

Datenkompression

Sommersemester 2016

Übungsblatt 2

Prof. Dr. E. Ohlebusch

Institut für Theoretische Informatik

J. Lorenz

Ausgegeben am 03.05.2016

Besprechung am 10.05.2016

Aufgabe 2.1

Bestimmen Sie für $\Sigma = \{a, b, c, d\}$ mit $P(a) = 0,7$ und $P(b) = P(c) = P(d) = 0,1$ einen Huffmancode. Wie viele Huffmancodes gibt es insgesamt für P ?

Wie kann sichergestellt werden, dass der Decodierer den gleichen Huffmancode verwendet, wie der Codierer und welche Information(en) müssen hierfür zusätzlich übertragen werden?

Aufgabe 2.2

Gegeben sei die folgende Wahrscheinlichkeitsfunktion. Berechnen Sie die Entropie und vergleichen Sie das Ergebnis mit den mittleren Codewortlängen aus Aufgabe 1.2.

x	d	i	l	m	n	r	u
P(x)	1/24	1/24	1/6	7/24	1/12	1/24	1/3

Aufgabe 2.3

Auf einem Alphabet Σ sei die Wahrscheinlichkeitsfunktion $P: \Sigma \rightarrow [0, 1]$ definiert. Für Zeichenketten aus Σ^m , deren Zeichen unabhängig voneinander gezogen werden, erhalten wir die Wahrscheinlichkeitsfunktion $P_m: \Sigma^m \rightarrow [0, 1]$, $(c_1, \dots, c_m) \mapsto P(c_1) \cdots P(c_m)$.

Beweisen Sie die Additivität der Entropie: $H(\Sigma^m, P_m) = mH(\Sigma, P)$.

Aufgabe 2.4

Schreiben Sie ein Programm, das für einen Eingabestring die mittlere Codewortlänge bei Shannon-Codierung, Shannon-Fano-Codierung bzw. Huffman-Codierung ausgibt. Ihr Programm sollte die relativen Buchstabenhäufigkeiten in der Eingabe als Wahrscheinlichkeitsverteilung verwenden.

Verwenden Sie hierbei diejenige Variante der Shannon-Fano-Codierung, bei der Listen von absteigend sortierten Wahrscheinlichkeiten in zwei Teile gespalten werden.

Aufgabe 2.5

Die Zeichen des Alphabetes $\{a, b, c\}$ mit der Anordnung $a < b < c$ treten mit den Wahrscheinlichkeiten $P(a) = 0,3$, $P(b) = 0,5$ und $P(c) = 0,2$ auf.

a) Berechnen Sie die arithmetische Codierung des Wortes `cabb` ($m = 4$).

b) Decodieren Sie ein Wort der Länge $m = 6$ aus dem Codewort `0100000000`.