

Datenkompression

Sommersemester 2014

Übungsblatt 1

Prof. Dr. E. Ohlebusch

Institut für Theoretische Informatik

S. Arnold

Ausgegeben am 30.04.2014

Besprechung in der Übung am 08.05.2014

Aufgabe 1.1

Beantworten Sie die untenstehenden Fragen für folgende Mengen von Codewörtern:

$$M_1 := \{1, 00, 01, 101\},$$

$$M_2 := \{00, 11, 101, 1101\},$$

$$M_3 := \{00, 011, 10, 010, 11\},$$

$$M_4 := \{00, 11, 001, 110\}.$$

- Welche Mengen erfüllen die Kraft-McMillan-Ungleichung?
- Welche Mengen sind Bildbereich eines eindeutig dekodierbaren Codes, und warum?

Aufgabe 1.2

Geben Sie für das Alphabet $\{a, e, i, o, u\}$ Präfixcodes mit den Längenfunktionen l_1 bzw. l_2 an!

x	a	e	i	o	u
$l_1(x)$	4	1	3	3	4
$l_2(x)$	2	2	3	3	2

Aufgabe 1.3

Auf dem Alphabet $\{a, e, i, o, u\}$ seien zwei Wahrscheinlichkeitsfunktionen P, P' definiert:

x	a	e	i	o	u
$P(x)$	1/8	1/4	1/8	7/16	1/16
$P'(x)$	5/32	3/16	5/32	3/8	1/8

- Bestimmen Sie für beide Fälle einen Präfixcode nach dem Shannon-Verfahren.
- Bestimmen Sie für beide Fälle einen Präfixcode nach dem Shannon-Fano-Verfahren.
- Bestimmen Sie für beide Fälle einen Präfixcode nach dem Huffman-Verfahren.
- Berechnen Sie die erwarteten Codewortlängen der Codes aus den Teilaufgaben a) bis c).

Aufgabe 1.4

Beim Shannon-Fano-Algorithmus werden wiederholt Listen (c_1, \dots, c_n) von Symbolen in zwei Teile $M_1 = (c_1, \dots, c_k)$ und $M_2 = (c_{k+1}, \dots, c_n)$ gespalten. Unter Umständen existiert aber eine andere Partitionierung $\tilde{M}_1 \cup \tilde{M}_2 = \{c_1, \dots, c_n\}$, $\tilde{M}_1 \cap \tilde{M}_2 = \emptyset$, die „gleichmäßiger“ ist, also: $|P(\tilde{M}_1) - P(\tilde{M}_2)| < |P(M_1) - P(M_2)|$. Wir betrachten eine Variante des Algorithmus, die unter *allen* Partitionierungen stets eine „gleichmäßigste“ wählt.

- Geben Sie eine Wahrscheinlichkeitsfunktion P an, bei der der modifizierte Algorithmus einen Code mit größerer erwarteter Codewortlänge generiert als der Shannon-Fano-Algorithmus.
- (*schwieriger!*) Geben Sie ein Beispiel an, bei dem der modifizierte Algorithmus einen Code mit kleinerer erwarteter Codewortlänge generiert als der Shannon-Fano-Algorithmus.