12. Übung zur Theoretischen Informationstechnik I

Prof. Dr. Rudolf Mathar, Prof. Dr. Anke Schmeink, Andreas Bollig, Christoph Schmitz, Milan Zivkovic

03.02.2012

Aufgabe 1. Sie sollen die Buchstabenbelegung der Telefontastatur neu gestalten. Sie haben als Vorgabe die Buchstaben so auf die Tasten 2 bis 9 zu verteilen, dass für einen typischen deutschen Text die erwartete Anzahl an Tastendrücken minimal ist. Sie dürfen dabei beliebig viele Buchstaben auf eine Taste legen und die Reihenfolge der Buchstaben darf beliebig gewählt werden. Es werden nur die Buchstaben A bis Z unterstützt, d.h. auch keine Zahlen, Leerzeichen, Satzeichen oder sonstige Sonderzeichen. Benutzen Sie zur Lösung der Aufgabe die in Tabelle 1 gegebene Häufigkeitsverteilung der Buchstaben in der deutschen Sprache.
Hinweis: Das Auswählen eines der Buchstaben, die sich auf derselben Taste befinden, wird durch mehrfaches Drücken erreicht. Befinden sich also beispielsweise die Buchstaben A, B und C auf der Taste 2, so wählt man durch einfaches Drücken der Taste 2 den Buchstaben A aus, durch zweifaches Drücken den Buchstaben B und durch dreifaches Drücken den Buchstaben C.

a) Geben Sie eine optimale Tastenbelegung und begründen Sie die Optimalität. Wie hoch ist die mittlere Codelänge?

b) Wie hoch ist die Entropie der Quelle?

c) Interpretieren Sie die Ergebnisse.

Hinweis: Die Häufigkeiten existieren als Textdatei zum Download auf der Webseite sowie im L^{2}P Lernraum.

E	N	I	S	R	A	T	D	H	U	L	C	G
17,40	9,78	7,55	7,27	7,00	6,51	6,15	5,08	4,76	4,35	3,44	3,06	3,01
M	O	B	W	F	K	Z	P	V	J	Y	X	Q
2,53	2,51	1,89	1,89	1,66	1,21	1,13	0,79	0,67	0,27	0,04	0,03	0,02

Tabelle 1: Häufigkeitstabelle der Buchstaben in der deutschen Sprache in Prozent

Aufgabe 2. Es seien $\mathcal{X} = \{x_1, \dots, x_m\}$ ein Quellalphabet, $\mathcal{Y} = \{y_1, \dots, y_d\}$ ein Kodealphabet und g ein eindeutig dekodierbarer Kode. Für $j = 1, \dots, m$ bezeichne $P(X = x_j) = p_j$ die Wahrscheinlichkeit für das Auftreten des Quellbuchstabens x_j und n_j bezeichne die Länge des Kodeworts $g(x_j)$. Zeigen Sie (z.B. mit der Ungleichung von McMillan), dass für die erwartete Kodewortlänge $\bar{n}(g)$ gilt:

$$\bar{n}(g) = \frac{H(X)}{\log d} \quad \Leftrightarrow \quad p_j = d^{-n_j} \text{ für alle } j = 1, \dots, m \text{ mit } p_j > 0.$$

Aufgabe 3. Gegeben sei ein gedächtnisloser binärer symmetrischer Kanal mit Ein- und Ausgabealphabet $\mathcal{X} = \mathcal{Y} = \{0, 1\}$. Die Menge der Eingabewörter sei $\mathcal{C} = \{(0, 0, 0), (1, 1, 1)\}$. Dabei trete $(0, 0, 0)$ mit Wahrscheinlichkeit $1/4$ sowie $(1, 1, 1)$ mit Wahrscheinlichkeit $3/4$ auf. Ferner sei $\epsilon = 1/3$.

a) Welche Ausgabewörter sind bei der Übertragung im Kanal möglich und wie groß sind deren Auftrittswahrscheinlichkeiten?

b) Wie groß ist die Wahrscheinlichkeit p_K für die fehlerfreie Übertragung eines Eingabeworts?

c) Wie groß ist die Wahrscheinlichkeit p_E dafür, ein Element aus $\mathcal{C}$ zu empfangen, das nicht gesendet wurde?

d) Geben Sie eine ML-Dekodierung $h_{ML} : \mathcal{Y}^3 \to \mathcal{C}$ an.

e) Geben Sie eine ME-Dekodierung $h_{ME} : \mathcal{Y}^3 \to \mathcal{C}$ an.