Aufgaben zum Tut am 25.06.2007

Thomas Pajor

26. Juni 2007

Aufgabe 1.

(a) Entwickeln Sie einen Huffman–Code zu der Zeichenkette „abracadabra“

(b) Berechnen Sie die relative Redundanz Ihres Codes

(c) Welche relative Redundanz hätte ein Shannon–Code für diese Zeichenkette?

Es vermöge nun der Huffman–Code aus der (Muster–)Lösung von Aufgabe (a).

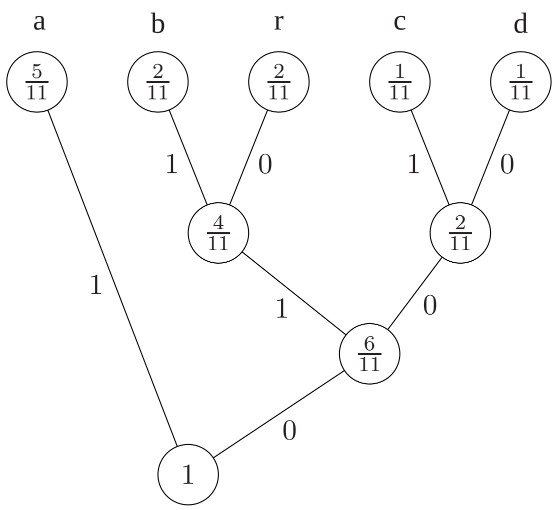
(d) Dekodieren Sie „000110111010001“

Lösung.

(a) Aus der Zeichenkette „abracadabra“ ergibt sich sofort das Alphabet $\mathcal{X} := \{a, b, c, d, r\}$. Eine Quellenstatistik liefert für die Zufallsvariable X die Wahrscheinlichkeitsverteilung $p(x)$ wie folgt

$$p(a) = \frac{5}{11} \qquad p(b) = p(r) = \frac{2}{11} \qquad p(c) = p(d) = \frac{1}{11}$$

Damit können wir einen Huffman–Baum nach dem in der Vorlesung beschriebenen Verfahren konstruieren



Damit folgt nun der Huffman–Code durch ablesen der Beschriftung an den Kanten auf dem Weg von der Wurzel zu den jeweiligen Blättern:

$$\begin{aligned} C(a) &= 1 \\ C(b) &= 011 \\ C(c) &= 001 \\ C(d) &= 000 \\ C(r) &= 010 \end{aligned}$$

(b) Wir wollen nun die relative Redundanz ausrechnen. Es gilt $l(a) = 1$ und $l(b, c, d, r) = 3$. Damit ist die Nominalinformation

$$\begin{aligned} L(C) &= \sum_{x \in \mathcal{X}} p(x) l(x) \\ &= \frac{5}{11} + 2 \cdot 3 \cdot \frac{2}{11} + 2 \cdot 3 \cdot \frac{1}{11} \\ &= \frac{23}{11} \text{ bit} \end{aligned}$$

Die Entropie von X ist[1]

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \approx 2.04 \text{ bit}$$

Die absolute Redundanz R_{abs} ist demnach

$$R_{\text{abs}} = L(C) - H(X) \approx 0.05 \text{ bit}$$

[1]Man benutze einen Rechenknecht, oder in einer Klausur die Logarithmustabellen

also sehr sehr gering. Die relative Redundanz, die eine anschauliche Aussage trifft, da sie normiert ist, liefert

$$R_{\text{rel}} = \frac{R_{\text{abs}}}{L(C)} = 1 - \frac{H(X)}{L(C)} \approx 0.024 = 2.4\%$$

(c) Nun wollen wir einen Shannon–Code C_2 benutzen um $\mathcal{X}$ zu kodieren. Die Länge $l_2(x)$ für ein Zeichen $x \in \mathcal{X}$ ist dabei gerade

$$l_2(x) := \left\lceil \log \frac{1}{p(x)} \right\rceil$$

Das liefert $l_2(a) = 2$, $l_2(b) = l(r) = 3$ und $l_2(c) = l_2(d) = 4$. Man sieht jetzt schon dass die Kodewortlängen schlechter sind als beim Huffman–Code aus Aufgabe (a). Wir berechnen die Nominalinformation

$$L(C_2) = 2 \cdot \frac{5}{11} + 2 \cdot 3\frac{2}{11} + 2 \cdot 4\frac{1}{11} = \frac{30}{11} \text{ bit} \approx 2.72 \text{ bit}$$

Die Entropie ist natürlich die gleiche wie oben, also gilt für die absolute und relative Redundanzen

$$R_{\text{abs}} = L(C_2) - H(X) \approx 0.686 \text{ bit} \qquad R_{\text{rel}} = 1 - \frac{H(X)}{L(C_2)} \approx 25.1\%$$

Man sieht jedoch, dass die Ungleichung

$$H(X) \leq L(C_2) < H(X) + 1$$

trotzdem erfüllt wird von dem Shannon–Code!

(d) Das Kodewort „000110111010001“ kann eindeutig dekodiert werden zu „000 1 1 011 1 010 001“ und damit zu „daabarc“.