes Josephus Nijtmans — TU Eindhoven 1992

SPEECH RECOGNITION BY
RECURSIVE STOCHASTIC MODELLING



SPEECH RECOGNITION BY
RECURSIVE STOCHASTIC MODELLING

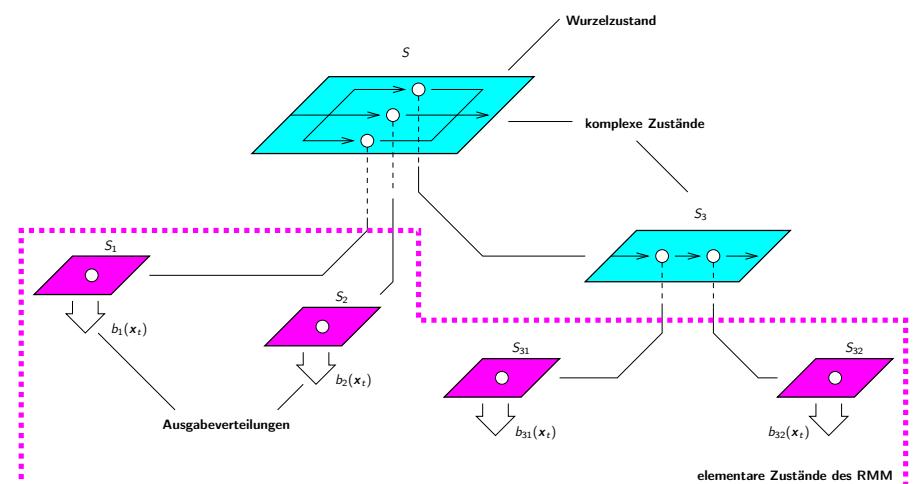
PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Technische Universiteit Eindhoven, op gezag van
de Rector Magnificus, prof. dr. J.H. van Lint,
voor een commissie aangewezen door het
College van Dekanen in het openbaar te
verdedigen op woensdag 6 mei 1992 om 16.00 uur

door
JOANNES JOSEPHUS NIJTMANS
Geboren te Oisterwijk

J.J. NIJTMANS

Rekursives Hidden Markov Modell



Was ist ein Rekursives Markov Modell ?

Definition

Ein RMM-Zustand über dem Ausgabealphabet $\mathcal{V} = \{v_1, \dots, v_M\}$ ist:

- Ein **elementarer** Zustand S ist durch eine Ausgabeverteilung

$$b_S : \mathcal{V} \rightarrow [0, 1], \quad \sum_{m=1}^M b_S(v_m) = 1$$

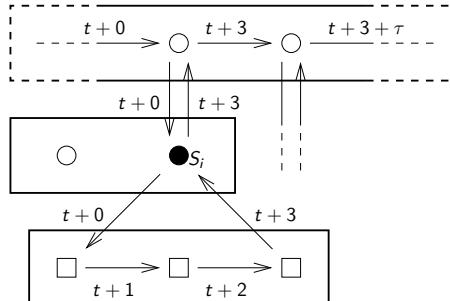
charakterisiert.

- Ein **komplexer** Zustand S besteht aus einem Tupel (S_1, \dots, S_N) von RMM-Zuständen, die im Sinne einer einfachen, stationären Markovkette stochastisch miteinander vernetzt sind. Das Wahrscheinlichkeitsgesetz von S ist durch die Parametermatrix

$$\mathbf{A} = \left(\begin{array}{c|ccccc} x & a_{11} & a_{12} & \dots & a_{1N} \\ \hline a_{1F} & a_{11} & a_{12} & \dots & a_{1N} \\ a_{2F} & a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{NF} & a_{N1} & a_{N2} & \dots & a_{NN} \end{array} \right), \quad \sum_{j=1}^N a_{ij} = 1 \quad (\forall i) \quad a_{iF} + \sum_{j=1}^N a_{ij} = 1$$

charakterisiert.

Die Zeitrechnung im RMM



- Komplexe Zustände
- Elementare Zustände

Bemerkung

Unterscheide zwischen

- dem **Betreten** von S zum Zeitpunkt t ,
- dem **Verlassen** von S zum Zeitpunkt $t + \Delta t + 1$ und
- dem **Verharren** in S von t bis einschließlich $t + \Delta t$. Zu jedem Zeitpunkt ist ein ganzer Zustandskeller **aktiv**:

$$S^* = S_{i_r} \prec S_{i_{r-1}} \prec S_{i_{r-2}} \prec \dots \prec S_{i_2} \prec S_{i_1} = S_e,$$

wobei S^* der Wurzelzustand des Gesamtmodells ist.

Die Wahrscheinlichkeitsparameter des RMM

Elementare Zustände

Wird der Zustand S zum Zeitpunkt $t \in \mathbb{N}$ betreten, so wird mit der Wahrscheinlichkeit

$$P(\mathbb{Y}_t = v_k \mid \mathbb{X}_t = S) = b_S(v_k)$$

ein Zeichen erzeugt und S wird zum Zeitpunkt $t + 1$ wieder verlassen.

Komplexe Zustände

Die Anfangswahrscheinlichkeiten a_{lj} , die Übergangswahrscheinlichkeiten a_{ij} und die Endewahrscheinlichkeiten a_{iF} haben die folgende Bedeutung:

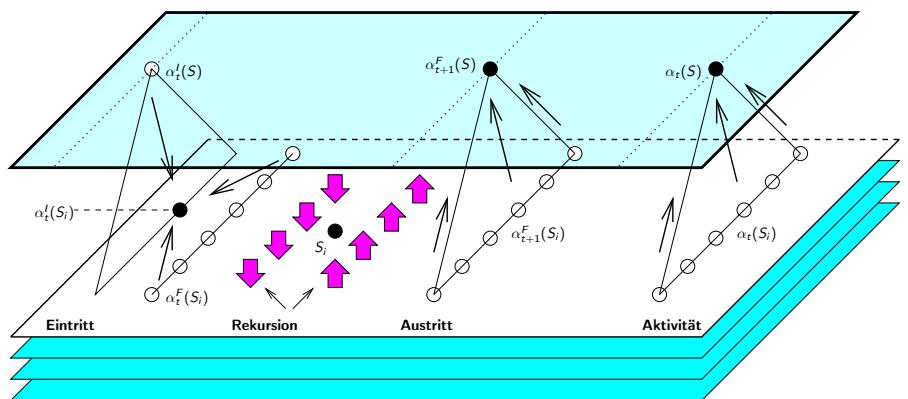
$$a_{lj} \stackrel{\text{def}}{=} P(S_j \text{ begonnen in } t \mid S \text{ begonnen in } t)$$

$$a_{ij} \stackrel{\text{def}}{=} P(S_j \text{ begonnen in } t \mid S_i \text{ beendet in } t)$$

$$a_{iF} \stackrel{\text{def}}{=} P(S \text{ beendet in } t \mid S_i \text{ beendet in } t)$$

Die *Unterzustände* (oder *Nachfolger*) S_1, \dots, S_N von S sind nicht notwendig paarweise verschieden (*gleich* vs. *identisch*).

Die Vorwärtswahrscheinlichkeiten II



Effiziente $O(N_{S^*} \cdot T)$ Berechnung simultan durch

- Iteration über die Zeit $t = 1, 2, \dots, T$ und
- Induktion über den Aufbau von S

Die Vorwärtswahrscheinlichkeiten I

Definition

Für jeden Zustand S und für jeden Zeitpunkt t definieren wir drei *Vorwärtswahrscheinlichkeiten*:

$$\alpha_t^I(S) \stackrel{\text{def}}{=} P(o_1, \dots, o_{t-1}, S \text{ begonnen in } t)$$

$$\alpha_t(S) \stackrel{\text{def}}{=} P(o_1, \dots, o_t, S \text{ aktiv in } t)$$

$$\alpha_t^F(S) \stackrel{\text{def}}{=} P(o_1, \dots, o_{t-1}, S \text{ beendet in } t)$$

Folgerung

Nach der Iteration $t = 1, \dots, T$ erhalten wir die Wahrscheinlichkeit, daß eine Zeichenfolge $\mathbf{o} \in \mathcal{V}^T$ vom RMM S^* erzeugt worden ist:

$$P(\mathbf{o} \mid S^*) = P(o_1, \dots, o_T, S^* \text{ beendet in } T+1) = \alpha_{T+1}^F(S^*)$$

Der geschachtelte Vorwärtsalgorithmus (RFA)

1 INDUKTIONSBEGINN (S^* Wurzel des RMM)

$$\alpha_t^I(S^*) = \begin{cases} 1 & t = 1 \\ 0 & t = 2..T \end{cases}$$

2 INDUKTIONBERANDUNG: $\alpha_1^F(S) = 0$ für alle Zustände S

3 INDUKTIONSENDE: (S ist elementar)

$$\alpha_{t+1}^F(S) = \alpha_t(S) = \alpha_t^I(S) \cdot b_S(o_t)$$

4 INDUKTIONSSCHRITT: (S ist komplex)

$$\alpha_t^I(S_j) = \alpha_t^I(S) \cdot a_{lj} + \sum_{i=1}^N \alpha_t^F(S_i) \cdot a_{ij} \quad (\forall j = 1..N)$$

(zwischendurch rekursiver Abstieg für S_j , $j = 1..N$)

$$\alpha_t(S) = \sum_{i=1}^N \alpha_t^I(S_i)$$

$$\alpha_{t+1}^F(S) = \sum_{i=1}^N \alpha_{t+1}^F(S_i) \cdot a_{iF}$$

Die Rückwärtswahrscheinlichkeiten

Definition

Für jeden Zustand S und für jeden Zeitpunkt t definieren wir drei Rückwärtswahrscheinlichkeiten:

$$\beta_t^I(S) \stackrel{\text{def}}{=} P(o_t, \dots, o_T \mid S \text{ begonnen in } t)$$

$$\beta_t(S) \stackrel{\text{def}}{=} \begin{cases} \beta_t^I(S) & S \text{ elementar} \\ \sum_j \beta_t(S_j) & S \text{ komplex} \end{cases}$$

$$\beta_t^F(S) \stackrel{\text{def}}{=} P(o_t, \dots, o_T \mid S \text{ beendet in } t)$$

Die Präsenzwahrscheinlichkeiten $\beta_t(S)$ haben in Rückwärtsrichtung offenbar keine anschauliche Bedeutung.

Folgerung

Nach der Iteration $t = T+1, \dots, 1$ erhalten wir die Wahrscheinlichkeit, daß eine Zeichenfolge $\mathbf{o} \in \mathcal{V}^T$ vom RMM S^* erzeugt worden ist:

$$P(\mathbf{o} \mid S^*) = P(o_1, \dots, o_T \mid S^* \text{ begonnen in } t=1) = \beta_1^I(S^*)$$

Sonstige Algorithmen für das RMM

- Viterbi-Algorithmus (RVA)**
berechnet die wahrscheinlichste Zustands(keller)folge
- Baum-Welch/EM-Algorithmus (RBWA)**

$$\gamma_t(S) = \alpha_t^I(S) \cdot \beta_t^I(S) / P(\mathbf{o} | S^*)$$

$$\xi_t^I(S_j) = \alpha_t^I(S) \cdot a_{lj} \cdot \beta_t^I(S_j) / P(\mathbf{o} | S^*)$$

$$\xi_t^F(S_i) = \alpha_{t+1}^F(S_i) \cdot a_{iF} \cdot \beta_{t+1}^F(S) / P(\mathbf{o} | S^*)$$

$$\xi_t(S_i, S_j) = \alpha_t^F(S_i) \cdot a_{ij} \cdot \beta_t^I(S_j) / P(\mathbf{o} | S^*)$$

- Skalierung** der Wahrscheinlichkeitswerte
erforderlich beim Analysieren „langer“ Muster ($T \gg 100$)
- Strahlsuche** für RFA, RBA, RVA
reduziert die Anzahl evaluierter (S, t) -Paare auf 1–10%

Der geschachtelte Rückwärtsalgorithmus (RBA)

(Algorithmus)

- 1 INDUKTIONSBEGINN (S^* Wurzel des RMM)

$$\beta_t^F(S^*) = \begin{cases} 1 & t = T+1 \\ 0 & t = 1..T \end{cases}$$

- 2 INDUKTIONBERANDUNG: $\beta_{T+1}^I(S) = 0$ für alle Zustände S

- 3 INDUKTIONSENDE: (S ist elementar)

$$\beta_t^I(S) = \beta_t(S) = \beta_{t+1}^F(S) \cdot b_S(o_t)$$

- 4 INDUKTIONSSCHRITT: (S ist komplex)

$$\beta_{t+1}^F(S_i) = a_{iF} \cdot \beta_{t+1}^F(S) + \sum_{j=1}^N a_{ij} \cdot \beta_{t+1}^I(S_j) \quad (\forall j = 1..N)$$

(zwischendurch rekursiver Abstieg für $S_j, j = 1..N$)

$$\beta_t(S) = \sum_{i=1}^N \beta_t(S_i)$$

$$\beta_t^I(S) = \sum_{j=1}^N a_{ij} \cdot \beta_t^I(S_j)$$

(summungIA)

RMM versus HMM-Netzwerk

Folgerung

Jedes RMM S^* besitzt ein $P(\cdot)$ -äquivalentes HMM $\lambda = \lambda(S^*)$

- Zustände von λ sind die elementaren Zustände von S^* .

- Ausgabeverteilung von $s_j = S_{\ell(j)}$ ist

$$b_j(v_k) \stackrel{\text{def}}{=} b_{S_{\ell(j)}}(v_k), \quad k = 1, 2, \dots, M$$

- Übergangswahrscheinlichkeiten von s_i nach s_j ergeben sich als Produkt

$$a_{ij} \stackrel{\text{def}}{=} \prod_{r=1}^{h(i)-1} a_{x_F}(S_{\ell_r}) \cdot a_{xx}(S_{\ell(i,j)}) \cdot \prod_{r=1}^{h(j)-1} a_{lx}(S_{\ell_r})$$

aller E/A/Ü-Wahrscheinlichkeiten der Verbindungswege

$$\begin{aligned} S_{\ell_0} &\prec S_{\ell_1} \prec S_{\ell_2} \prec \dots \prec S_{\ell_{h(i)-1}} \prec S_{\ell_{h(i)}} = S_{\ell(i)} \\ S_{\ell'_0} &\prec S_{\ell'_1} \prec S_{\ell'_2} \prec \dots \prec S_{\ell'_{h(j)-1}} \prec S_{\ell'_{h(j)}} = S_{\ell(j)} \end{aligned}$$

zwischen den elementaren Zuständen (über $S_{\ell(i,j)} = S_{\ell_0} = S_{\ell'_0}$).

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_j(x_t)$?
oooE-Zustände
ooooooooDekodierung
ooooooooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
●ooooooooooooWelche $b_j(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooooooo

Spezialisierte Zustandstypen ?

Architektur eines hierarchischen HMM-Systems

Mathematische Theorie des RMM

Spezialisierte Typen von RMM-Zuständen

Elementar · Syntagma/Paradigma · Loop · etc.

Ausgabeverteilungen der Zustände

Elementare Zustände des RMM

Geschachtelte symbolische Beschreibung

1. Die Wahrscheinlichkeitsparameter eines RMM werden automatisch (EM) aus Daten gelernt.
2. Die Struktur eines RMM wird vom Entwickler im Hinblick auf die Domäne festgelegt.
3. Im RMM ist die Modellstruktur *vorwiegend* in den Modellparametern repräsentiert.
4. Die explizite Angabe von Übergangswahrscheinlichkeiten durch die Entwicklerin ist unzumutbar & nicht erstrebenswert.
 - ⇒ Die zu Grunde liegende Modelltopologie muß dem Benutzer explizit zugänglich sein!

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
○○ooooooooooooWelche $b_j(x_t)$?
oooE-Zustände
ooooooooDekodierung
ooooooooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
○○○ooooooooooooWelche $b_j(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooooooo

Spezialisierte Zustandstypen I

E elementar

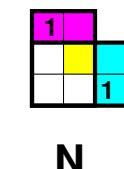
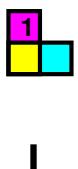
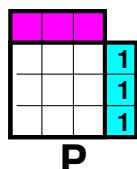
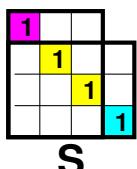
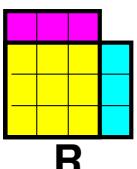
keine Unterzustände · eine Ausgabeverteilung

R RMM, ohne Restriktion

komplexer RMM-Zustand · $N \geq 1$ Kinder und $(N+1)^2 - 1$ Ü-W'keiten

C collection

$N \geq 1$ Kinder · keine Übergänge/Parameter



Spezialisierte Zustandstypen II

S seriell, syntagmatisch

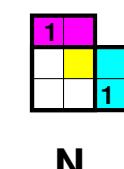
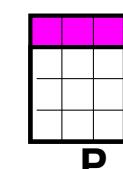
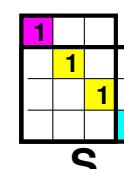
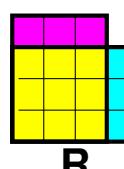
$N \geq 1$ Kinder und keine (!) freien Parameter ($a_{ij} = \delta_{i=j-1}$)

P parallel, paradigmatisch

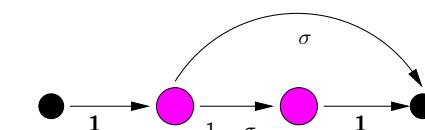
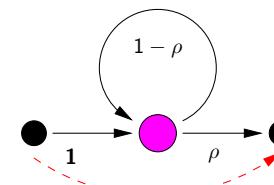
$N \geq 1$ Kinder und N Auswahlwahrscheinlichkeiten $a_{l,i}$

L loop, Wiederholung

ein Kindzustand · ein freier Parameter: Fluchtwahrsch. $a_{1,F}$

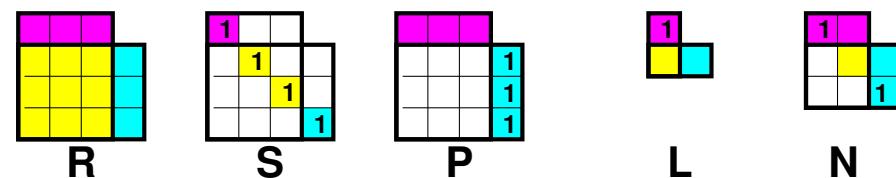


Spezialisierte Zustandstypen III



N bypass, $\frac{1}{2}$ -Bakis₂

zwei Kindzustände · S_2 mit Wahrsch. $\sigma = a_{1,F}$ passierbar
notwendig wegen $a_{I,F} = 0$ (Verbot nullproduktiver Zustände)



Spezialisierte Zustandstypen IV

D duration, Dauer

N serielle Kindzustände · freie Einsprungwahl $a_{I,j} > 0 (\forall j)$

B Bakis-LR-Modell

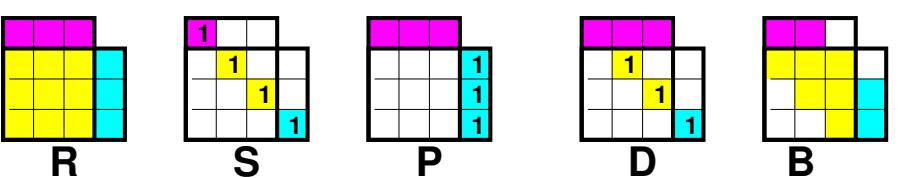
Ü-Wahrscheinlichkeiten $a_{i,i}$, $a_{i,i+1}$ und $a_{i,i+2}$

T tree, DAG ohne u.g. Zyklen

Ü-Wahrscheinlichkeiten $a_{\pi_j,j}$ und $a_{j,F}$ sowie a_{I,j_0}

A azyklische gerichtete Graphen, DAG

Ü-Wahrscheinlichkeiten $a_{i,j}$ für alle $I, 1 \leq i \leq j \leq N, F$



Beispiel: elementare Zustände

Isadora-Deklarationen

```
create 5
insert E          goethe
insert E          schiller
insert E          lessing
>RMM /dev/stdout
```

RMM-Ausziss

```
dgraph<S,T> 3
goethe 0]      E:-+
lessing 0]     E:-+
schiller 0]    E:-+
dgraph<S,T> 4
goethe 0]      DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
schiller 0]    DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
lessing 0]      DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
blueprint 3]   goethe schiller lessing DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
```

Beispiel: syntagmatische Zustände

Isadora-Deklarationen

```
create 5
insert E          null
insert E          sieben
insert S          007    null null sieben
>RMM /dev/stdout
```

RMM-Ausziss

```
dgraph<S,T> 3
null 0]       E:-+
sieben 0]     E:-+
007 3]        null null sieben S:+-
dgraph<S,T> 3
null 0]       DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
sieben 0]     DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
blueprint 2]  null sieben DISTRIBUTION 1
5           0.2 0.2 0.2 0.2 0.2
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooo●oooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
oooooooo

Beispiel: paradigmatische Zustände

Isadora-Deklarationen

```
create 5
insert E      sekt
insert E      selters
insert P      alternative    sekt selters
insert P      ramadan       selters selters selters
insert P      harald_juhnke sekt sekt sekt sekt
>RMM      /dev/stdout
```

RMM-Ausriß

```
dgraph<S,T> 5
sekt 0]   E:+-
selters 0] E:-+
alternative 2]
    sekt selters P:-+DISTRIBUTION 1
2        0.5 0.5
harald_juhnke 4]
    sekt sekt sekt sekt P:-+DISTRIBUTION 1
4        0.25 0.25 0.25 0.25
ramadan 3]
    selters selters selters P:-+DISTRIBUTION 1
3        0.333333 0.333333 0.333333
dgraph<S,T> 3
sekt 0]   DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
selters 0] DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
blueprint 2]
    sekt selters DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
```

Beispiel: wiederholende Zustände

Isadora-Deklarationen

```
create 5
insert E      body
insert L      loop   body
>RMM      /dev/stdout
```

RMM-Ausriß

```
dgraph<S,T> 2
body 0]   E:+-
loop 1]   body L:-+DISTRIBUTION 1
2        0.5 0.5
dgraph<S,T> 2
body 0]   DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
blueprint 1] body DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooo●ooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
oooooooo

Beispiel: aggregierende (kollektive) Zustände

Isadora-Deklarationen

```
insert C SAMMLUNG aus jedem hund ein dorf
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo●ooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
oooooooo

Beispiel: vollständig vernetzte Zustände

Isadora-Deklarationen

```
create 5
insert E      state
insert R      RMM      state state state state
>RMM      /dev/stdout
```

RMM-Ausriß

```
dgraph<S,T> 2
state 0]   E:+-
RMM 4]   state state state state R:-+5]
DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
DISTRIBUTION 1
4        0.25 0.25 0.25 0.25
dgraph<S,T> 2
state 0]   DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
blueprint 1] state DISTRIBUTION 1
5        0.2 0.2 0.2 0.2 0.2
```

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
oooooooooooo●Welche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
ooooooooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
oooooooooooo●Welche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooooooo

Beispiel: optionale Rechtsnachfolger

Realisierung mit „Bordmitteln“

```
insert S vvb Victor      von      Bülow
insert S v-vb Victor    «sil»   von      Bülow
insert S vv-b Victor    von     «sil»   Bülow
insert S v-v-b Victor  «sil»   von     «sil»   Bülow

insert P Loriot        vvb     v-vb   vv-b    v-v-b
```

Isadora-Deklarationen

```
create 5
insert E    «sil»
insert E    x
insert S    Victor  x x x x x x
insert S    von    x x x
insert S    Bülow x x x x
insert N    Victor+ Victor  «sil»
insert N    von+   von    «sil»
insert S    Loriot Victor+ von+   Bülow
>RMM /dev/stdout
```

Isadora-Deklarationen

```
create 5
insert E elem
insert L loop elem
insert S sss loop loop loop
insert D ddd loop loop loop
>RMM /dev/stdout
```

RMM-Ausriß

```
digraph<S,T> 4
elem 0] E:-+
loop 1] elem L:-+DISTRIBUTION 1
2          0.5 0.5
ddd 3] loop loop loop D:-+DISTRIBUTION 1
3          0.333333 0.333333 0.333333
sss 3] loop loop loop S:+-
digraph<S,T> 2
elem 0] DISTRIBUTION 1
5          0.2 0.2 0.2 0.2 0.2
blueprint 1] elem DISTRIBUTION 1
5          0.2 0.2 0.2 0.2 0.2
```

Architektur
oooo Theorie RMM
oooooooooooo Zustandstypen
ooooooooooooooo Welche $b_f(x_t)$?
ooo E-Zustände
oooooooo Dekodierung
oooooooooooo

Architektur
oooo Theorie RMM
oooooooooooo Zustandstypen
oooooooooooo● Welche $b_f(x_t)$?
●oo E-Zustände
oooooooo Dekodierung
oooooooooooo

Architektur eines hierarchischen HMM-Systems

Mathematische Theorie des RMM

Spezialisierte Typen von RMM-Zuständen

Ausgabeverteilungen der Zustände

Diskret: hart/weich/offen · Stetig: gaußsch/naiv/etcetera

Elementare Zustände des RMM

Geschachtelte symbolische Beschreibung

Diskrete Wertebereiche für Ausgabesequenzen o_1, \dots, o_T

Diskrete Ausgabezeichen/ketten aus $\mathcal{V} = \{v_1, \dots, v_M\}$

- **hart quantisierte Daten** (DD-HMM)

$$o_t \in \mathcal{V} \quad \text{und} \quad b_S(v_m) \stackrel{\text{def}}{=} b_{S,m}$$

- **weich quantisierte Daten** (SC-HMM)

$$o_t \in [0, 1]^M \quad \text{und} \quad b_S(\mathbf{p}) \stackrel{\text{def}}{=} \sum_{m=1}^M p_m \cdot b_{S,m}$$

- **offener Symbolvorrat** (Wörter/Zeichenketten)

$$o_t \in \mathcal{V}^* \quad \text{und} \quad b_S(w) \stackrel{\text{def}}{=} \begin{cases} b_{S,m} & \exists m \leq \tilde{M} : w = v_m \\ \delta_{S,\tilde{M}} & w \text{ „neu“} \end{cases}$$

Stetige Wertebereiche für Ausgabesequenzen o_1, \dots, o_T Stetige Ausgabevektoren des \mathbb{R}^N , $N \in \mathbb{N}$

$$x_t \in \mathbb{R}^N \quad \text{und} \quad b_S(x_t) \stackrel{\text{def}}{=} \mathcal{N}(x_t | \mu_S, S_S)$$

• Probleme mit der multivariaten Gaußdichte

- Fluch der Dimension
- Plage des Datensammelns
- Schlechte Kondition
- Modellannahmen

Rechen/Speicheraufwand $O(N^2)$

$$T \not\asymp N \Rightarrow \det S = 0$$

$$T \gg N \Rightarrow \text{Generalisierung?!?}$$

Symmetrie, expon. Abklingen, Unimodalität

• Derivate des Normalverteilungsmodells

- Erzwinge Nulleinträge in der Kovarianzmatrix S Blockdiagonalform
- Erzwinge Nulleinträge in der Konzentrationsmatrix S^{-1} Markovnetze
- Beschränkung auf dominierende Hauptachsen PPCA, FA
- Faktorisiere S in dünn besetzte Matrizen Bayesnetze

Spezielle Strukturierung der Kovarianzmatrix

• Unterschiedliche Besetzung der Kovarianzmatrix (BDF)

$$S = \text{diag}(\sigma_1, \dots, \sigma_N), \quad S = \sigma \cdot E, \quad S = 1 \cdot E$$

• Probabilistische Hauptkomponentenanalyse (PPCA)

$$S = \underbrace{UDU^\top}_M + \underbrace{U'D'U'^\top}_{N-M} \approx UDU^\top + \sigma \cdot U'U'^\top =: S^*$$

• Faktorenanalyse (FA)

$$S^{-1} \approx \Phi^\top \Phi + \Psi^2, \quad \Phi \in \mathbb{R}^{M \times N}, \quad \Psi = \text{diag}(\psi_1, \dots, \psi_N)$$

• Bayesnetze in Baumgestalt (TABN)

$$f(\mathbf{x}) = \prod_{i=1}^N f(x_i | x_1, \dots, x_{i-1}) = \prod_{i=1}^N f(x_i | x_{\pi(i)})$$

Architektur eines hierarchischen HMM-Systems

Mathematische Theorie des RMM

Spezialisierte Typen von RMM-Zuständen

Ausgabeverteilungen der Zustände

Elementare Zustände des RMM

Basis-PDF · geerbte PDF · eingefrorene Zustände ·
 PDF-Hierarchie

Geschachtelte symbolische Beschreibung

Wo kommen eigentlich die kleinen Ausgabeverteilungen her ?

1. Zu Beginn gibt es genau eine, uniforme PDF ('big bang').
2. Jeder neue E-Zustand bekommt eine Kopie davon.
3. Im Verlauf des Lernens verändern (EM) sich ihre Parameter.
4. Neue E-Zustände dürfen auch explizit erben.
5. die großindustrielle Herstellung:
komplexe Zustände rekursiv einfrieren

Wie funktioniert „Einfrieren“ ?

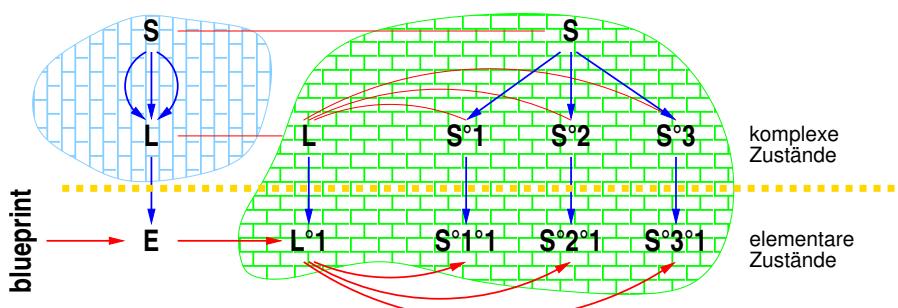
- S elementar:
Keine Aktion — S ist „sein eigenes Gefriergut“.
- S komplex mit Unterzuständen S_1, \dots, S_N :
 - Alle Unterzustände S_n , $n = 1..N$, werden eingefroren.
 - Von jedem (gefrosten!) Kind S_n wird eine tiefe Kopie S'_n angefertigt.
 - In S wird jeder Unterzustand S_n durch seine Kopie S'_n substituiert.

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
o●ooooooooDekodierung
ooooooooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
oo●ooooooooDekodierung
oooooooo

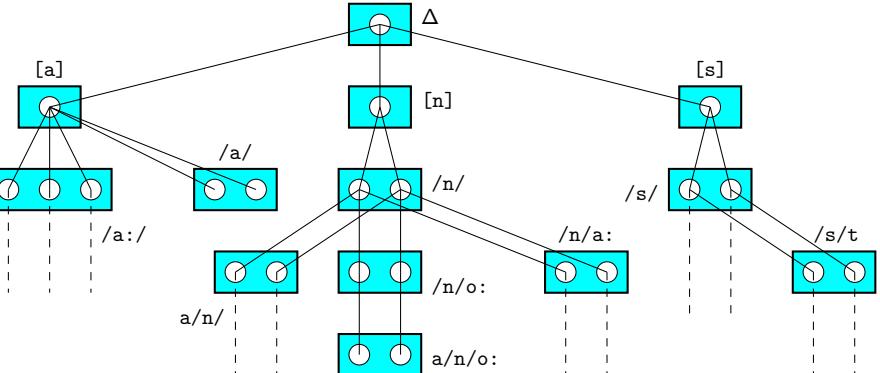
Beispiel zur Tiefkühltechnik

Isadora-Deklarationen

```
insert E ele
insert L loop ele
insert S seq loop loop loop
freeze seq
```



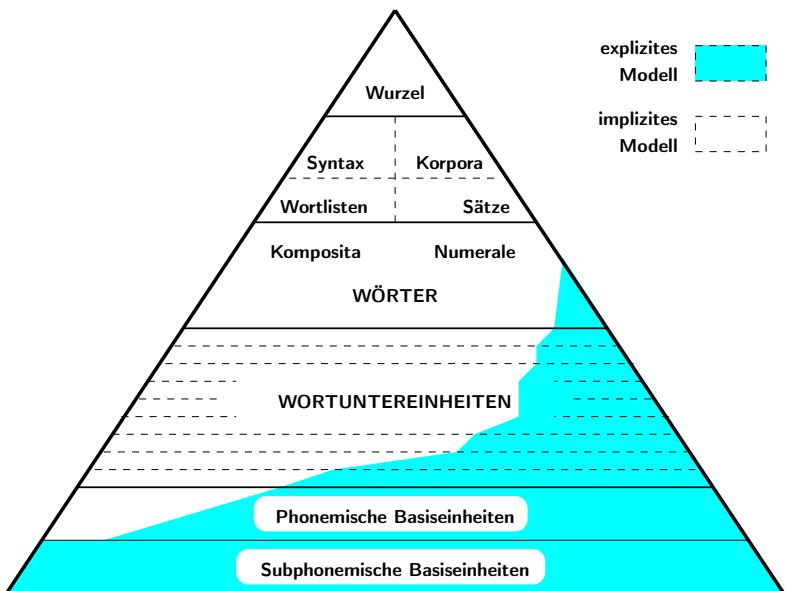
Interpolationsbaum der Ausgabeverteilungen



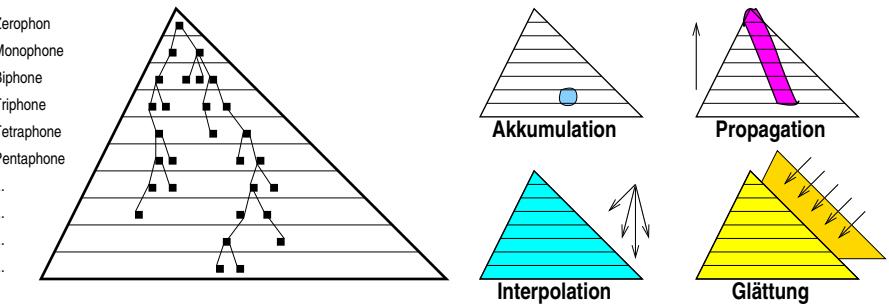
- Das Trainingsmaterial einer PDF ist die **Vereinigungsmenge** des Materials ihrer Spezialisierungen.
- Die Big-Bang-PDF trägt die Statistik **aller** Lernmuster.
- Speziellere** PDFs sind präzise Modelle.
- Generellere** PDFs sind robuste Modelle.

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooo●ooooDekodierung
ooooooooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooo○●ooooDekodierung
oooooooo

Eingefrorene Spracheinheiten



A.P.I.S. Lernverfahren



- (Algorithmus)
- A** AKKUMULATION aller a posteriori Erwartungswerte via Baum-Welch
 - P** PROPAGATION der Statistiken von unten nach oben
 - I** INTERPOLATION der Verteilungsparameter mit den Generalisierungen
 - S** SMOOTHING der Verteilungsparameter via MAP-Schätzung
- (sumdogia)

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo●ooo

Dekodierung
oooooooo

Architektur
oooo

Theorie RMM
oooooooooooo

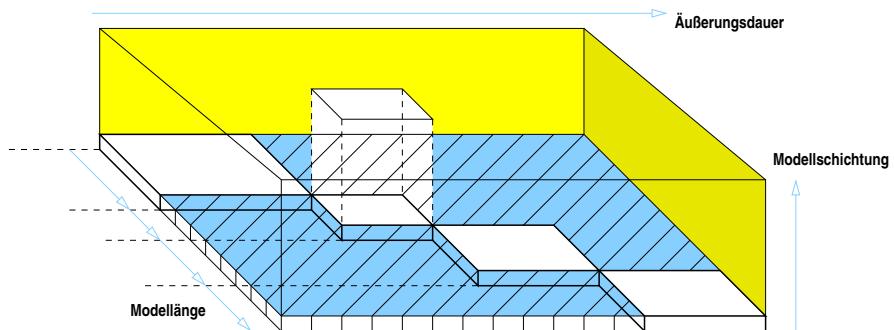
Zustandstypen
oooooooooooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo●oo

Dekodierung
oooooooo

Die Puppenstubenwelt des Baum-Welch-Algorithmus



Drei Dimensionen des maschinellen Lernverfahrens:

- Ausdehnung der Beobachtungsdaten
- Zustände des Modells
- Höhe der Generalisierungshierarchie
- ⇒ segmental entscheidungsüberwachtes Lernen

Daten, Erzeugungswahrscheinlichkeit, Vorbesetzung

Lesen/Schreiben von Zeitreihendaten

```
<EVS <filename>
>EVS <filename>
```

Berechnen der Vorwärts- und Rückwärtswahrscheinlichkeiten

```
algorun forward <state>
algorun backward <state>
algorun FORWARD <state1> <state2> <state3> ...
algorun BACKWARD <state1> <state2> <state3> ...
```

Überwachtes Lernen der PDF elementarer Zustände

```
...
insert E whitespace
$ cat /tmp/data/*blank*.evs > nothing.evs
<EVS nothing.evs
learn DETERMINISTIC whitespace
$ rm -f nothing.evs
...
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo●ooo

Dekodierung
oooooooo

Beschreibungsüberwachtes Lernen

Lernen komplexer Zustände mit dem RBWA

```
learn RESET
<EVS <file>
algorithm training <state>
learn ESTIMATE
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_f(x_t)$?
ooo

E-Zustände
oooooooo●ooo

Dekodierung
oooooooo

Beschreibungsüberwachtes Lernen mit dem RBWA

```
learn RESET
<EVS <file1>
algorithm training <state1>
<EVS <file2>
algorithm training <state2>
...
<EVS <fileL>
algorithm training <stateL>
learn ESTIMATE
```

Entscheidungsüberwachtes Lernen

Hoher Aufwand bei langen Lernsequenzen

```
<EVS ./digits.evs
algorithm TRAINING vier sieben eins eins
```

Segmentweise Durchführung des RBWA

```
<EVS ./digit-1.evs
algorithm training vier
<EVS ./digit-2.evs
algorithm training sieben
<EVS ./digit-3.evs
algorithm training eins
<EVS ./digit-4.evs
algorithm training eins
```

Beispiele aus Schrift, Sprache, Genom:

```
algorithm TRAINING | - 4 -- 7 -- 1 -- 1 -
algorithm TRAINING *stille* sein oder nicht sein *stille*
algorithm TRAINING C G C C G T A T A G C C C G A G C T A T A
```

Automatisierung des entscheidungsüberwachten Lernens

```
algorithm directed <state>
algorithm DIRECTED <state1> <state2> <state3> ...
```

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_j(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
oooooooo

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_j(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
●oooooooo

Architektur eines hierarchischen HMM-Systems

Mathematische Theorie des RMM

Spezialisierte Typen von RMM-Zuständen

Ausgabeverteilungen der Zustände

Elementare Zustände des RMM

Geschachtelte symbolische Beschreibung

Analysezustände · Symbol. Beschreibung · RVA · Opazität ·
Strahlsuche

Der rekursive Viterbi-Algorithmus

... findet die wahrscheinlichste Folge von **Zustandskellern** ...

- **Analyseaufgabe**

durch geeignet strukturierten Zustand repräsentiert:
Bigramme · reguläre Ausdrücke · endliche CFG

- **Strahlsuche**

manches wird komplizierter

- **Analyseergebnis**

ein (vollständiger) Parsebaum der Eingabe
· mit Konstituenteninformation
· und absoluten Zeit- bzw. Positionsmarken

- **Vorsicht, Datenfriedhof!**

Wer will das schon ?

Explizite Kontrolle über die Granulation der Analyse ...

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_j(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
○○○○○○○

Beispiele für (NLP) Analysezustände I

Einfache hierarchische Wortmodelle

S: Hamburg ham burg
S: ham /h/ /a/ /m/
S: burg /b/ /U/ /r/ /k/
S: /k/ [CLOS] [BURST] [ASP]
S: /r/

Aussprachevarianten

P: zwei /zwei/ /zwo/
S: /zwei/ /ts/ /v/ /aI/
S: /zwo/ /ts/ /v/ /o:/

Phrasenstrukturgrammatiken (rekursionsfrei)

S: <S> <NP> <VP>
S: <VP> <V> <NP>
P: <NP> Johann Marion Hasso ...
P: <V> liebt schlägt küßt ...

Ziffernfolgenerkennung

S: null /n/ /U/ /l/
S: eins /aI/ /n/ /s/
... ...
S: neun /n/ /OY/ /n/
P: ZIFFER null eins zwei drei ... neun
R: ZIFFERNFOLGE ZIFFER
S: *SATZ* <Stille> ZIFFERNFOLGE <Stille>

Architektur
oooo

Theorie RMM
oooooooooooo

Zustandstypen
oooooooooooo

Welche $b_j(x_t)$?
ooo

E-Zustände
oooooooo

Dekodierung
○○●○○○○

Beispiele für (NLP) Analysezustände II

Stichprobenbeschreibung

S: *7001* heute ist schönes Frühlingswetter
S: *7002* die Sonne lacht
... ...
S: -7001- SILENCE *7001* SILENCE
... ...
S: *LERNEN* -7001- -7002- -7003-

Triphondarstellung

S: heute /h/0 h/OY/t Y/t@ t@/
S: ist /I/s I/s/t s/t/
S: schönes /S/2 S/2:/n 2/n@

Zahlwörtergrammatik

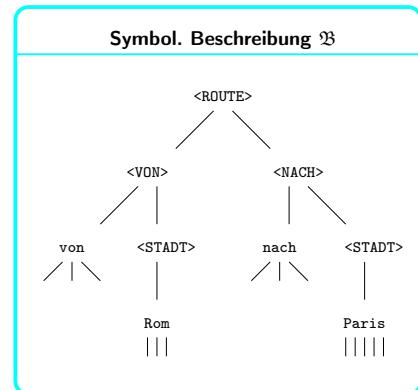
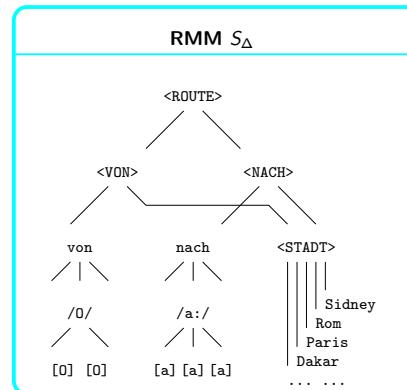
P: [1-9] eins zwei drei vier ... sieben acht neun
P: X-zehn zehn elf zwölf dreizehn ... neunzehn
P: X-zig zwanzig dreißig vierzig ... neunzig
P: und-PREF ein zwei drei vier ... sieben acht neun
S: X-und und-PREF und
P: X-UND X-und *NIL*
S: X-und-X-zig X-UND X-zig
P: [1-99] [1-9] X-zehn X-und-X-zig *NIL*
S: X-hundert und-PREF hundert
P: X-HUNDERT *NIL* hundert X-hundert
S: [1-999] X-HUNDERT [1-99]
P: tausend-PREF *NIL* ein [1-999]
S: X-tausend tausend-PREF tausend
P: X-TAUSEND *NIL* X-tausend
S: [1-999999] X-TAUSEND [1-999]
P: *Zahlen* [1-99] [1-999] [1-999999]

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
ooo●ooo

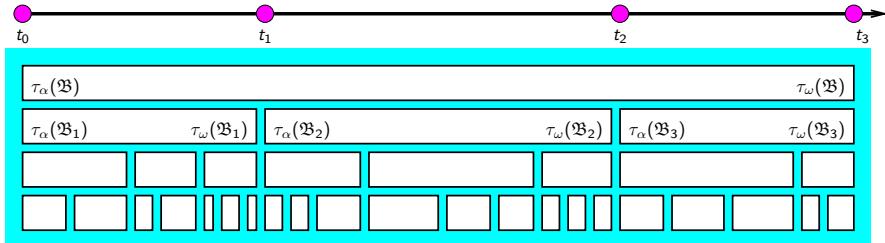
Symbolische Beschreibung

Isadora-Deklarationen

```
insert S <ROUTE> <VON> <NACH>
insert S <VON> von <STADT>
insert S <NACH> nach <STADT>
insert P <STADT> Sydney Rom Paris Dakar ...
...
... ... ... ... ...
```

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooo●ooo

Hierarchische Segmentierung im RMM



Analyseresultat ist ein Baum von Hypothesen der Form:

$$(S^b, \tau_\alpha^b, \tau_\omega^b) \in \mathcal{S} \times \mathbb{N} \times \mathbb{N}$$

Die Auflösung (Granularität) der Analyse wird „gedeckelt“ durch Markieren uninteressanter Zustände als **opak** (undurchsichtig).

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
ooo●ooo

Der rekursive Viterbi-Algorithmus

- 1 ELEMENTAR. Wenn S elementar ist, berechne

$$\vartheta_{t+1}^F(S) = \vartheta_t(S) = \vartheta_t^I(S) \cdot b_S(x_t).$$

Andernfalls ist S komplex und Schritte (2)–(5) kommen zum Zuge:

- 2 EINTRITT. Für jeden Nachfolger S_j berechne bzw. setze

$$\vartheta_t^I(S_j) = \max \left\{ \vartheta_t^I(S) \cdot a_{lj}, \vartheta_t^F(S_1) \cdot a_{1j}, \dots, \vartheta_t^F(S_N) \cdot a_{Nj} \right\}$$

$$\psi_t(S_j) = \begin{cases} S & \text{falls } \vartheta_t^I(S_j) = \vartheta_t^I(S) \cdot a_{lj} \\ S_i & \text{falls } \vartheta_t^I(S_j) = \vartheta_t^F(S_i) \cdot a_{ij} \end{cases}$$

- 3 REKURSION. Wende RVA in t auf alle Zustände S_j an.

- 4 AKTIVITÄT. Berechne $\vartheta_t(S) = \max_i \vartheta_t(S_i)$

- 5 AUSTRITT. Berechne bzw. setze

$$\vartheta_{t+1}^F(S) = \max_i (\vartheta_{t+1}^F(S_i) \cdot a_{iF})$$

$$\psi_{t+1}^F(S) = \operatorname{argmax}_{S_i} (\vartheta_{t+1}^F(S_i) \cdot a_{iF})$$

Architektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooo●oooArchitektur
ooooTheorie RMM
ooooooooooooZustandstypen
ooooooooooooWelche $b_f(x_t)$?
oooE-Zustände
ooooooooDekodierung
oooo●ooo

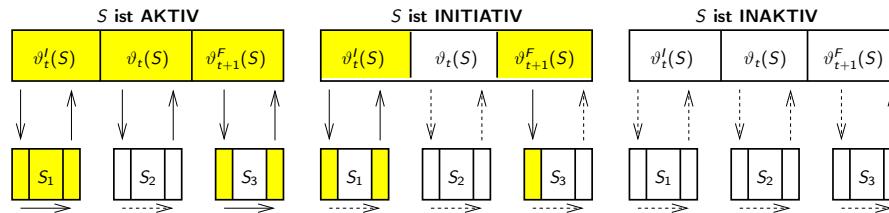
Opazität von Zuständen & lokales Gedächtnis

$$\begin{array}{lll} \psi : \{1, \dots, T\} \times \{S_1, \dots, S_N\} & \rightarrow & \{S, S_1, \dots, S_N\} \\ \psi^F : \{1, \dots, T\} & \rightarrow & \{S_1, \dots, S_N\} \\ \tau : \{1, \dots, T\} & \rightarrow & \{1, \dots, T\} \end{array}$$

Lokaler Speicherplatzbedarf für transparente Zustände

TYP	ψ	ψ^F	τ	Speicherplätze
R	×	×	-	$(N + 1) \cdot T$
P	-	×	-	T
L	×	-	-	T
S	-	-	-	0
E	-	-	×	T

Strahlsuche im RMM



Klassische Strahlsuche im HMM

Gitterpunkte (s_j, t) mit $v_t(j) \neq 0$ bzw. $v_t(j) \geq \beta \cdot \Lambda_t$ heißen **aktiv**, sonst **inaktiv**.

Hierarchische Strahlsuche im RMM

Der (komplexe, nicht elementare) Gitterpunkt (S, t) heißt

- **aktiv** falls $v_t(S) \neq 0$
- **initiativ** falls $v_t(S) = 0$, aber $v_t^I(S) \neq 0$
- **inaktiv** falls $v_t(S) = 0$ und auch $v_t^I(S) \neq 0$

SPEZIELLE MUSTERANALYSESYSTEME

Schrift- und Spracherkennung mit Hidden-Markov-Modellen

Vorlesung im Wintersemester 2018

Prof. E.G. Schukat-Talamazzini

Stand: 1. August 2018

Teil XI

RMM/ISADORA Anwendungsbeispiele

RMM-Schnittstelle
oooooo

Motivsuche
oooooooo

KFZ-Kennzeichen-Erkennung
oooooooooooo

Automatische Spracherkennung
oooooooooooooooooooo

RMM-Schnittstelle
●ooooo

Motivsuche
oooooooo

KFZ-Kennzeichen-Erkennung
oooooooooooo

Automatische Spracherkennung
oooooooooooooooooooo

Objektorientierte RMM-Schnittstelle

C++-Klasse für RMMs · Schnittstelle zur Sprache 'R'

Motivdetektion in DNA-Strängen

KFZ-Kennzeichen-Erkennung

Automatische Spracherkennung

Attribute

Keinerlei öffentliche Variablen !!

Konstruktoren

Standardkonstruktor: erzeugt ein „leeres“ RMM
Kopierkonstruktor, Untermodellkonstruktor, ...

Deserialisierung

Methoden

- **Projektionen**

Navigation, Inspektion, Reports, ...
Serialisierung
Wahrscheinlichkeitsbewertung, Symbolische Beschreibung von Zeitreihen

- **Hilfskonstruktoren**

Erzeugung eines neuen RMM-Zustands
Umschalten von Zustandsattributen
elementar · komplex
Durchführen eines RBW-Lernschritts
±trainable, ±opaque
reset · push* · A/P/I/S

RMM-Objekte ausschließlich als implizite Parameter !

RMM-Schnittstelle
oooooo

Motivsuche
oooooooo

KFZ-Kennzeichen-Erkennung
oooooooooooo

Automatische Spracherkennung
oooooooooooooooooooo

Objektorientierte RMM-Schnittstelle

Motivdetektion in DNA-Strängen

Aufgabe · Synthesedaten · Modellentwurf · Resultat

KFZ-Kennzeichen-Erkennung

Automatische Spracherkennung

Motivdetektion in DNA-Strängen

Gegeben:

1. Eine größere Anzahl von DNA-Beispielen $o \in \{A, C, G, T\}^*$
2. Die Vermutung, daß es in den Daten eine typische, wiederkehrende Sequenz („Motiv“) gibt
3. Eine begründete Annahme über die (ungefähre) Länge M des Motivs
4. Eine Annahme über das Auftretensmuster, z.B. je Datensatz genau einmal

Gesucht:

- ein **hartes** Motiv $q \in \{A, C, G, T\}^M$ oder
- ein **weiches** Motiv $p : [1..M] \times \{A, C, G, T\} \rightarrow [0, 1]$

Synthetische Beispieldaten

- $N = 100$ DNA-Stränge der Länge $T = 64$ werden zufällig erzeugt.
- Basen sind unabhängig und identisch mit $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ verteilt.
- Jeder Strang enthält das Motiv (Länge $M = 7$) an einer zufälligen Position.
- Motive werden generiert nach dem Wahrscheinlichkeitsprofil

	q_1	q_2	q_3	q_4	q_5	q_6	q_7
A	0.1	0.1	0.1	0.1	0.7	0.1	0.7
C	0.1	0.1	0.1	0.1	0.1	0.1	0.1
G	0.7	0.7	0.7	0.1	0.1	0.1	0.1
T	0.1	0.1	0.1	0.7	0.1	0.7	0.1

(ein weiches „GGGTATA“)

Beispielstränge

Sequenz #1

T G G G A T T G G A C C G T G A C A C A G G T C A G G A G G T G G
 A T T A A C A G G G T G T A T G G T T C T A T T T C G C C C G

Sequenz #2

T G G T G C T A T A G T C C A C A T G C G G C G T A G T T A G G C
 C A C G T A A G G G G T A T A G A T C A C G G T A G T T T A G

Sequenz #3

T A T C C C T C A A C T A C G G G T T A T C G A A A C T A A G G T
 C T A C T A C C T C C C C A G C A C C T A T T C C T A G T G A

Sequenz #4

C A C G G C T C A T A C A C A G T A G C A T C G C C G G G T A G A
 C T C C C A T T A C C G A T T A A C T A G C C C C A T A A A A

Sequenz #5

G G C C A C C G G T A T T C A T C T C T G T T A C G G G C A G C G
 A G T A C T C C C T T C G A G C G A A A C C C G A C T C G G G

Modellierung von Motiv & Sequenzen

Isadora-Deklarationen

```

create  @ALPHASIZE
insert  E      s
insert  L      «garbage»    s
insert  S      «motif»      s s s s s s s

freeze  «garbage»
freeze  «motif»

insert  S      OOPS        «garbage» «motif» «garbage»
Opaque  -      OOPS
Opaque  +      «garbage»
Opaque  +      «motif»

$ource  learn.isa   @BAUMWELCHSTEPS
$ource  segment.isa

>HSD   /dev/tty    Verbose
>RMM   /dev/tty
quit

```

Beispielstränge mit unüberwachter Segmentierung

Sequenz #1

T G G G A T T G G A C C C G T G A C A C A G G T C A G G A G G T G G
A T T A A C A G G G T G T A T G G T T C T A T T T C G C C C G

Sequenz #2

T G G T G C T A T A G T C C A C A T G C G G C G T A G T T A G G C
C A C G T A A G G G G T A T A G A T C A C G G T A G T T T A G

Sequenz #3

T A T C C C T C A A C T A C G G G T T A T C G A A A C T A A G G T
C T A C T A C C T C C C C A G C A C C T A T T C C T A G T G A

Sequenz #4

C A C G G C C T C A T A C A C A G T A G C A T C G C C G G G T A G A
C T C C C A T T A C C G A T T A A C T A G C C C C A T A A A A

Sequenz #5

G G C C A C C G G T A T T C A T C T C T G T T A C G C G C A G C G
A G T A C T C C C T T C G A G C G A A A C C C G A C T C G G G

30 Baum-Welch-Iterationen später ...

... liegt ein RMM vor, das die Eingabedaten u.a. wie folgt analysiert:

Die ersten fünf symbolischen Beschreibungen

1-64	OOPS	
=>	1-40	«garbage»
=>	41-47	«motif»
=>	48-64	«garbage»
1-64	OOPS	
=>	1-41	«garbage»
=>	42-48	«motif»
=>	49-64	«garbage»
1-64	OOPS	
=>	1-29	«garbage»
=>	30-36	«motif»
=>	37-64	«garbage»
1-64	OOPS	
=>	1-26	«garbage»
=>	27-33	«motif»
=>	34-64	«garbage»
1-64	OOPS	
=>	1-6	«garbage»
=>	7-13	«motif»
=>	14-64	«garbage»

Die Parameter des resultierenden RMM

RMM-Ausriß

```

dgraph<S,T> 10
«garbage»*1 0] DISTRIBUTION 5700
+8.76808e-05 4 A 0.247876 C 0.251157 G 0.252354 T 0.248175

«motif»*1 0] DISTRIBUTION 100
+0.00487805 4 A 0.116797 C 0.153957 G 0.666914 T 0.0379428

«motif»*2 0] DISTRIBUTION 100
+0.00487805 4 A 0.0963097 C 0.0465919 G 0.758758 T 0.0739498

«motif»*3 0] DISTRIBUTION 100
+0.00487805 4 A 0.0883085 C 0.0766578 G 0.642302 T 0.168341

«motif»*4 0] DISTRIBUTION 100
+0.00487805 4 A 0.0553429 C 0.0953407 G 0.101258 T 0.723669

«motif»*5 0] DISTRIBUTION 100
+0.00487805 4 A 0.645064 C 0.172578 G 0.118494 T 0.039474

«motif»*6 0] DISTRIBUTION 100
+0.00487805 4 A 0.0958529 C 0.111119 G 0.151422 T 0.617145

«motif»*7 0] DISTRIBUTION 100
+0.00487805 4 A 0.799781 C 0.0341988 G 0.0677086 T 0.0739214

s 8] «garbage»*1 «motif»*1 «motif»*2 «motif»*3 ... ... ... «motif»*7 DISTRIBUTION 6400
+7.80945e-05 4 A 0.251152 C 0.234752 G 0.264897 T 0.248809

blueprint 1] s DISTRIBUTION 6400
+7.80945e-05 4 A 0.251152 C 0.234752 G 0.264897 T 0.248809

```

Objektorientierte RMM-Schnittstelle

Motivdetektion in DNA-Strängen

KFZ-Kennzeichen-Erkennung

Konfiguration, Zeichenmodelle, Erkennungsmodelle

Automatische Spracherkennung

Kraftfahrzeug-Kennzeichen



Die Konfiguration des Erkenners

FALSE Grauwertlinearisierung

- 40** Bildfensterhöhe
- 9** Bildfensterbreite
- 2** Bildfensterfortschaltung
- 24** PCA-Zieldimension

b_S(x_t) Tree Bayesian Net**10⁻³** MAP-Schätzer (EQSS)**TRUE** entscheidungsüberwachtes RBWT

- 10** Baum-Welch-Iterationen

10 #_{min} Dedizierte Modelle

Die Erzeugung der RMM-Zustände

```

Feb 08, 08 19:12 kfv-symbols.sh Seite 1/1
#!/bin/sh
## source ./kfz.defs
## LETTER/DIGIT ALPHABETS
## (no 'S' and 'R')
lower="echo "abcdefghijklmnopqrstuvwxyz" | sed "s/J/g"
upper="echo "ABCDEFGHIJKLMNOPQRSTUVWXYZ" | sed "s/J/g"
digit="echo "0123456789" | sed "s/J/g"
space="echo " " | sed "s/J/g"
brace="echo "[" | sed "s/J/g"
other="echo "]" | sed "s/J/g"
EOF

# TOPOLOGIES: LINEAR LEFT-RIGHT RMMs OF VARYING SIZE
cat <>EOF
L p* p*
LSS p* p* p*
SS <digit> p* p* p* p*
SS <letter> p* p* p* p*
SS <brace> p* p* p* p*
SS <other> p* p* p* p*
SS <space> p* p* p* p*
EOF

# THE BASIC UNITS OF KFZ APPLICATION
for c in $digit; do echo "$c $c <digit>" done
for c in $space; do echo "$c $c <space>" done
for c in $brace; do echo "$c $c <brace>" done
for c in $other; do echo "$c $c <other>" done

# SUPERLETTERS 'X/X' IN ORDER TO TIE OUTPUT STATISTICS
for c in $lower; do echo "$c $c <lower>" done
for c in $upper; do echo "$c $c <upper>" done
for c in $lower; do
    C="echo $c | tr "[lower]" "[upper]"
    echo "$C $C $c$C"
    echo "$C $C $c$C"
done

# SOME IMPORTANT PARADIGMATIC SUBSETS ZB USED IN THE RECOGNITION MODELS
cat <>EOF
P <digit> $digit
P <space> $space
P <brace> $brace
P <brace> $upper
P <lower> $lower
P <letter> $lower $upper
P <symbol> $digit $upper
P <any> $digit $upper $other $space
EOF

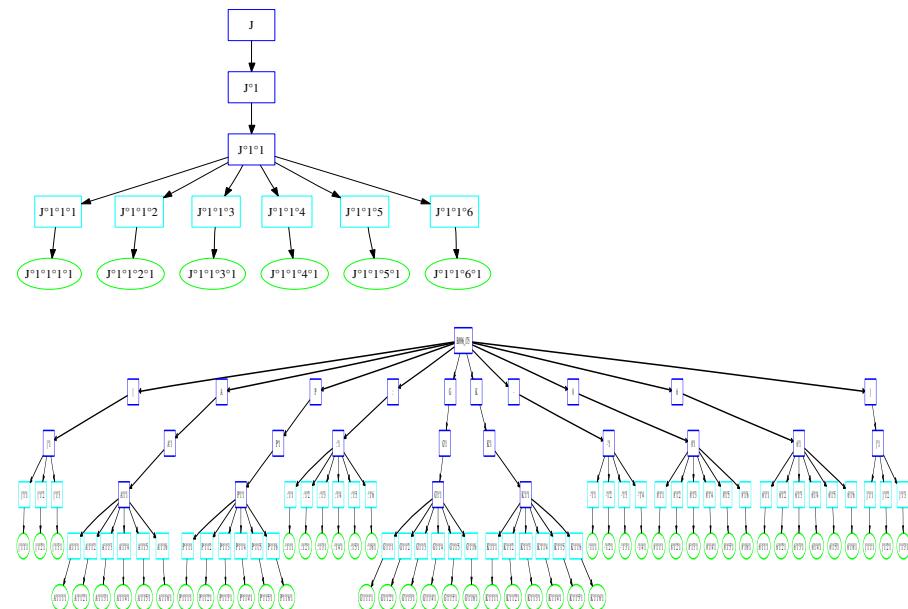
```

File location: /home/schukat/TEX/FOLIEN/Hidden-Markov-Modelle-07/img

Die Menge der RMM-Zustände

RMM-Schnittst.
ooooo

Beispiele für RMM-Zustände: 'J' und 'AP : GK 86'



Schrifteinheiten und ihre Häufigkeiten

Lerndatensammlung
ca. 2600 Kfz-Kennzeichen

Darstellung im RMM

Häufigkeitszählung

Traversieren ab «data»

> 10 Auftreten/Zustand

584 elementare Zustände

RMM-Zustände zur Musteranalyse

P «neuer_Zeichensatz» { } : 0 1 2 3 4 5 6 7 8 9
 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Ö Ü

P «alter_Zeichensatz» { } - : 0 1 2 3 4 5 6 7 8 9
 a b c d e f g h i j k l m n o p q r s t u v w x y z ö ü

P «voller_Zeichensatz» { } - - : 0 1 2 3 4 5 6 7 8 9
 a b c d e f g h i j k l m n o p q r s t u v w x y z ö ü
 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Ö Ü

Wiederholende Zustände für Zeichenketten

L «neues_KK» «neuer_Zeichensatz»
L «altes_KK» «alter_Zeichensatz»

L «irgend KK» «voller Zeichensatz»

Alternative für alte und neue Kennzeichen

Beispielresultate für '«alt|neu»'

```
{ A Z E : M · 3 2 3 } { t k · 3 4 i 8 2 7 } { A P : P G · 5 1 · }
{ O S : M R · 6 7 9 } { S H K : H Y · 8 3 } { W M : Z H 4 2 · }
{ M A : R Y · 1 2 7 } { M : S F · 6 4 8 5 } { A N A : S J · 3 0 0
{ L U : K · 6 8 6 2 } { B : C J · 9 1 4 4 } { H S T : V A · 8 8 }
{ G R Z : C B · 5 1 } { S H K : K H · 6 0 0 } { B N : D T · 5 0 2 9
{ B : D Y · 9 5 1 8 } { B T F : G · 8 1 8 } { B : N M · 6 1 4 4 }
{ H D : S K · 9 1 } { H H : P Z · 1 2 0 4 } { M E : H Y · 1 1 0 }
{ G A P : K L · 3 1 2 } { B L K : A K · 1 5 6 } { M B : H X · 5 3 7
{ F R : N D · 2 6 9 } { P L : V J · 2 5 · } { P F : P · 6 1 1 0 }
{ E F : H T · 1 5 2 } { E F 0 6 2 2 1 } { F G : S K · 6 2 2 }
· Z I : A N · 5 9 · } { B T F : S · 7 6 4 } { E F : L E · 1 3 5 }
{ · A P : J D · 4 7 } { A P : K R · 1 3 } { B Z : F L · 5 5 7 }
} · L : H · 4 0 4 0 } { J : A E · 8 0 · } { E S W : L · 7 4 9 }
{ S O K : K · 7 4 7 } { B : E E · 1 5 9 1 } { W I : C · 6 6 2 9 }
{ P L : A T · 3 4 2 } { B : E H · 4 7 0 7 } { R S : W W · 5 5 · }
{ K S : H K · 1 7 5 } { 3 3 } U } K S · } { M T L : C V · 5 0 0 }
```

Objektorientierte RMM-Schnittstelle

Motivdetektion in DNA-Strängen

KFZ-Kennzeichen-Erkennung

Automatische Spracherkennung

Sprachdaten, Spracheinheiten, Zustände, Ausgabeverteilungen

RMM-Zustände zur Kennzeichensyntax

Buchstaben und Ziffern

P	«upper»	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Ö Ü
P	«digit»	1 2 3 4 5 6 7 8 9 0
P	«nonzero»	1 2 3 4 5 6 7 8 9

TÜV-Plaketten und Trennweiß

P	«space»	· #
S	«plaq»	· : ·

Ortsangabe und alphabetischer Schlüssel

D	«ort»	«upper» «upper» «upper»
D	«alfacode»	«upper» «upper»

Numerischer Schlüssel

D	«num123»	«digit» «digit» «digit»
N	«numcode»	«nonzero» «num123»

Neues EU-Kennzeichen (Deutschland)

S	«kennzeichen»	{ «ort» «plaq» «alfacode» «space» «numcode» }
---	---------------	---

Die Konfigurationsschritte des Erkennungssystems

- /series/nis National Institute of Standards Sprachsignale
- /series/par Partiturdaten: alle verfügbaren Transkriptionen
- /series/apl Akustisch-phonetische Transkription
- /series/nwl Normative Worttranskription
- /series/ufv Sprachdaten in Mel-Frequenz-Cepstrum/Ableitung
- /series/lfv MFCC mit Lauttranskription
- /rules/turns Dialoge, Dialogschritte und Transliterationen
- /rules/voc Trainingsdatenvokabular · Erkennungsvokabular
- /rules/pic Kontextabhängige Spracheinheiten („Polyphon“)
- /models/init Initiales Gesamt-RMM
- /models/super Überwachte Anpassung weniger Ausgabe-PDFs
- /models/mono Lernen der Monophon-Modelle
- /models/poly Lernen der Polyphon-Modelle $(F_{\min} = 1000)$
- /models/grow Lernen der Polyphon-Modelle $(F_{\min} = 100)$
- /models/Grow Lernen der Polyphon-Modelle $(F_{\min} = 10)$
- /models/test Erkennungstest auf Validierungsdaten

Die VERBMOBIL VM1-Dialogdaten

Maße und Gewichte

- 63 dt. Dialoge unterschiedlicher Sprecherpaare
 - 1 840 Dialogschritte (*turns*)
 - 39 560 gesprochene Wortvorkommen
 - 153 901 gesprochene Lautvorkommen
 - 1 461 930 Sprachframes zu 10 msec (4 Stunden 3½ Minuten)
 - 57 015 270 Cepstrumkoeffizienten bzw. Δ^1/Δ^2

Worteinheiten

- 2363 verschiedene Wörter

1373 ich + 1025 wir + 973 das + 857 ja + 783 dann + 681 und + 662 da + 556 es + 528 mir + 513 am

Lauteinheiten

- 46 verschiedene Laute (Phoneme)

305175 «p:» · 109606 n · 83101 s · 70664 t · 66253 m · 63490 a: · 60796 6 · 49834 a · 46644 al ·	34614 i: · 34173 @ · 33444 l · 33197 f · · 11720 g · 9033 n · 8246 h · 4508 Y · 4343 OY · 3254	Schnittstelle	Motivsuche	KFZ-Kennzeichen-Erkennung	Automatische Spracherkennung
oooooooooooo	oooooooooooo	oooooooooooo	oooooooooooo	oooooooooooo	oooooooooooo

Wortschätzte, Ziffernwörter, Buchstabieralphabet

C	«VOX»	«digit» «spell» «vm1» «VM1»
P	«digit»	null eins zwei drei vier fünf sechs sieben acht neun
P	«spell»	\$A \$B \$C \$D \$E \$F \$G \$H \$I \$J \$K \$L \$M \$N \$O \$P \$Q \$R \$S \$T \$U \$V \$W \$X \$Y \$Z
P	«vm1»	ab aber ablehnen absa acht achte «äh» «ähm» allerdings Allerheiligen als also am an anderen anderes zweitägiges zweite zweiten zweiundzwanzigsten zwischen
P	«VM1»	«%» a A \$A ab abend Abend Abenden Abendessen abends aber abfassen abge abgehakt abgehandelt abgeklär abgeklärt abgemacht zwo zwölf zwölfta zwölften zwölfter zwote zwoten zwoundzwanzigsten Zyklus Zylinder

Dialogturns und ihre Worttranskription

S	g071axx0-000-TIS	ja guten Tag dann fange ich einfach mal an und wollte Sie mal fragen wie das aussieht <ähm> wir müssten also insgesamt drei Arbeitssitzungen festlegen zwei davon müssten wir zweitätig machen <ähm> wann hätten Sie denn dafür mal Zeit
S	g071axx0-001-HAH	ja also für den eintägigen wenn wir den als erstes erledigen wollten quasi wäre mir ganz recht Montag der achte November
S	g071axx0-002-TIS	Montag der achte das sieht bei mir ganz schlecht aus denn die Woche vom Samstag dem sechsten an da bin ich weg <ähm> die Woche davor wie sähe das aus erste Novemberwoche
S	g071axx0-003-HAH	ja am ersten ist Allerheiligen da ist wohl nicht so sehr schön am zweiten Dienstag dem zweiten das wäre mir noch recht
S	g071axx0-004-TIS	<ähm> Dienstag würde mir gut passen <ähm> das heißt Moment allerdings erst nachmittags das wird dann wahrscheinlich ein bisschen schwierig Dienstag mittwochs <äh> is sieht das bei mir sch schwierig aus da habe ich tagsüber Termine <ähm> wie sieht das bei Ihnen am Donnerstag aus

Aussprachewörterbuch

Aachen	Q'a:x@n
Ab	Q'ap
Abbruch	Q'apbRUx
Abend	Q'a:b@nt
Abendöffnung	Q'a:b@nt9fnUN
Abendbrot	Q'a:b@nt#bRö:t
Abende	Q'a:b@nd@
Abenden	Q'a:b@nd@n
Abendessen	Q'a:b@nt#QEs@n
...
Kulturfreak	kUlt' u:6fri:k
Kulturrauptstadt	kUlt' u:6#h@uptStat
Kulturhochburg	kUlt' u:6ho:xbU6k
Kulturkalender	kUlt' u:6#kalEнд6
Kulturleben	kUltu:6#le:b@n
Kulturprogramm	kUlt' u:6#pro:gräm
Kulturveranstaltungen	kUlt' u:6fE6QanStaltUN@n
Kulturwochen	kUlt' u:6#vÖx@n
Kunde	k'Und@
Kunden	k'Und@n
...
zwoten	tsv'o:@n
zwoter	tsv'o:t6
zwoundvierzig	tsv'o:#QUnt#fI6tsIC
zwoundzwanzig	tsv'o:#QUnt#tsväntsIC
zwoundzwanzigste	tsv'o:#QUnt#tsväntsICst@
zwoundzwanzigsten	tsv'o:#QUnt#tsväntsICst@n
zwoundzwanzigster	tsv'o:#QUnt#tsväntsICst@6

Füllmuster und Stillemuster

Füllmuster

E [?]
L [?] * [?]
S /?/ [?] [?] [?]
L ?...? /?/
C FILLER [?] [?] * /?/ ?...?

Stillemuster

E [-]
L [-] * [-]
S /-/ [-]
S /-/ [-] [-]
S /--/ [-] [-] [-]
L /--*/ [-]
P /-?-/ /--/ /-/ /-/
S |-- /-?-/
S --| /-?-/
C SILENCE [-] [-] * /- /-/ /--/ /--*/ /-?-/ |-- --|

Strukturierung von Polyphonmodellen (8 648 Modelle)

Wortaufbau aus Polyphonen

S Abend /Q/a:b@nt Q/a:/b@nt Qa:/b/@nt Qa:b@/n/t Qa:b@n/t/
S Advent /a/t a/t/ t/v/E tv/E/n vE/n/t En/t/
S achtzehn /a/xtse:n a/x/ttte:n ax/t/tse:n axt/t/se:n axtt/s/e:n ts/e:/n tse:/n/
S

Polyphonaufbau durch Generalisierung

S @/n/d	@/n/
S b@/n/d@	@/n/d
S n/d/	/d/
S n/d@	n/d/
S @n/d@	n/d@
S @n/d/@n	@n/d/@
S d/@/	@/
S d/@/n	d/@/
S nd/@/n	d/@/n
S @nd/@/n	nd/@/n
S
S Qa/x/tQUnttsvantsICst	Qa/x/tQUnttsvan
S Qa/x/tQUnttsvantsICst@	Qa/x/tQUnttsvantsICst
S Qax/t/QUnttsvantsICst	Qax/t/QUnttsvan
S

Strukturierung von Monophonmodellen

P MONOPHONE /@/ /2:/ /6/ /9/ /a:/ /a/ /b/ /C/ /d/ /e:/ /e/ /E:/ /E/
/f/ /g/ /h/ /i:/ /I/ /j/ /k/ /l/ /m/ /n/ /N/ /o:/ /O/ /p/ /Q/
/r/ /s/ /S/ /t/ /u:/ /u/ /U/ /v/ /x/ /y:/ /Y/ /z/ /Z/

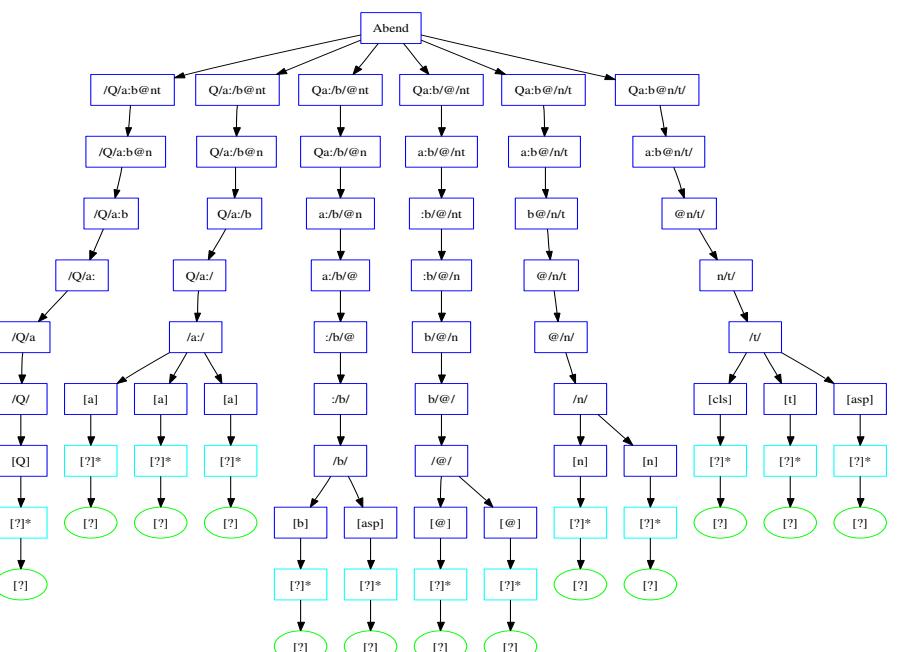
Subphonenzustände

S [d] [?] *
S [e] [?] *
S [E] [?] *
S [f] [?] *
S [cls] [?] *
S

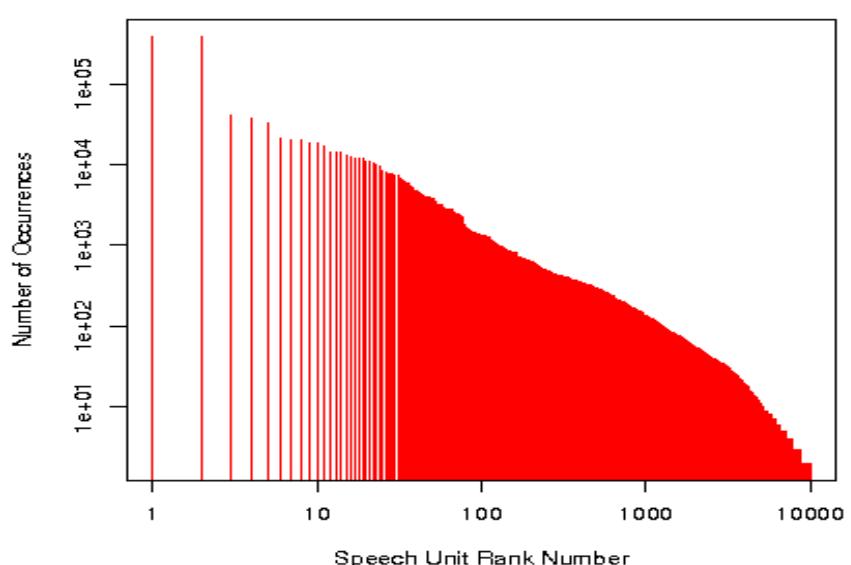
Monophon Zustände

S /d/ [d] [asp]
S /e:/ [e] [e] [e]
S /e/ [e] [e]
S /f/ [f] [f]
S /h/ [h] [h]
S /k/ [cls] [k] [asp]
S

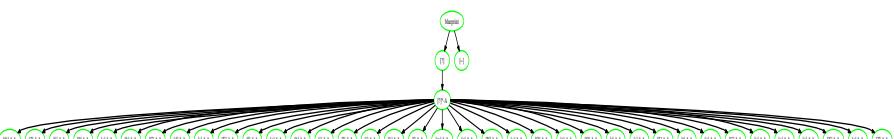
RMM-Struktur des Zustands 'Abend'



Spracheinheiten und ihre Korpushäufigkeit



Überwachtes Lernen der subphonemischen Zustände



Phonemische Ausrichtung via Bootstrap-Erkenner
Königsdichte 'blueprint' · Stilledichte '[-]' · Sprachdichte '[?]',

226648	a	219212	n	166202	s	141328	t
132506	m	121592	6	105288	E	93288	aI
69228	i	68346	Ø	66888	I	66394	f
64858	e	60552	d	57118	C	45756	v
45708	Q	44610	l	42746	x	42488	z
40640	o	37984	k	35334	aU	33012	U
32968	Ø	29446	u	27568	b	27520	N
26882	r	26016	j	25070	S	23440	g
18066	p	16492	h	9016	Y	8686	OY
6508	9	4644	y	3318	2	108	Z
36	aN						

Entscheidungsüberwachtes Baum-Welch-Training

Monophonzustände

42 neue Spracheinheiten

157 elementare Zustände

13 272 Zustände

Polyphonzustände ($F_{min} = 1000$)

315 neue Spracheinheiten

262 elementare Zustände

13 733 Zustände

Polyphonzustände ($F_{min} = 100$)

1 903 neue Spracheinheiten

3 211 elementare Zustände

26 380 Zustände

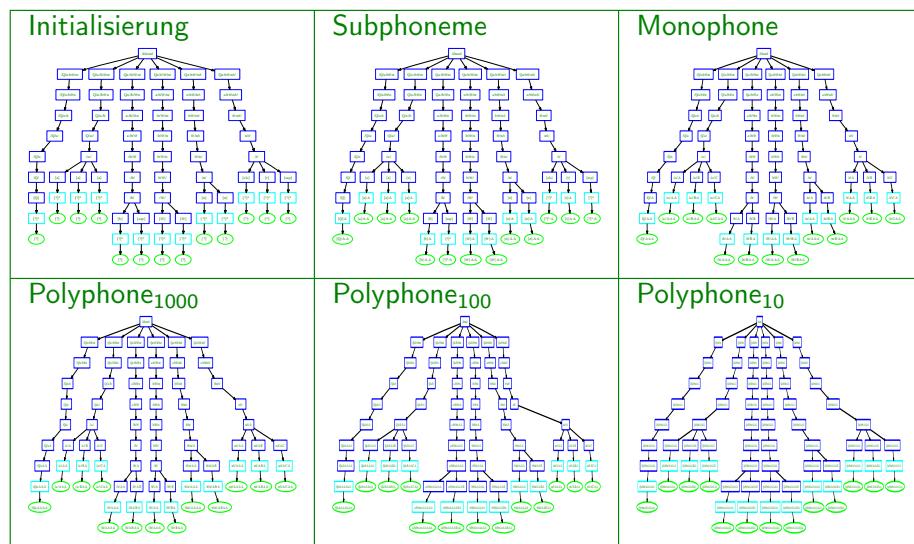
Polyphonzustände ($F_{min} = 10$)

15 622 neue Spracheinheiten

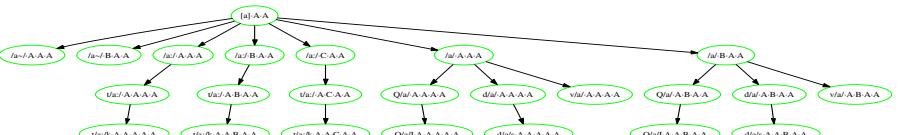
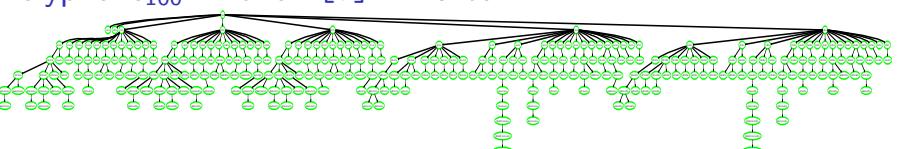
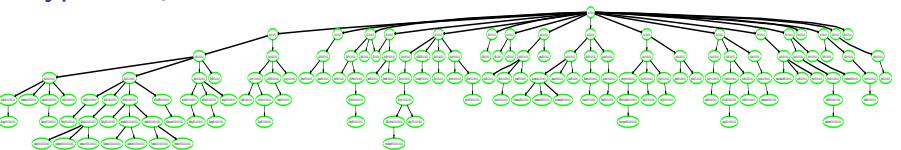
15 898 elementare Zustände

87 734 Zustände

Wortmodelle für das Wort 'Abend'



Elementarzustände für das Phonem '/a:/'

Polyphone₁₀₀₀ — alle '/a/'-DichtenPolyphone₁₀₀ — alle '/a/'-DichtenPolyphone₁₀ — nur '/a:/·A'-Dichten

Worterkennung ohne linguistisches Modell

vm1/g071axx0020TIS (712 Frames)

- | | | | | |
|------|-----------------|----------|---------|---------|
| [1] | ' --' | 'Hut' | 'ja' | 'da' |
| [5] | 'bin' | 'ich' | 'voll' | 'mit' |
| [9] | 'einverstanden' | 'is' | 'meine' | 'warum' |
| [13] | 'sollte' | 'sch' | 'das' | 'noch' |
| [17] | 'ablehnen' | 'durf' | 'Hut' | 'Hut' |
| [21] | 'belassen' | 'das' | 'dabei' | '«häs»' |
| [25] | 'darf' | 'vielen' | 'Dank' | 'der' |
| [29] | 'ne' | '-- ' | | |

Dekodierwahrscheinlichkeiten

$$\log P(x|S) = -53964.3 = -53929.8 - -72.7486 + -107.29$$