Übungen Musterklassifikation
Sommersemester 2000
Prof. Dr. Stefan Posch

Institut für Informatik
Universität Halle

Blatt 9

Aufgabe 9.1

Betrachten Sie das Problem, durch Tippfehler "gestörte" bekannte Wörter mit Hilfe von Hidden-Markov-Modellen zu rekonstruieren. Als Fehler kommen Ersetzungen, Löschungen und Einfügungen einzelner Buchstaben in Frage.

(a) Was entspricht bei diesem Problem den Emissionen des Modells?

(b) Wie könnte die Modellierung aussehen, um aus fehlerhaften Schreibungen von `Markov` wie z.B. `Markoff` oder `Msrkow` auf die korrekte Orthographie rückschließen zu können?

Wie können die benötigten Modellparameter bestimmt werden?

(c) Bei der Rekonstruktion längerer Texte muß damit gerechnet werden, daß ggf. auch Wortzwischenräume fehlen und der Rückschluß daher *nicht* für jeweils ein einzelnes Wort erfolgen kann. Wie müssen die Modellbildung und das Textrekonstruktionsverfahren modifiziert werden?

Aufgabe 9.2

Bei der automatischen Erkennung gesprochener Sprache liegen *keine* diskreten symbolischen Einheiten als Observationssymbole vor.

Wie kann die oben gewählte einfache Modellierung prinzipiell verändert werden, um für die automatische Spracherkennung eingesetzt werden zu können?