
Vorlesung Sprachverarbeitung

Übung 1

Einführung

1. Nennen Sie einige Beispiele für den Einsatz von Spracherkennungssystemen.
2. Worin unterscheiden sich die einzelnen Anwendungen, bzw. was gibt es für unterschiedliche Anforderungsbedingungen?
3. Skizzieren Sie die typische Block-Architektur eines statistischen Spracherkennungssystems. Erklären Sie die einzelnen Elemente.

Wahrscheinlichkeiten

1. Wie groß ist die Wahrscheinlichkeit von Kopf(K) oder Zahl(Z) beim Wurf einer Münze?
2. Die Münze werde dreimal hintereinander geworfen. Wie groß ist die Wahrscheinlichkeit für (KKK) und (ZKZ)?
3. Die Münze werde wieder dreimal geworfen. Wie groß ist die Wahrscheinlichkeit das zweimal (K) und einmal (Z) geworfen wird?
4. Wieder drei Würfe. Der erste Wurf zeige nun (K). Wie groß ist nun die Wahrscheinlichkeit insgesamt zweimal (K) und einmal (Z) zu werfen?

Sprachmodellierung

1. Fragen zur Sprachmodellierung
 - (a) Was ist ein Sprachmodell, wodurch unterscheiden sich Bi- und Trigramme?
 - (b) Was bedeutet Perplexität?
 - (c) Worauf muss bei der Erstellung von Sprachmodellen geachtet werden?
 - (d) In welchen Systemen können Sprachmodelle mit Historie sinnlos sein?
2. Gegeben sei der folgende Trainingscorpus:

ich schaue zum himmel. die sonne brennt und blendet. die luft ist heiß weil die sonne seit tagen vom himmel brennt. die haut brennt vom sand und brennt vom öl am strand.

 - (a) Wie groß ist das Vokabular V ?
 - (b) Wieviele Bi- und Trigramme sind theoretisch möglich?
 - (c) Berechnen Sie die Zeroogramm Wahrscheinlichkeiten
 - (d) Bilden Sie nun mit Hilfe von Zero-, Uni- und Bigrammen die Auftrittswahrscheinlichkeit für den folgenden Testcorpus:

die sonne brennt vom himmel
 - (e) Wie groß ist die Perplexität des Testcorpus für Zero-, Uni- und Bigramme?