12. Übung zur Theoretischen Informationstechnik I

Prof. Dr. Rudolf Mathar, Fabian Altenbach, Michael Reyer

29.01.2010

Aufgabe 1. Die diskreten Zufallsvariablen X_1 und X_2 seien identisch verteilt aber nicht notwendigerweise unabhängig. Ferner sei

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$

a) Zeigen Sie, dass $\rho = \frac{I(X_1,X_2)}{H(X_1)}$ gilt.

b) Zeigen Sie, dass $0 \leq \rho \leq 1$ gilt.

c) In welchem Fall gilt $\rho = 0$?

Aufgabe 2. Sei X eine diskrete Zufallsvariable mit Träger $\mathcal{X} = \{x_1, \dots, x_m\}$ und $g : \mathcal{X} \to \mathcal{Y}$, $\mathcal{Y} = \{y_1, \dots, y_n\}$, eine Abbildung. Zeigen Sie

$$H(g(X)) \leq H(X).$$

Wann gilt Gleichheit?

Aufgabe 3. Ein PF-Kode g^* mit den Wortlängen $n_1^*, \dots, n_m^*$ heißt *optimal*, wenn

$$\bar{n}(g^*) = \sum_{i=1}^{m} p_i n_i^* \leq \sum_{i=1}^{m} p_i n_i = \bar{n}(g)$$

für alle PF-Kodes g mit Wortlängen $n_1, \dots, n_m$ gilt.

a) Zeigen Sie, dass für einen optimalen Kode g^* folgendes gilt:

$$p_i > p_j \Rightarrow n_i^* \leq n_j^*.$$

Ein Algorithmus zur Konstruktion optimaler PF-Kodes ist das *Huffman-Verfahren*. In diesem Verfahren wird ein Binärbaum erzeugt, wobei die Blätter mit den Quellbuchstaben belegt werden. Der Weg von der Wurzel zu den Blättern legt das Kodewort für den Buchstaben fest, das heißt „links“ entspricht einer Null und „rechts“ einer Eins im Kodewort. Die Konstruktionsvorschrifr lautet folgendermaßen.

1. Liste die Symbole des Quellalphabets mit fallenden Wahrscheinlichkeiten auf und interpretiere die Symbole im Folgenden als Bäume.
2. Fasse die Bäume mit den geringsten Wahrscheinlichkeiten zu einem neuen Baum zusammen, d.h. an einen Knoten wird links bzw. rechts jeweils einer dieser Bäume gehangen und der Knoten (der neue Baum) wird mit der Summe der Wahrscheinlichkeiten markiert. Die ursprünglichen Bäume werden aus der Liste entfernt und der resultierende Baum wird gemäß seiner Wahrscheinlichkeit in die bisherige Liste einsortiert.
3. Wiederhole 2. bis nur noch ein Binärbaum übrig ist. Dieser repräsentiert den Huffmann-Code.

b) Bestimmen Sie für $\mathcal{X} = \{x_1, \ldots, x_8\}$ mit $p = (0.25, 0.2, 0.15, 0.15, 0.12, 0.05, 0.04, 0.04)$ und $\mathcal{Y} = \{0, 1\}$ einen optimalen PF-Kode.

c) Berechnen Sie die erwartete Kodewortlänge des Kodes und vergleichen Sie diese mit $H(X)$.

Aufgabe 4. 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.

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.

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