

Fragenkatalog zur Prüfungsvorbereitung zur Vorlesung „Automatic Speech Recognition“ WS2017/18

Einführung

1. Nennen Sie einige Anwendungen für Automatische Spracherkennung!
2. Warum ist Spracherkennung schwierig?
3. Welche Gründe gibt es für den Einsatz von Automatischer Spracherkennung?
4. Warum ist gesprochene Sprache nicht immer das geeignetste Mittel, um mit Computern zu interagieren?

Phonetische Grundlagen

5. Wie kann man sich die menschliche Sprachproduktion vorstellen?
6. Wie funktioniert die Schallwahrnehmung im menschlichen Ohr?
7. Welcher Frequenzbereich ist für Menschen hörbar?
8. Wodurch entsteht die Sprachgrundfrequenz (fundamental frequency)?
9. In welchem Bereich liegt etwa die Sprachgrundfrequenz von Männern/Frauen?
10. Nach welchen beiden messbaren Größen kann man die Vokale recht gut unterscheiden?
11. Was ist ein Phonem?
12. Was ist ein Allophon?
13. Was versteht man unter Koartikulation?
14. Wozu dient Prosodie bzw. Intonation?
15. Was versteht man unter Formanten?
16. Wie kann man aus dem Spektrum eines Sprachsignals die Grundfrequenz ermitteln?
17. Wie kann man aus einem Breitbandspektrogramm die Grundfrequenz ermitteln?
18. Wie kann man in einem Schmalbandspektrogramm die Grundfrequenz ermitteln?
19. Wie kann man direkt aus dem Zeitsignal die Grundfrequenz eines Sprachsignals ermitteln?

Mustererkennung

20. Womit beschäftigt sich das Gebiet der Mustererkennung (pattern recognition)?
21. Was versteht man unter der Klassifikation einfacher Muster (simple patterns)? Nennen Sie Beispiele?
22. Beschreiben Sie den grundsätzlichen Aufbau eines Klassifikationssystems!

Digitalisierung und Merkmale

23. Welche grundlegenden Entscheidungen müssen getroffen werden, bevor ein analoges Signal digitalisiert wird?
24. Was besagt das Abtasttheorem (sampling theorem)? Beispiel?
25. Welche Form hat i.d.R. die Kennlinie eines mit 8 Bit quantisierten Signals, und warum?
26. Worin liegt der Vorteil eines mit 8 Bit quantisierten Signals gegenüber einem mit 16 Bit quantisierten Signal?

27. Wie entsteht der sogenannte Leck-Effekt (spectral leakage) und wie lässt er sich reduzieren? Beispiel?
28. Welche typische Fenstergröße nutzt man in der automatischen Spracherkennung? Was wären die Vor- und Nachteile breiterer bzw. schmalerer Fenster?
29. Was ist der Unterschied zwischen einem Breitband- (wide band) und einem Schmalbandspektrogramm (narrow band spectrogram) und wozu nutzt man diese?
30. Welche Merkmale (features) werden in der automatischen Spracherkennung überwiegend eingesetzt? Wie errechnet man sie?
31. Welche Merkmale ergänzt man, um den zeitlichen Verlauf der MFCCs besser zu erfassen? Wie errechnet man sie?

Klassifikation

32. Welche Verfahren zur Klassifikation von Merkmalvektoren kennen Sie und wodurch zeichnen sich diese aus?
33. Beschreiben Sie den Nächster-Nachbar-Klassifikator!
34. Welche Formel ist bei der statistischen Klassifikation von zentraler Bedeutung? Erläutern Sie diese?
35. Welches Verfahren kennen Sie, mit denen man aus Stichproben von Merkmalsvektoren unüberwacht Kodebücher schätzen kann?
36. Wie funktioniert der k-Means-Algorithmus?
37. Was versteht man unter einer Gaußschen Mischverteilung (GMM)?
38. Wie funktioniert der EM-Algorithmus?

Deep Learning

1. Was ist ein Perzeptron?
2. Wie sehen die Schwellwert-Funktionen bei künstlichen neuronalen Netzen aus?
3. Was versteht man unter einem Feed-Forward-Netzwerk?
4. Was versteht man unter einem MLP?
5. In welcher Weise wirkt sich die Verwendung von neuronalen Netzen auf die Wahl von geeigneten Merkmalen für die Spracherkennung aus?

Dynamic Time Warping

6. Wozu dient der DTW-Algorithmus?
7. Erläutern Sie den DTW-Algorithmus!
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) .

Hidden-Markov-Modelle

9. Welche Parameter besitzt ein diskretes HMM?
10. Welche Arten von HMMs haben wir noch kennengelernt und worin unterscheiden sich diese?
11. Welche HMM-Topologien sind für die Spracherkennung geeignet?
12. Was versteht man unter der Produktionswahrscheinlichkeit und wie lässt sich diese naiv errechnen?

13. Wie ist die Grundidee eines effizienten Algorithmus zur Bestimmung der Produktionswahrscheinlichkeit?
14. Wozu dient der Viterbi-Algorithmus und worin besteht seine Grundidee?
15. Beschreiben Sie die Grundidee der Schätzung von HMM-Parametern anhand einer Beobachtungsfolge (Stichprobe)?
16. Was versteht man unter einer Maximum-Likelihood-Schätzung?
17. Welche Wortuntereinheiten nimmt man i.d.R. für die Spracherkennung?
Erläutern Sie die Idee dahinter!

Sprachmodelle (Language Models)

18. Was versteht man unter einem N-Gramm?
19. Wie erhält man ML-Schätzwerte für N-Gramme?
20. Warum sind diese in der Praxis von Nachteil?
21. Wie lassen sich die Schätzwerte glätten?
22. Welche weitere Möglichkeit gibt es, die Parameterzahl zu reduzieren?
23. Wie errechnet sich die sog. Test-Set-Perplexität einer Stichprobe, gegeben ein Sprachmodell?
24. Wie kann man diese interpretieren?
25. Wie kann man mit Sprachmodellen Themen (topics) klassifizieren?
26. Wie kann man mit Sprachmodellen Sprachen klassifizieren?
27. Wie lassen sich rekurrente Neuronale Netze als Sprachmodelle einsetzen?

Dekodierung

28. Was ist ein konfluenter Zustand (confluent state)?
29. Was versteht man unter Beam Search (Strahlsuche)?
30. Wie kann man bei sehr großen Wortschätzen den Wortschatz sinnvoll organisieren?
31. Welche weitere Methode zur Beschleunigung des Dekodiervorgangs haben wir kennengelernt?
32. Welche beiden heuristischen Parameter führt man ein, die von der reinen Lehre der Bayes-Formel abweichen?

End-to-End Deep Learning

33. Erläutern Sie die Grundidee von DeepSpeech!
34. In welcher Weise können Language Models bei diesem Ansatz in den Erkennungsvorgang einbezogen werden?