Fragenkatalog zur Vorlesung „Automatic Speech Recognition“

Einführung

1. Nennen Sie einige Anwendungen für Automatische Spracherkennung!
   - Speech Recognition in Windows 8
   - Voice-directed warehousing (“Voice Picking”)
   - IVR - Interactive Voice Response
   - Deutsche Bahn Timetable Information System
   - Spoken Dialogue System - Berti
   - Voice-controlled Navigation
   - Siri
   - Google Glass
   - Amazon Echo (Alexa)
   - Semi-automatic video Subtitling
   - Allg: Medizin, Ausbildung, Behindertenhilfe, Übersetzung, Militär, Industrie & Logistik

2. Warum ist Spracherkennung schwierig?

Variabilität	**Ambiguität**
• Gleiche Wörter können sehr unterschiedlich klingen (abhängig vom Alter, Geschlecht, Pitch, Akzent…) • Eine gewisse Bedeutung kann mit ganz versch. Wörtern und Phrasen vermittelt werden.	• Das Verständnis einer Äußerung erfordert oft Hintergrundwissen, Wissen über den Dialogkontext, Kenntnisse über den Dialogpartner, etc. • Auch Interaktionen zwischen Menschen führen oft zu Missverständnissen (z. B. Ironie oder nicht?). • Viele Witze basieren auf sprach. Ambiguität.

3. Welche Gründe gibt es für den Einsatz von Automatischer Spracherkennung?
   - Hände und Augen freie Interaktion (im Auto)
   - Uneingeschränkte Bewegungsfreiheit der Arme, Hände und Beine (Voice Picking)
   - Extrem geringer Platzbedarf. (Kleine Geräte wie Apple Watch oder Google Glass lassen wenig Raum für alternative Benutzeroberflächen)
   - Die Spracheingabe kann in Kombination mit anderen Eingabegeräten extrem effizient sein
   - VUIs (Voice User Interfaces) können in vollständiger Dunkelheit verwendet werden
   - VUIs können Menschen mit Behinderungen helfen Computer zu bedienen

4. Warum ist gesprochene Sprache nicht immer das geeignetste Mittel, um mit Computern zu interagieren?
   - Beispiel: Computer soll Programm ausführen
     - Mit der Maus: Weniger Klicks (Hand + Augen Koordination)
     - Mit Spracheingabe: Direktion mündlich sehr aufwendig
   - Voice User Interfaces (VUI) können die Privatsphäre des Benutzers verletzen und können für andere Personen nerven, z.B. Im Büro, in den Öffis oder in einem Restaurant.
   - Naive Menschen denken sie interagieren mit intelligenten Maschinen → Fehlgeschlagene interaction und frustrierte Nutzer

Phonetische Grundlagen

5. Wie kann man sich die menschliche Sprachproduktion vorstellen?
   - Lunge generiert Strom an Luft(druck)
   - Stimmapparat (Stimmbänder) im Kehlkopf schwingen mit Grundfrequenz (<180 Hz männlich, > 180 Hz weiblich)
   - Nase- und Mundraum, sowie Zunge, Lippen modulieren Luftstrom zu gewünschten Klang

6. Wie funktioniert die Schallwahrnehmung im menschlichen Ohr?
   - Aussenohr: Luftkanal, Trommelfell schwingt, wandelt Schalldruck in mechanische Schwingung um
   - Mittelohr: Hammer, Amboss, Steigbügel: Übertragen/Verstärken Bewegung auf Membran in Gehörschnecke
   - Innenohr: Membran überträgt Schwingungen auf Flüssigkeit, verteilt Schwingung in Gehörschnecke, Basilar Membran hat Empfindlichkeit gegen verschiedene Frequenzbereiche an verschiedenen Stellen, wird durch Schwingungen in Flüssigkeit angeregt, kleine Haar-Zellen regen Nervensignale an
   - Mensch kann zwischen 20 und 20kHz Signale hören
   - Teilt Akustisches Signal in 24 verschiedene Bänder
7. Nach welchen beiden messbaren Größen kann man die Vokale recht gut unterscheiden?
   - Die ersten beiden Formanten und deren Frequenzverlauf

8. Was ist ein Phonem?
   - Kleinste Klangeinheit (=Laut), die zwei Wörter einer Sprache unterscheiden kann. Basiert auf Linguistik, nicht Akustik. Ein Phonem kann mehreren Buchstaben/Wortteilen zugeordnet werden

“Example: In German, [χ] (Rache), [x] (Kuchen), [ç] (Milch) are different realizations of the same phoneme /x/.”

9. Was ist ein Allophon?
   - Menge von gesprochenen Klängen, die ein Phonem bilden. Einzelnes Allophon kann mehrere Phoneme erzeugen

“Example: In German, the allophone [χ] can be used in Rache (phoneme /x/) and in Kragen (phoneme /r/)”

10. Was versteht man unter Koartikulation?
    - Angrenzende Phoneme/Buchstaben klingen in aktuellen Laut mit, aktueller Laut wird aufgrund angrenzender Laute verändert (meist weil einfacher auszusprechen, z.B. “dt” wird nicht einzeln gesprochen)

11. Wozu dient Prosodie bzw. Intonation?

Prosodie: zusätzliche Information durch Betonung (über die Akustik)
   - Akzentuierung: untersch. Betonung eines Worts, kann untersch. Bedeutung hervorrufen
   - Satzart: Aussage vs. Frage vs. Wiederholung (zB am Telefon)
   - Betonung / Tonfall: Pausen und Wortdehnung, Tonhöhe, Melodie und Rhythmus, Druck beim Reden

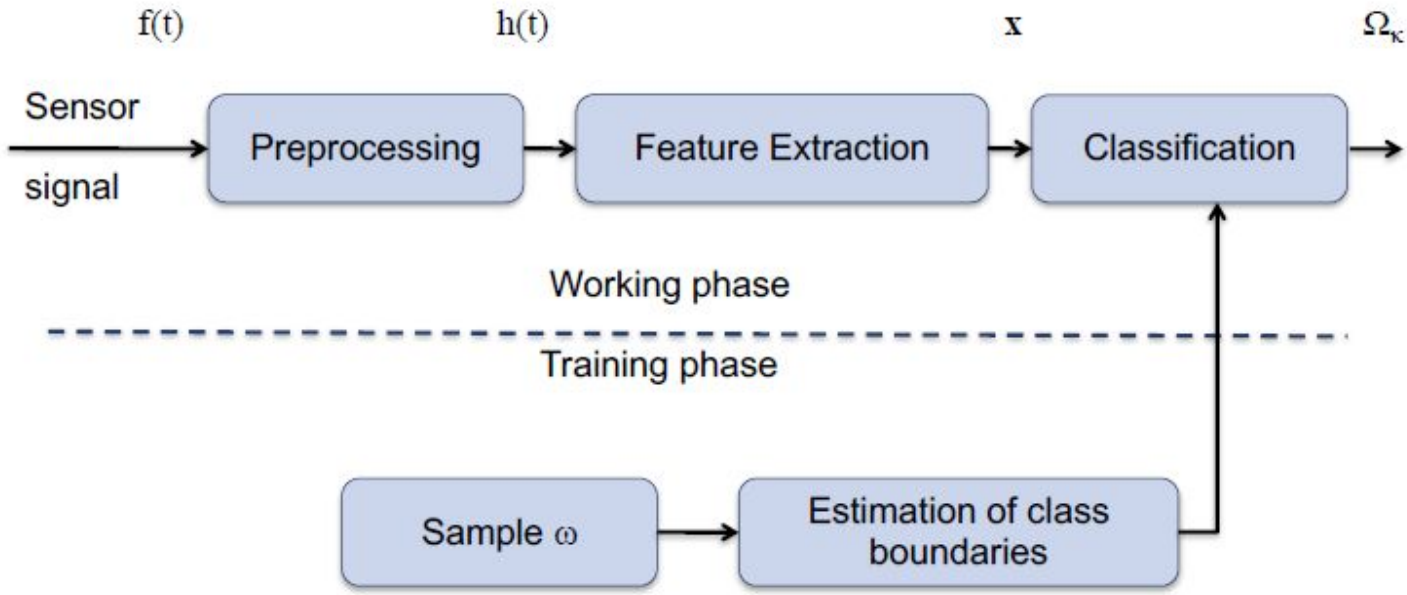
Mustererkennung

12. Womit beschäftigt sich das Gebiet der Mustererkennung?
    - Wir definieren die Mustererkennung als den Prozess der automatischen Umwandlung eines Sensorsignals in eine aufgabenspezifische symbolische Beschreibung.

13. Was versteht man unter der Klassifikation einfacher Muster? Nennen Sie Beispiele?
    - Das Muster ist Reperäsentant eines Objekts der realen Welt
    - Jedes Objekt kann genau einer Klasse (Anz. endlich) zugeordnet werden

- Beispiele:
  - Gesichtserkennung
  - Spracherkennung (geschrieben, wie auch gesprochen)

14. Beschreiben Sie den grundsätzlichen Aufbau eines Klassifikationssystems!



Digitalisierung und Merkmale

15. Welche grundlegenden Entscheidungen müssen getroffen werden, bevor ein analoges Signal digitalisiert wird?

- Samplingrate (um Aliasing zu vermeiden: ½ T, bzw. doppelte Frequenz 2f)
- Quantisierung (endliche Abstufung der Amplitude)

16. Was besagt das Abtasttheorem? Beispiel?

- Samplingrate muss min. die doppelte Frequenz (2*f) bzw. die Hälfte einer Periode (1/2T) betragen um ein kontinuierliches Signal korrekt wiederzugeben (Aliasing vermeiden!)
- Bsp: 2000 samples / second → signal < 1000Hz

17. Welche Form hat i.d.R. die Kennlinie eines mit 8 Bit quantisierten Signals, und warum?

Werte sind in der Sprache exponentiell verteilt ⇒ logarithmische Verzerrung ist effizienter:

- Auflösung kleiner Amplituden erhöht
- Bereiche darüber werden komprimiert
- Gleichverteilung der Werte
- Klingt besser (Menschen empfinden hohe und tiefe Signale als Gutklingend, die Mitten Frequenzen werden oft als störend empfunden)

18. Worin liegt der Vorteil eines mit 8 Bit quantisierten Signals gegenüber einem mit 16 Bit quantisierten Signal?

- 8 Bit Signal mit logarithmischer Skala (Telefon, A-law, u-law) hat gleichverteilten Wertebereich (Da hohe Amplituden seltener vorkommen und diese bei logarithmischer Skala zusammengefasst werden) und somit bessere Qualität (da genauer in niedrigen Amplituden als lineare Skala) als 8-bit linear
- Auf 8 Bit quantisiertes Signal benötigt nur halb so viel Speicher (und Übertragungsrate) als 16 Bit Daten

19. Wie entsteht der sogenannte Leck-Effekt und wie lässt er sich reduzieren? Beispiel?

- Entsteht zB. beim ausschneiden eines Fensters und Anwendung der DFT (dafür wird ein kontinuierliches Perioden Signal benötigt → beim “aneinanderhängen” der Perioden kann es vorkommen, dass diese keine “weichen, perfekten” Überhänge haben, sondern Kanten → Leck-Effekt -> Einstreuung von vorher nicht vorhandenen Frequenzen)
- Reduzierbar durch anwenden eines Hamming-Fensters, welches die Enden des jeweiligen Perioden-Signals gegen 0 drückt (also abschwächt)

20. Welche typische Fenstergröße nutzt man in der automatischen Spracherkennung? Was wären die Vor- und Nachteile breiterer bzw. schmalerer Fenster?

- **Wideband spectrograms:**
  - low frequency resolution, high temporal resolution
  - **vertical lines** indicate individual **pitch periods**
  - useful for analyzing the temporal structure of plosives
  - useful for determining **formants**, which are characteristic of certain vowels
- **Narrowband spectrogram:**
  - high frequency resolution, low temporal resolution
  - **horizontal bands** indicate the **harmonics** (positive integer multiples) of the fundamental frequency (perceived as pitch), e.g. at 150 Hz, 300 Hz, 450 Hz, 600 Hz, …
  - useful for determining and tracking the fundamental frequency, e.g. for sentence mode classification

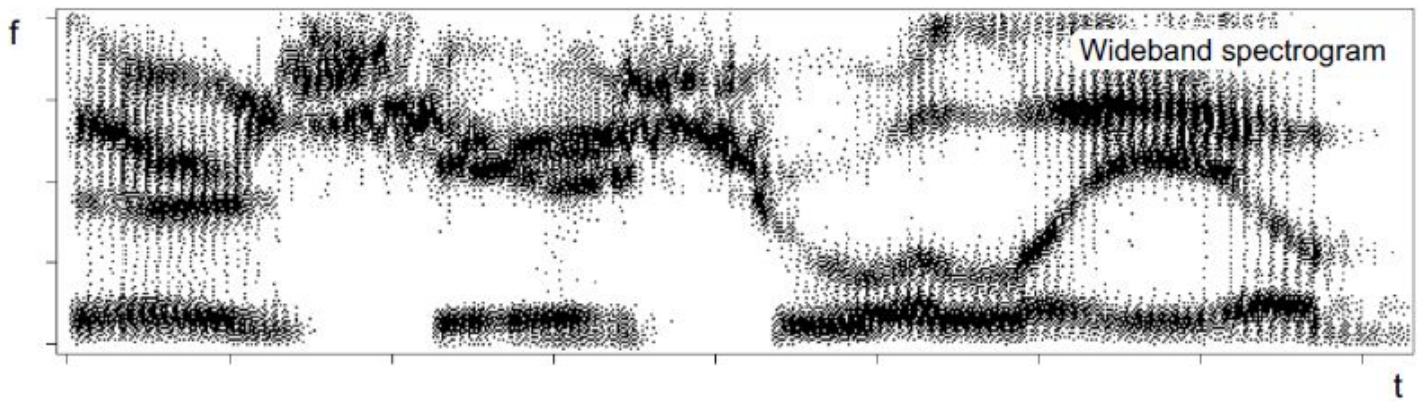
???

- **Typisch sind 256 Abtastwerte; geschoben wird um 10ms pro Fenster**
- **25,6 ms wenn (Abhängig von Abtastrate)**

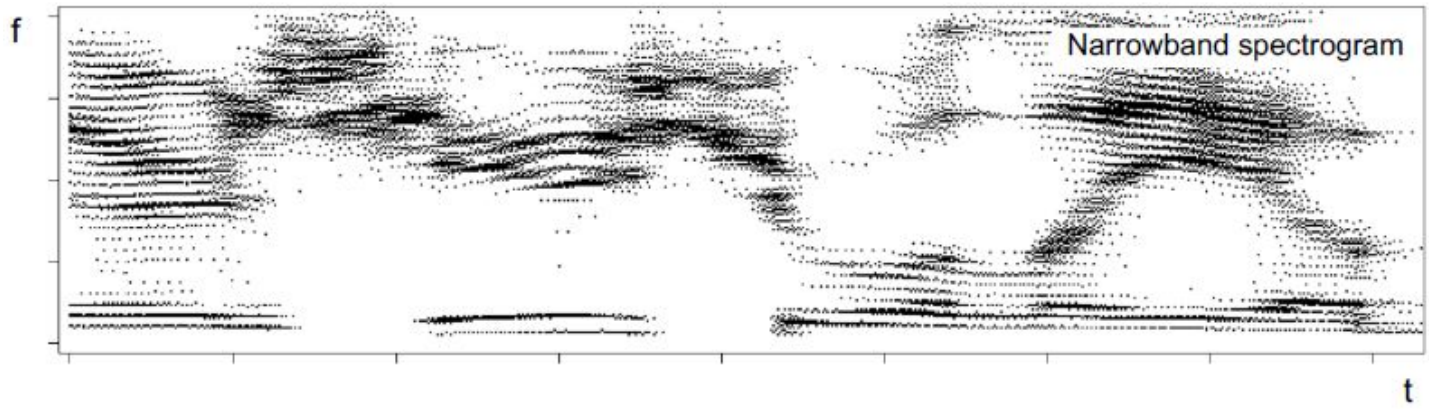
	Vorteil	**Nachteil**
Schmales Fenster 64 Samples Wideband	Hohe Zeitauflösung	geringe Frequenzauflösung
Breites Fenster 512 Samples narrowband	Hohe Frequenzauflösung	schlechte Zeitauflösung -> “Verwischen” der Werte → evtl.Plosives (Explosivlaut ‘t’, ‘p’,...) schwierig zu erkennen

21. Was ist der Unterschied zwischen einem Breitband- und einem Schmalbandspektrogramm und wozu nutzt man diese?

- Breitband: Erkennung einzelner Phoneme bzw. der Formanten derer (besser für Spracherkennung)



- Schmalband (narrowband): “Zählen” der Grundfrequenz sehr einfach (horizontale Linien) z.B. für Notenerkennung und Pitch Verlauf also Fragesatz Aussagesatz etc...

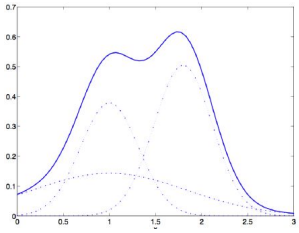


22. Welche Merkmale werden in der automatischen Spracherkennung überwiegend eingesetzt? Wie errechnet man sie?
    - Mel Frequency Cepstral Coefficients (MFCCs): Kombination der Vorteile der Mel Filterbank + cepstrum
    - Audio Signal -> Sample Fenster ausschneiden-> Hamming Fenster Anwenden -> FFT -> mel filter bank -> mel frequency coefficients ;
    - mel filter bank -> log -> Cosinus Transf. -> Cepstrum
23. Welche Merkmale ergänzt man, um den zeitlichen Verlauf der MFCCs besser zu erfassen? Wie errechnet man sie?
    - MFCC's (13 Statische Merkmale) um Dynamische Merkmale (dynamic features) erweitern
    - Steigung jedes einzelnen Wertes eines MFCC Vektors berechnen = 1. Ableitung -> +13 neue Merkmale
    - 2. Ableitung -> +13 neue Merkmale
    - Ergibt dynamischen Merkmalsvektor mit 39 Werten (13 statische + 26 dynamische)
    - Dyn. Merkmale zur Vorhersage der Entwicklung der Merkmale im Zeitverlauf

Klassifikation

24. Welche Verfahren zur Klassifikation von Merkmalvektoren kennen Sie und wodurch zeichnen sich diese aus?
    - Nicht-parametrische: Verwenden ganzen Trainingsdatensatz (z.B. Vergleich mit jedem Trainingswert), z.B. Nearest Neighbor
    - Verteilungsfreie: Explizite Grenzen zwischen Klassen, z.B. Linear, Support Vector Machine, Neuronale Netze !
    - Probabilistische: Dichtefunktion bestimmt Wahrscheinlichkeit, dass Test-Sample zu einer Klasse gehört, Gauss
    - Neuronal Netze (siehe Folie 150) ja hat er nur extra aufgeführt weils ein extra thema sein sollte gehören aber zu Verteilungsfreie Klassifikatoren (wurde bei Besprechung Fragenkatalog erwähnt) internet machts möglich: es gibt auch Probablistische Neuronale Netze allerdings hat er die nirgends erwähnt
25. Beschreiben Sie den Nächster-Nachbar-Klassifikator!
    - Es wird für jedes Sample im Trainings-Datensatz der (z.B: Euklidische) Abstand zum Test-Sample berechnet, und die Klasse vom Trainings-Sample mit dem geringsten Abstand übernommen
26. Welche Formel ist bei der statistischen Klassifikation von zentraler Bedeutung? Erläutern Sie diese?
    - Bayes Formel (Erklären können für Vektoren, Wörter, einzelne Merkmale...)
    - $$p(\Omega_K|X)=\frac{p(\Omega_K)\cdot p(X|\Omega_K)}{p(X)}$$ (Hier: Wahrscheinlichkeit, dass X zur Klasse Omega gehört, berechnet sich aus a priori Wahrscheinlichkeit der Klasse Omega * Wahrscheinlichkeit von X in Abhängigkeit von Omega durch Grundwahrscheinlichkeit von X über alle Klassen)

27. Welches Verfahren kennen Sie, mit denen man aus Stichproben von Merkmalsvektoren unüberwacht Kodebücher schätzen kann?
    - kMeans
    - EM-Algorithmus
28. Wie funktioniert der k-Means bzw. LBG-Algorithmus?
    1. (Zufällige) Auswahl von k Feature Vektoren als Mittelpunkte
    2. Zuordnung aller Feature-Vektoren zu Mittelpunkt mit kleinstem Abstand
    3. Neu-Berechnen der Mittelpunkte/Standardabweichung innerhalb einer Klasse
    4. Wiederhole 2-3, bis Aufteilung stabil
29. Was versteht man unter einer Gaußschen Mischverteilung?
    - Alle Feature-Vektoren werden aus einer Kombination verschiedener Gauß-Verteilungen generiert (z.B. Feature F wurde zu 60% von G1, 30% von G2 und 10% von G1 produziert)



30. Wie funktioniert der EM-Algorithmus zu Kodebuchschätzung?
    1. Initialisiere Mittelpunkte, Standardabweichungen (zufällig) aus Feature-Vektoren, ähnlich k-Means
    2. Expectation-Step: Berechne für jeden Feature-Vektor für jede Klasse eine Gewichtung (die Wahrscheinlichkeit der Klasse abhängig vom Feature-Vektor)
    3. Maximization-Step: Mittelpunkte/Standardabweichungen, sowie Wahrscheinlichkeiten der Klassen aktualisieren. Darin gehen ALLE Feature-Vektoren abhängig von den vorher berechneten Gewichtungen ein
    4. Wiederhole 2-3 bis zu Abbruchkriterium

Deep Learning

1. Was ist ein Perzeptron?
    - Vereinfachte Simulierung eines einzelnen Neurons (Rosenblatt 1957)
    - ein Perzeptron hat als Eingang die Gewichte w und Werte x der mit ihm verbundenen Perzeptronen
    - Als Ausgang wird die Summe aller eingehenden Werte mal deren Gewichte berechnet und anschließend ein Bias b abgezogen. $f = \sum_{0}^{i} (w_i * x_i) - b$

2. Wie sehen die Schwellwert-Funktionen bei künstlichen neuronalen Netzen aus?
    - Standard: Stufe 0 auf 1 -> schlecht da keine Zwischenwerte
    - daher Sigmoid-Funktion (monoton, kontinuierlich, differenzierbar) verwendet

3. Was versteht man unter einem Feed-Forward-Netzwerk?
    - Verbindung nur in eine Richtung (keine Rekursion, vgl. RNN - recurrent neural networks)

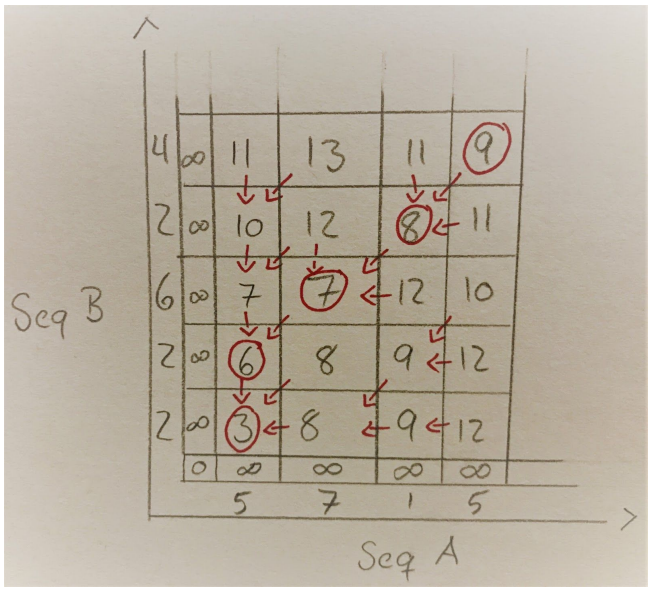
4. Was versteht man unter einem MLP?
    - Multilayer Perzeptron: NN organisiert in Schichten auch als DNN Bezeichnet

5. In welcher Weise wirkt sich die Verwendung von neuronalen Netzen auf die Wahl von geeigneten Merkmalen für die Spracherkennung aus?
    - Aufbereitung der Merkmale wird stetig minimiert → NN lernt besser und schneller als Mensch
    - Trend: Immer “rohere” Daten als Input; Direkt daten aus des FFT Spektrums verwenden.

Dynamic Time Warping

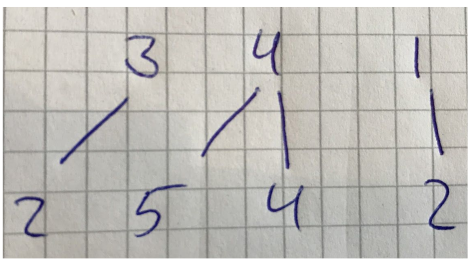
6. Wozu dient der DTW-Algorithmus?
   - Klassifikation/Erkennung von einzelnen Wörtern

7. Erläutern Sie den DTW-Algorithmus!
   1. Seq A = x-Achse (Testwort Werte); Seq B = y-Achse (Trainingswort Werte)
   2. Jeweils erste Spalte und erste Zeile mit ∞ initialisieren und Ursprung hat Werte 0
   3. Distanz der Gegenüberstehenden Zahlen berechnen (hier als erstes | 2 - 5 | = 3) + min(links, links-unten, unten))



8. Wie erhält man die zeitliche Zuordnung zwischen dem Test- und dem Referenzsignal? Rechnen Sie ein kurzes Beispiel durch, z.B. d(3-4-1, 2-5-4-2) .

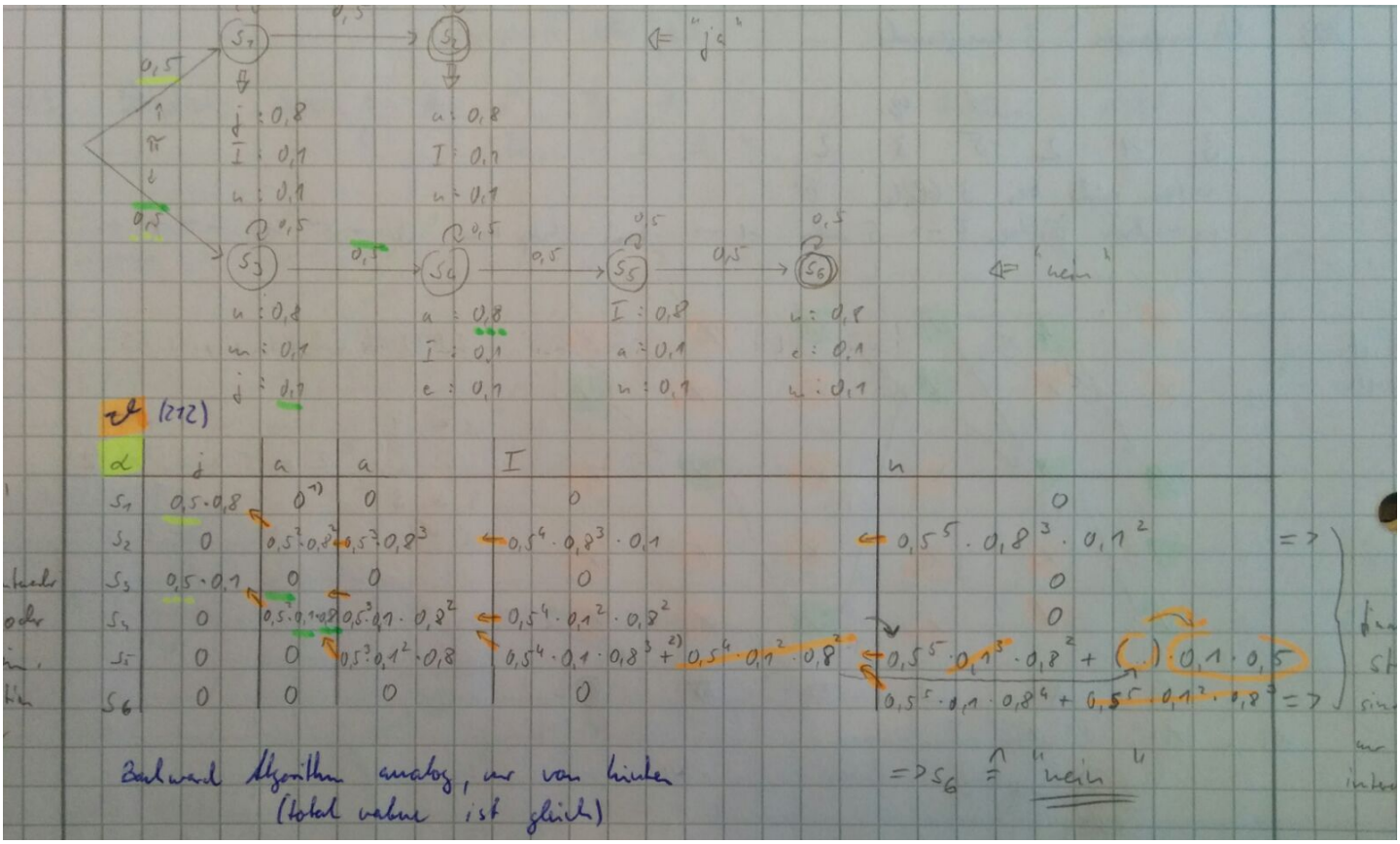
1	∞	4	6	5	**3**
4	∞	3	**2**	**2**	4
3	∞	**1**	3	4	5
	0	∞	∞	∞	∞
		2	5	4	2



Hidden-Markov-Modelle

9. Welche Parameter besitzt ein diskretes HMM?

- π: Die Startwahrscheinlichkeiten der Anfangspfade
- A: Die Übergangswahrscheinlichkeiten (zwischen allen Zuständen)
- B: Die Ausgabewahrscheinlichkeiten (für jeden Zustand und jedem generierten Symbol bzw Phonem)



10. Welche Arten von HMMs haben wir noch kennengelernt und worin unterscheiden sich diese?
    - DNN-HMM: Ausgabewahrscheinlichkeiten B werden durch Neuronales Netzwerk bestimmt
    - GMM-HMM: Ausgabewahrscheinlichkeiten B werden durch Gauß Mixture Model bestimmt

11. Welche HMM-Topologien sind für die Spracherkennung geeignet?
    - Left-to-Right: Von einem Zustand aus kann in alle Zustände rechts davon über gegangen werden -> überspringung überall möglich -> schlecht für lange wörter -> buchstaben werden ausgelassen
    - Bakis-Model: Jeder Zustand hat Übergänge zum nächsten und übernächsten (d.h. es kann immer ein Zustand übersprungen werden) hier ist es einfach haben oder habn zu erkennen.
    - Linear: jeder Zustand hat nur Übergänge zu direkten Nachfolgern (keine Auslassung möglich) -> haben und habn muss dann in der Verteilungsdichte funktion des jeweiligen zustands modelliert sein. Ansonsten wird habn nicht erkannt (verkürzung)

12. Was versteht man unter der Produktionswahrscheinlichkeit und wie lässt sich diese naiv errechnen?
    - Die Wahrscheinlichkeit P(X|λ), dass X aus diesem HMM generiert wurde
    - Berechnen: Summe über alle (Anfangs-Pfade * das Produkt aller Übergangs- und Ausgabewahrscheinlichkeiten, die das Wort bilden)
    - $$P(X|\lambda) = \sum_{q \in Q^T} P(X, q|\lambda) = \sum_{q \in Q^8} \pi_{q_1} \cdot b_{q_1}(x_1) \cdot \prod_{t=2}^{T} a_{q_{t-1} q_t} \cdot b_{q_t}(x_t)$$
    - Oder einfach: Berechnen aller Pfade, die das gesuchte Wort abbilden können (brute Force)
    - (siehe Bild unten Frage 9)

13. Wie ist die Grundidee eines effizienten Algorithmus zur Bestimmung der Produktionswahrscheinlichkeit?
    - ?? Forward-Algorithmus mit Alpha-Matrix <-

- (Oder: Backward-Algorithmus mit Beta-Matrix (beide sind gleichbedeutend) : geht aber nur wenn die gesamte Wortinformation schon vorhanden ist also man schon rückwärts rechnen kann weil die Folge der Buchstaben schon da ist;)

14. Wozu dient der Viterbi-Algorithmus und worin besteht seine Grundidee?
    - dient zur Berechnung der wahrscheinlichsten Zustandssequenz durch Maximierung anstatt der Bildung der Summe im Rekursionsschritt (basierend auf Forward-Algorithmus)
    - Gleiche wie Forward nur dass nur der Zustand mit der höchsten Wahrscheinlichkeit in den Pfad kommt (also kein + wenn man von 2 Zuständen in den jetzigen kommen kann sondern einfach der Zustand mit der höheren Wahrscheinlichkeit wird verwendet)
15. Beschreiben Sie die Grundidee der Schätzung von HMM-Parametern anhand einer Beobachtungsfolge (Stichprobe)?
    - Baum-Welch Training (Kombination des Forward und Backward Algorithmus):
        - Alpha-Matrix berechnen
        - Beta-Matrix berechnen
        - iterativ vorgehen

16. Was versteht man unter einer Maximum-Likelihood-Schätzung?
    - zB bei Baum-Welch-Verfahren:
        - versucht Parameter zu finden, welcher die max. Wahrscheinlichkeit bei den Trainingsdaten erhält
        - versucht Parameter zu finden, der die Wahrscheinlichkeit maximiert, dass die Stichprobe vom Modell erzeugt wurde -> den Parameter bestimmen, der die Trainingsstichprobe maximal wahrscheinlich macht
        - → Gefahr: Zu sehr auf die Trainingsdaten angepasst

17. Welche Wortuntereinheiten nimmt man i.d.R. für die Spracherkennung? Erläutern Sie die Idee dahinter!
    - Triphone:
        - keine (!) drei aufeinanderfolgende Laute!!
        - sondern: einzelner Laut, beeinflusst von beiden Nachbarlauten (Im Context)
        - zB “sieben” klingt wie “sie[bn]” : phoneme I hat vorgänger s und nachfolger e -> HMM mit 3 States pro phoneme!
        - Für jedes Triphone gibt es ein Sub-HMM welches wiederverwendet werden kann um größere HMM’s für Wörter zu erstellen
        - Die Sub-HMM’s werden relativ robust trainiert da die triphone in verschiedenen wörtern also mehrmals innerhalb des vokabulars vorkommen

Sprachmodelle (Language Models)

18. Was versteht man unter einem N-Gramm?
    - Tupel aufeinanderfolgender Wörter mit N Elementen

19. Wie erhält man ML-Schätzwerte für N-Gramme?
    - Zähler: Anzahl der vorgekommen Kombination
    - Nenner: Anz. des vorgänger Wortes (ohne nächstes Wort)
    - **Beispiel**: #(“to Chicago”) = 2; #(“to”) = 4 → P^ (“Chicago” | “to”) = 2 / 4 = 0,5 = 50%

20. Warum sind diese in der Praxis von Nachteil?
    - Nicht existente Kombination werden nie erkannt (P = 0)!
        - zB im Text kommt kein “to Boston” vor: #(“to Boston”) = 0; #(“to”) = 4 → P^ (“Boston” | “to”) = 0 / 4 = 0%

21. Wie lassen sich die Schätzwerte glätten?
    - Laplace Glättung: Vokabulargröße L (# unterschiedlich. Wörter im Text) mit in den Nenner addieren, sowie Zähler plus 1 rechnen → nie mehr P = 0; reduzierung von P für tatsächliche Wortkombinationen
    - Jeffrey: wie Laplace, nur mit ½ und L/2
    - Besser: Backoff smoothing

22. Welche weitere Möglichkeit gibt es, die Parameterzahl zu reduzieren?
    - Wortkategorien: Kombinationen von Wörtern, welche in ähnlichen grammatikalischen Kontexten verwendet werden können (zB Personen-Namen, Städtenamen, Wochentage, Nummern)

23. Wie errechnet sich die sog. Test-Set-Perplexität einer Stichprobe, gegeben ein Sprachmodell?
    - 10 Ziffern Vokabular;
    - w = “eins fünf drei zwei fünf”; P(w) = 1/10 * 1/10 * 1/10 * 1/10 * 1/10 = 1 / 10^5 (10 = versch. Mögl.)
    - PP(w) = (1 / 10^5) ^ (- ⅕) = 10
        - m = # Wörter in w

$$PP(\mathbf{w}) = P(\mathbf{w})^{-\frac{1}{m}} = \frac{1}{\sqrt[m]{P(\mathbf{w})}}$$

24. Wie kann man diese interpretieren?
    - Je geringer der Wert, desto Wahrscheinlicher, dass ein Wort produziert wird

25. Wie kann man mit Sprachmodellen Themen (topics) klassifizieren?
    - Topic Klassifikation:
        - mit themenspezifischen Texten trainieren (Modelle generieren)
        - Test-Set Perplexität für alle Themen-Modelle berechnen → kleinste PP(w) gewinnt

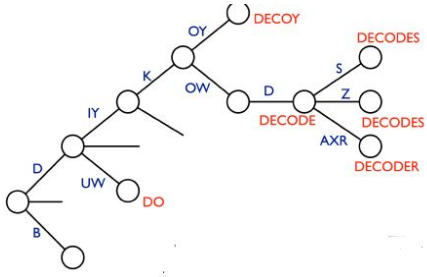
26. Wie kann man mit Sprachmodellen Sprachen klassifizieren?
    - Trainieren mit Buchstaben in versch. Sprachen (ergibt n-Grams für Buchstabenkombinationen)

Dekodierung

27. Was ist ein konfluenter Zustand (confluent state)?
    - Extra Zustand in HMMs, ohne Ausgabe von Symbol, wird verwendet, um HMM von vorne neu zu beginnen -> Erkennung von aneinander gereihten Wortteilen

28. Was versteht man unter Beam Search (Strahlsuche)?
    - Für bessere Effizienz werden bei jedem Schritt unwahrscheinliche Pfade verworfen (Gefahr: Ausschluss von evtl. passenderen Pfaden schon zu Beginn da erstes Teilwort nicht passt / falsch erkannt wurde)

29. Wie kann man bei sehr großen Wortschätzen den Wortschatz sinnvoll organisieren?
    - Prefix Pronunciation Tree: Baumstruktur nach aufeinanderfolgenden Phonemen, logarithmische Suchzeit



30. Welche weitere Methode zur Beschleunigung des Dekodiervorgangs haben wir kennengelernt?
    - Mehrphasen Dekodierung
        a. bi-Gram Graph: Gitterstruktur (Word Lattice), wenige Pfade
        b. tri-Gram Sprachmodelle

31. Welche beiden heuristischen Parameter führt man ein, die von der reinen Lehre der Bayes-Formel abweichen?
    - Insertion Penalty: Faktor, der die Wahrscheinlichkeit von Wortgrenzen verringern soll, um längere Wörter zu erzwingen
    - Language Model Gewichtung: Erhöht den Einfluss des Sprachmodells auf das Ergebnis