Informationstheorie, FS 2021 Prof. Joachim M. Buhmann
ETH Zürich, Dept. Informatik Fabian Laumer

Lösungen zu Übungsblatt 7

7.1

In dieser Aufgabe geht es um die Decodierung von Lempel-Ziv-Codes. Gehen Sie davon aus, dass der Decodierer den vom Codierer verwendete Code $\phi : \mathcal{X} \to \{0,1\}^*$ für das "letzte Symbol" kennt, wobei $\mathcal{X} = \{s_1, \ldots, s_M\}$ das Eingabealphabet ist.

a) Geben Sie den allgemeinen Algorithmus für die Decodierung eines Codeworts $\mathbf{b} \in \{0,1\}^*$ an. *Hinweis:* Überlegen Sie sich zunächst, nach wie vielen Symbolen des Codeworts jeweils ein $|$ eingefügt werden muss.

Wir bezeichnen mit D_i die Version des beim Codieren entstandenen Wörterbuchs nach dem i-ten Schritt, das also ausser dem leeren Wort alle Phrasen bis einschliesslich der i-ten enthält; es gilt $|D_i| = i + 1$. Solange $2^{n-1} + 1 \leq |D_i| \leq 2^n$ erfüllt ist (was äquivalent zu $2^{n-1} \leq i \leq 2^n - 1$ ist), können wir die $(i+1)$-te Phrase mit $n + k$ Bits codieren, wobei $k = \lceil \log M \rceil$. Daraus ergibt sich die Folge der Längen der den Phrasen entsprechenden Abschnitte in $\mathbf{b}|$) : Der erste Abschnitt hat k Bits, dann folgen für jedes n mit $1 \leq n \leq n_{\max}$ genau 2^{n-1} Abschnitte mit $n + k$ Bits, wobei $n_{\max} = \max\{N \mid k + \sum_{n=1}^{N-1} 2^{n-1}(n+k) < |\mathbf{b}|\}$, und am Ende entweder noch maximal $2^{n_{\max}-1}$ Abschnitte mit $n_{\max}+k$ Bits oder maximal $2^{n_{\max}-1}-1$ Abschnitte mit $n_{\max}+k$ bits und ein Abschnitt mit $n_{\max}$ Bits.

Auf der Vorlesungs-Website findet sich eine Python-Implementierung des Decodierungsalgorithmus basierend auf diesen Überlegungen.

b) Decodieren Sie 00100101010011001001000, für den Fall, dass $\mathcal{X} = \{a, b, c\}$ und dass $\phi : \mathcal{X} \to \{0,1\}$ durch $a \mapsto 00$, $b \mapsto 01$, $c \mapsto 10$ gegeben ist.

Die Aufteilung des zu decodierenden Strings in die den Phrasen entsprechenden Abschnitte ist

$$, 00|1, 00|10, 10|10, 01|100, 10|010, 00.$$

Das decodierte Wort ist $a|aa|aac|aab|aabc|aaa$ (ohne die $|$).

7.2

Berechnen Sie die Kapazität einer verrauschten Schreibmaschine mit Alphabet $\{A, \ldots, Z, \square\}$ und $P(Y = \square | X = A) = P(Y = A | X = A) = P(Y = B | X = A) = \frac{1}{3}$ etc. (wie in der Vorlesung). Geben Sie mindestens zwei Verteilungen P_X über die Eingabe an, die die Kapazität erreichen.

Für jede Verteilung P_X über die Eingabe gilt

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) \\ &= H(Y) - \sum_x P_X(x) H(Y|X=x) \\ &\leq \log 27 - \log 3 \\ &= \log 9, \end{aligned}$$

denn $H(Y) \leq \log 27$ und $H(Y|X=x) = \log 3$ für alle x. Mit der Gleichverteilung $P_X(x) = \frac{1}{27}$ wird diese Schranke erreicht, und damit ist die Kapazität des Kanals gleich $\log 9$.

Eine weitere Verteilung P_X, die die Kapazität erreicht, erhält man, indem man 9 Symbole wählt, für die sich die Mengen der möglichen Ausgabesymbole nicht überlappen (z.B. $A, D, G, \dots$), und $P_X(x) = \frac{1}{9}$ für all diese Symbole und $P_X(x) = 0$ für alle übrigen setzt. Dann gilt $H(X) = \log 9$ und $H(X|Y=y) = 0$ für alle y, und folglich $I(X;Y) = \log 9$.

7.3

In der Vorlesung haben wir gesehen, dass die Kapazität des binären Auslöschungskanals mit Auslöschungswahrscheinlichkeit p durch $C = 1 - p$ gegeben ist. Zeigen Sie, dass bei Verwendung von Feedback (d.h. der Empfänger kann die nochmalige Übertragung eines ausgelöschten Bits verlangen) tatsächlich durchschnittlich $\frac{1}{1-p}$ Bits gesendet werden müssen, um ein Quellbit eindeutig decodieren zu können.

Die Wahrscheinlichkeit, dass ein Quellbit nach dem n-ten Übertragung erstmals nicht ausgelöscht wird, beträgt $p^{n-1}(1-p)$. Damit ist die erwartete Anzahl der Wiederholungen, die nötig sind, bis ein Quellbit eindeutig decodiert werden kann, gegeben durch

$$\sum_{n=1}^{\infty} n p^{n-1}(1-p) = \frac{1}{1-p},$$

wobei wir die Identität $\sum_{n=1}^{\infty} n p^n = \frac{p}{(1-p)^2}$ verwendet haben (vgl. Aufgabe 2.3). Im Schnitt muss jedes Quellbit also $\frac{1}{1-p}$ mal gesendet werden.

7.4

Betrachten Sie zwei Kanäle K_1 und K_2 mit Kapazitäten C_1 bzw. C_2. Diese beiden Kanäle können parallel als ein Kanal K verwendet werden, der als Eingabe (X_1, X_2) nimmt und als Ausgabe (Y_1, Y_2) liefert (*Parallelschaltung* von K_1 und K_2). Sei C die Kapazität von K.

Die Kanäle sind in Abbildung **??** illustriert. Man beachte, dass dies nicht dasselbe ist wie ein Kanal mit $|\mathcal{X}_1| + |\mathcal{X}_2|$ Input- und $|\mathcal{Y}_1| + |\mathcal{Y}_2|$ Outputsymbolen.

a) Beweisen Sie $I(X_1, X_2; Y_1, Y_2) = I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2)$

Der Kanal K wird definiert durch die durch $P(y_1, y_2|x_1, x_2) = P(y_1|x_1)P(y_2|x_2)$ gegebene Familie von bedingten Verteilungen. Die gemeinsame Verteilung von X_1, X_2, Y_1, Y_2 erfüllt daher $P(x_1, x_2, y_1, y_2) = P(x_1, x_2)P(y_1|x_1)P(y_2|x_2)$. Daraus folgt insbesondere, dass $P(x_2, y_1|x_1) = P(x_2|x_1)P(y_1|x_1)$ ist, d.h. $X_2 \rightarrow X_1 \rightarrow Y_1$ ist eine Markovkette

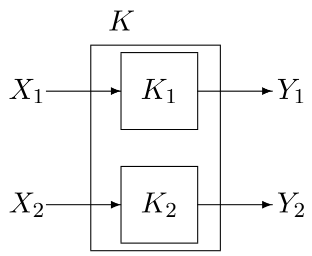

Abbildung 1: Zwei Kanäle in Parallelschaltung

und somit $I(X_2; Y_1|X_1) = 0$, was äquivalent zu $I(X_1, X_2; Y_1) = I(X_1; Y_1)$ ist. Damit erhalten wir zunächst

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\ &= I(X_1; Y_1) + I(X_1, X_2; Y_2|Y_1), \end{aligned}$$

wobei im ersten Schritt die Kettenregel zum Einsatz kam. Zu zeigen ist also noch, dass $I(X_1, X_2; Y_2|Y_1) = I(X_2; Y_2) - I(Y_1; Y_2)$ gilt; das folgt aus

$$\begin{aligned} I(X_1, X_2; Y_2|Y_1) + I(Y_1; Y_2) &= I(X_1, X_2, Y_1; Y_2) \\ &= I(X_1, Y_1; Y_2|X_2) + I(X_2; Y_2) \\ &= I(X_2; Y_2), \end{aligned}$$

wobei wir in den ersten beiden Schritten die Kettenregel verwendet haben und im letzten Schritt die Tatsache, dass $(X_1, Y_1) \to X_2 \to Y_2$ eine Markovkette ist (das sieht man wiederum anhand der obigen Formel für die gemeinsame Verteilung).

b) Beweisen Sie $C = C_1 + C_2$.

Mit a) folgt

$$\begin{aligned} C &= \max_{P_{X_1 X_2}} I(X_1, X_2; Y_1, Y_2) \\ &= \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\leq \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) \\ &\leq \max_{P_{X_1 X_2}} I(X_1; Y_1) + \max_{P_{X_1 X_2}} I(X_2; Y_2) \\ &= C_1 + C_2. \end{aligned}$$

Andererseits folgt mit a) auch

$$\begin{aligned} C &= \max_{P_{X_1 X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &\geq \max_{P_{X_1} P_{X_2}} I(X_1; Y_1) + I(X_2; Y_2) - I(Y_1; Y_2) \\ &= \max_{P_{X_1}} I(X_1; Y_1) + \max_{P_{X_2}} I(X_2; Y_2) \\ &= C_1 + C_2. \end{aligned}$$

Dabei haben wir ausgenutzt, dass das Maximum nur kleiner werden kann, wenn wir anstatt über alle Verteilungen $P_{X_1X_2}$ nur über diejenigen Verteilungen maximieren, bei denen X_1 und X_2 unabhängig sind. In diesem Fall sind auch Y_1 und Y_2 unabhängig und damit ist $I(Y_1, Y_2) = 0$.

Insgesamt folgt also $C = C_1 + C_2$.