3. Woche
Information, Entropie, Kryptographische Codierung

Informationsgehalt einer Nachricht

Intuitiv: Je kleiner die Quellws, desto „wichtiger“ oder „strukturierter“ die Information, bzw. höher der Informationsgehalt.

Zur Struktur:

Ws von „A“ ist 6,51%, Ws von „t“ ist 6,15%, Ws von „e“ ist 17,40% und Ws von „m“ ist 2,53 %.

Falls Buchstaben unabhängig (gedächtnisloser Kanal), ist Ws von „Atem“ gleich $0,0651 \cdot 0,0615 \cdot 0,174 \cdot 0,0253 = 1,76 * 10^{-5} = 0.00176\%$

(Eigentlich ist die deutsche Sprache kein gedächtnisloser Kanal: Nach den Daten im Leipziger Wortschatz: Ws von „Atem“ $< 0.0008\%$.)

Intuitiv betrachtet, enthält „Atem“ mehr Information als „A“, „t“, „e“, oder „m“.

Wir wollen eine Informations–„abmessende“ Funktion nach dieser Intuition definieren.

Informationsgehalt einer Nachricht

Forderungen für eine Funktion, die Information „abmessen“ soll.

1. $I(p) \geq 0$: Der Informationsgehalt soll positiv sein.
2. $I(p)$ ist stetig in p: Kleine Änderungen in der Ws p sollen nur kleine Änderungen von $I(p)$ bewirken.
3. $I(p_i) + I(p_j) = I(p_i p_j)$:
   - X = Ereignis, dass a_i und a_j nacheinander übertragen werden. Annahme war: gedächtnislose Quelle, also unabhängige Ereignisse.
   - Informationsgehalt von X: $I(p_i) + I(p_j)$, $\mathrm{Ws}(X) = p_i p_j$

Satz zur Struktur von $I(p)$

Jede Funktion $I(p)$ für $0 < p \leqslant 1$, die obige Bedingungen erfüllt, ist der Form

$$I(p) = C \log \frac{1}{p}$$

für eine positive Konstante C.

Beweis: Form von $I(p)$

- Forderung 3 liefert $I(p^2) = I(p) + I(p) = 2\,I(p)$.
- Induktiv: $I(p^n) = n\,I(p)$ für alle $n \in \mathbb{N}$ und alle $0 < p \leqslant 1$.
- Substitution $p \to p^{\frac{1}{n}}$ liefert: $I(p) = nI\left(p^{\frac{1}{n}}\right)$ bzw. $I\left(p^{\frac{1}{n}}\right) = \frac{1}{n}\,I(p)$
- Damit gilt für alle $q \in \mathbb{Q}$: $I(p^q) = q\,I(p)$.
- Für jedes $r \in \mathbb{R}$ gibt es eine Folge q_i mit $\lim\limits_{n\to\infty} q_n = r$.
  Aus der Stetigkeit von $I(p)$ (Annahme!) folgt

$$I(p^r) = I\Big(\lim_{n\to\infty} p^{q_n}\Big) = \lim_{n\to\infty} I\left(p^{q_n}\right) = \lim_{n\to\infty} q_n I(p) = r\,I(p)$$

- Fixiere $0 < q < 1$. Für jedes $0 < p \leq 1$ gilt

$$I(p) = I\left(q^{\log_q p}\right) = I(q)\log_q p = -I(q)\log_q\left(\frac{1}{p}\right) = -I(q)\frac{\log_2 \frac{1}{p}}{\log_2 q}$$

$$= \boldsymbol{C\log_2\frac{1}{p}} \quad \text{wobei} \quad C = -I(q)\cdot\frac{1}{\log_2(q)} > 0\ .$$

□

Definition Information $I(p)$

Definition $I(p)$

Die Information $I(p)$ eines Symbols mit Quellws $p > 0$ beträgt

$$I(p) = \log_2 \frac{1}{p}.$$

Die Einheit der Information bezeichnet man als Bit.

Hier ist der Logarithmus zur Basis 2 gemeint (auch wenn nicht explizit geschrieben).

Beispiele für Information

- $Q = \{0, 1\}$ mit $p_1 = p_2 = \frac{1}{2}$. Dann ist $I(\frac{1}{2}) = 1$, d.h. für jedes gesendete Symbol erhält der Empfänger 1 Bit an Information.
- $Q = \{0, 1\}$ mit $p_1 = 1, p_2 = 0$. Dann ist $I(1) = 0$, d.h. der Empfänger enthält 0 Bit an Information pro gesendetem Zeichen.
- Beweis für „Ein Bild sagt mehr als 1000 Worte!“
  - Beamer-Bild: Auflösung $1024 * 768$, 24-Bit Farben
    Ssenario: Der Dozent will Informationsfunktionen erklären.
    - $2^{1024*768*24}$ mögliche Folien. Annahme: Jede gleich wahrscheinlich!
    - Information in Bit: $I(2^{-1024*768*24}) = 1024 * 768 * 24 = 18.874.368$
  - Selbe Erklärung in Worten im Buch:
    $\leq$ 1000 Worte, $\leq$ 10.000 Worte Vokabular
    - Information: $I(10.000^{-1000}) < 13.288$

Entropie einer Quelle

Definition Entropie einer Quelle

Sei Q eine Quelle mit Quellws $P = \{p_1, \ldots, p_n\}$. Die Entropie von Q ist

$$H(Q) = \sum_{i=1}^{n} p_i I(p_i) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = -\sum_{i=1}^{n} p_i \log p_i \quad .$$

- Für $p_i = 0$ definieren wir $p_i \log \frac{1}{p_i} = 0$. Also ist $p_i \log \frac{1}{p_i} = 0$ gdw $p_1 = 0$ oder 1.
- Entropie ist die durchschnittliche Information pro Quellsymbol.
  - $P = \{\underbrace{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}}_{n \text{ Symbole}}\} : H(Q) = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n$
  - $P = \{\underbrace{\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}}_{n \text{ Symbole}}, 0\} : H(Q) = \sum_{i=1}^{n} \frac{1}{n} \log n = \log n$
  - $P = \{1, 0, 0, \ldots, 0\} : H(Q) = 1 \cdot \log 1 = 0$

N.B.: man kann $H(Q)$ als auch $H(P)$ für eine Quelle Q mit Ws-Verteilung P schreiben.

Sätze über Entropie und Codewortlänge

Unser nächstes Ziel ist, folgende Sätze zu beweisen:

Satz Schranken für $H(Q)$

Sei Q eine Quelle mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Dann gilt

$$0 \leqslant H(Q) \leqslant \log n.$$

Weiterhin gilt $H(Q) = \log n$ gdw alle $p_i = \frac{1}{n}$ für $i = 1, \ldots, n$ und $H(Q) = 0$ gdw $p_i = 1$ für ein $i \in \{1, \ldots, n\}$.

D.h. die Beispiele der vorigen Folie sind extremal.

Codierungstheorem von Shannon (1948)

Sei Q eine Quelle für $\{a_1, \ldots, a_n\}$ mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Sei C ein kompakter Code für Q. Dann gilt für die erwartete Codewortlänge

$$H(Q) \leqslant E(C) < H(Q) + 1.$$

Wechsel zu anderer Ws-Verteilung

Lemma Wechsel Ws-Verteilung

Seien $P = \{p_1, \ldots, p_n\}$ eine Ws-Verteilung und $Q = \{q_1, \ldots, q_n\}$ mit $\sum_{i=1}^{n} q_i \leqslant 1$. Dann gilt

$$\sum_{i=1}^{n} p_i I(p_i) \leqslant \sum_{i=1}^{n} p_i I(q_i).$$

Gleichheit gilt genau dann, wenn $p_i = q_i$ für alle $i = 1, \ldots, n$.

Nützliche Ungleichung für das Rechnen mit logs:

$$x - 1 \geqslant \ln x = \log x \cdot \ln 2 \qquad \text{für alle } x > 0$$

Gleichheit gilt gdw $x = 1$.
$\ln(x)$ beweichnet den natürlichen Logarithmus, $\log(x)$ ist zur Basis 2.

Beweis des Lemmas

$$\begin{aligned}
\sum_{i=1}^{n} p_i I(p_i) - \sum_{i=1}^{n} p_i I(q_i) &= \sum_{i=1}^{n} p_i \left(\log \frac{1}{p_i} - \log \frac{1}{q_i} \right) \\
&= \sum_{i=1}^{n} p_i \log \frac{q_i}{p_i} \\
&\leqslant \frac{1}{\ln 2} \sum_{i=1}^{n} p_i \left(\frac{q_i}{p_i} - 1 \right) \\
&= \frac{1}{\ln 2} \left(\underbrace{\sum_{i=1}^{n} q_i}_{\leqslant 1} - \underbrace{\sum_{i=1}^{n} p_i}_{=1} \right) \leqslant 0 \ .
\end{aligned}$$

Gleichheit gilt gdw $\frac{q_i}{p_i} = 1$ für alle $i = 1, \ldots, n$. □

Untere und obere Schranken für $H(P)$

Satz Schranken für $H(P)$

Sei Q eine Quelle mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Dann gilt

$$0 \leqslant H(Q) \leqslant \log n.$$

Weiterhin gilt $H(Q) = \log n$ gdw alle $p_i = \frac{1}{n}$ für $i = 1, \ldots, n$ und $H(Q) = 0$ gdw $p_i = 1$ für ein $i \in \{1, \ldots, n\}$.

- Sei $P' = \{\frac{1}{n}, \ldots, \frac{1}{n}\}$ die Gleichverteilung.
- Nach Lemma zum Wechsel von Ws-Verteilungen gilt

$$H(Q) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leqslant \sum_{i=1}^{n} p_i \log \frac{1}{p'_i} = \log n \sum_{i=1}^{n} p_i = \log n.$$

- Gleichheit gilt gdw $p_i = p'_i = \frac{1}{n}$ für alle i.

Untere Schranke für $H(P)$

Verwenden Ungleichung $\log x \geqslant 0$ für $x \geqslant 1$. Gleichheit gilt gdw $x = 1$. Nun,

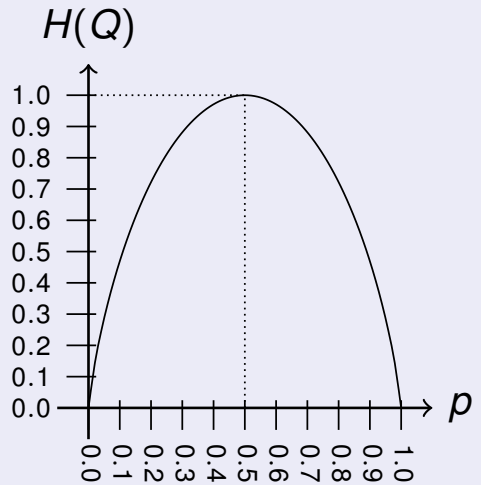
$$H(Q) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \geqslant 0,$$

mit Gleichheit gdw alle $p_i \log \frac{1}{p_i} = 0$, d.h. $p_i = 1$ für ein $i \in \{1, \ldots, n\}$ und $p_j = 0$ für alle $j \neq i$. □

Beispiel

Binäre Quelle $Q = \{a_1, a_2\}$ mit $P = \{p, 1-p\}$

$$H(Q) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \quad .$$



$H(Q)$ heißt *binäre Entropiefunktion*.

Codierungstheorem von Shannon

Codierungstheorem von Shannon (1948)

Sei Q eine Quelle für $\{a_1, \ldots, a_n\}$ mit Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Sei C ein kompakter Code für Q. Dann gilt für die erwartete Codewortlänge

$$H(Q) \leqslant E(C) < H(Q) + 1.$$

Beweis: $H(Q) \leqslant E(C)$

- Bezeichnen Codewortlängen $\ell_i := |C(a_i)|$ und $q_i := 2^{-\ell_i}$.
- Satz von McMillan: $\sum_{i=1}^{n} q_i = \sum_{i=1}^{n} 2^{-\ell_i} \leqslant 1$.
- Lemma Wechsel Ws-Verteilung liefert

$$\begin{aligned} H(Q) &= \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq \sum_{i=1}^{n} p_i \log \frac{1}{q_i} \\ &= \sum_{i=1}^{n} p_i \log 2^{\ell_i} = \sum_{i=1}^{n} p_i \ell_i = E(C). \end{aligned}$$

$E(C) < H(Q) + 1$

- Seien $\ell_1, \ldots, \ell_n$ Codewortlängen mit $\sum_{i=1}^{n} 2^{-\ell_i} \leqslant 1 = \sum_{i=1}^{n} p_i$.
- Satz von McMillan garantiert Existenz von Code C' für

$$2^{-\ell_i} \leqslant p_i \quad \Leftrightarrow \quad -\ell_i \leqslant \log p_i \quad \Leftrightarrow \quad \ell_i \geqslant \log \frac{1}{p_i}.$$

- Wählen $\ell_i \in \mathbb{N}$ für alle i minimal mit obiger Eigenschaft, d.h.

$$\log \frac{1}{p_i} \leqslant \ell_i < \log \frac{1}{p_i} + 1.$$

  Ein Code C' mit dieser Eigenschaft heißt *Shannon-Fano Code*.
- Für jeden kompakten Code C gilt

$$\begin{aligned} E(C) \leqslant E(C') &= \sum_{i=1}^{n} p_i \ell_i < \sum_{i=1}^{n} p_i \left(\log \frac{1}{p_i} + 1 \right) \\ &= \sum_{i=1}^{n} p_i \log \frac{1}{p_i} + \sum_{i=1}^{n} p_i = H(Q) + 1 \ . \end{aligned}$$

□

Besser als kompakte Codes?

Nun wissen wir:

1. Die Huffman-Codierung liefert einen kompakten Code C.
2. Für diesen code C gilt $H(Q) \leqslant E(C) < H(Q) + 1$.

Falls wir Pech haben ist $E(C)$ fast gleich $H(Q) + 1$.

Können wir Codes bauen, die eine „kompaktere als eine kompakte“ Codierung liefern?

Wir fangen mit einem Beispiel an.

Codieren einer binären Quelle

Szenario: Binäre Quelle Q mit $P = \{\frac{1}{4}, \frac{3}{4}\}$ mit

$$H(Q) = \frac{1}{4} \cdot \log 4 + \frac{3}{4} \cdot \log \frac{4}{3} \approx 0.811.$$

- Huffman-Codierung von Q:
  $C(a_1) = 0, C(a_2) = 1$ mit $E(C) = 1$.

- **Problem:** Jedes Symbol hat Codewortlänge $\geqslant 1$. Also $E(C) \geqslant 1$.
  Wie kann mann die Ungleichheit der Wahrscheinlichkeiten ausnutzen, um die erwartete Codewortlänge zu reduzieren?

- **Idee:** Codieren *Zweierblöcke* von Quellsymbolen.

Quellenerweiterungen von Q

- Betrachten $Q^2 = \{a_1a_1, a_1a_2, a_2a_1, a_2a_2\}$ mit Quellws

$$p_1 = \frac{1}{16}, p_2 = p_3 = \frac{3}{16}, p_4 = \frac{9}{16}.$$

- Huffmann-Codierung von Q^2 liefert

$$C(a_1a_1) = 000, C(a_1a_2) = 001, C(a_2a_1) = 01, C(a_2a_2) = 1$$

mit $E(C) = 3 \cdot \frac{4}{16} + 2 \cdot \frac{3}{16} + \frac{9}{16} = \frac{27}{16}$.

- Jedes Codewort codiert zwei Quellsymbole, d.h. die durchschnittliche Codewortlänge pro Quellsymbol ist

$$E(C)/2 = \frac{27}{32} = 0.84375.$$

- *Übung:* Für Q^3 erhält man 0.82292.
  Wie viel können wir dies noch verbessern?
  Wie nah kommen wir zur Entropie?

k-te Quellenerweiterung Q^k

Definition k-te Quellenerweiterung

Sei Q eine Quelle mit Alphabet $A = \{a_1, \ldots, a_n\}$ und Ws-Verteilung $P = \{p_1, \ldots, p_n\}$. Die k-te Quellenerweiterung Q^k von Q ist definiert über dem Alphabet A^k, wobei $a = a_{i_1} \ldots a_{i_k} \in A^k$ die Quellws $p_{i_1} \cdot p_{i_2} \cdot \ldots \cdot p_{i_k}$ besitzt.

Satz Entropie von Q^k

Sei Q eine Quelle mit k-ter Quellenerweiterung Q^k. Dann gilt

$$H(Q^k) = k \cdot H(Q).$$

Beweis für $H(Q^k)$

$$
\begin{aligned}
H(Q^k) &= \sum_{(i_1,\ldots,i_k)\in[1..n]^k} p_{i_1}\cdots p_{i_k} \log \frac{1}{p_{i_1}\cdots p_{i_k}} \\
&= \sum_{(i_1,\ldots,i_k)\in[1..n]^k} p_{i_1}\cdots p_{i_k} \left(\log \frac{1}{p_{i_1}} + \cdots + \log \frac{1}{p_{i_1}}\right) \\
&= \sum_{(i_1,\ldots,i_k)\in[1..n]^k} p_{i_1}\cdots p_{i_k} \log \frac{1}{p_{i_1}} + \cdots + \sum_{(i_1,\ldots,i_k)\in[1..n]^k} p_{i_1}\cdots p_{i_k} \log \frac{1}{p_{i_k}}
\end{aligned}
$$

Betrachten wir den ersten Summanden

$$
\begin{aligned}
\sum_{(i_1,\ldots,i_k)\in[1..n]^k} p_{i_1}\cdots p_{i_k} \log \frac{1}{p_{i_1}} &= \sum_{i_1\in[1..n]} p_{i_1} \log \frac{1}{p_{i_1}} \cdot \sum_{i_2\in[1..n]} p_{i_2} \cdots \sum_{i_k\in[1..n]} p_{i_k} \\
&= \sum_{i_1\in[1..n]} p_{i_1} \log \frac{1}{p_{i_1}} \cdot 1 \cdots 1 = H(Q).
\end{aligned}
$$

Analog liefern die anderen $n-1$ Summanden jeweils $H(Q)$. $\square$.

Anwendung des Satzes von Shannon auf Quellenerweiterungen

Korollar zu Shannons Codierungstheorem

Sei Q eine Quelle mit k-ter Quellenerweiterung Q^k. Sei C ein kompakter Code für Q^k. Dann gilt

$$H(Q) \leqslant \frac{E(C)}{k} < H(Q) + \frac{1}{k}.$$

- Anwendung von Shannon's Codierungstheorem auf Q^k liefert

$$H(Q^k) \leqslant E(C) < H(Q^k) + 1.$$

  Anwenden von $H(Q^k) = kH(Q)$ und teilen durch k liefert die Behauptung.
- Für wachsenden k, $E(C)/k$ wird beliebig nah zu $H(Q)$.

Bedingte Entropie

- Sei X, Y Zufallsvariablen
- Definieren $\mathrm{Ws}(x) = \mathrm{Ws}(X = x)$ und $\mathrm{Ws}(x, y) = \mathrm{Ws}(X = x, Y = y)$.
- X, Y heißen unabhängig $\Leftrightarrow \mathrm{Ws}(x, y) = \mathrm{Ws}(x) \cdot \mathrm{Ws}(y)$

Definition Bedingte Entropie

Wir bezeichnen die Größe $H(Y \mid X)$

$$:= \sum_x \mathrm{Ws}(x) H(Y|X = x) = \sum_x \mathrm{Ws}(x) \left(\sum_y \mathrm{Ws}(y|x) \log \frac{1}{\mathrm{Ws}(y|x)} \right)$$

$$= \sum_x \sum_y \mathrm{Ws}(x, y) \log \frac{1}{\mathrm{Ws}(y|x)}$$

als bedingte Entropie von Y gegeben X.

Eigenschaften bedingter Entropie

Rechenregeln für die bedingte Entropie

1. Kettenregel:
   $H(X,Y) = \sum_x \sum_y \mathrm{Ws}(x,y) \log \frac{1}{\mathrm{Ws}(x,y)} = H(X) + H(Y \mid X)$
   (Übung)
2. $H(Y \mid X) \leqslant H(Y)$. Gleichheit gilt gdw X, Y unabhängig sind.
   (ohne Beweis)
3. Folgerung aus 1. und 2.: $H(X,Y) \leqslant H(X) + H(Y)$.

Kryptographische Codierung

Szenario:

- Drei Klartexte: a, b, c mit Ws $p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$.
- Zwei Schlüssel k_1, k_2 gewählt mit Ws jeweils $\frac{1}{2}$
- Verschlüsselungsfunktionen:
  - $e_{k_1}: a \mapsto d,\ b \mapsto e,\ c \mapsto f$
  - $e_{k_2}: a \mapsto d,\ b \mapsto f,\ c \mapsto e$

- Seien P,C Zufallsvariablen für den Klar- und Chiffretext.
- Erhalten Chiffretext d, Plaintext muss a sein.
- Erhalten Chiffretext e, Plaintext muss b oder c sein.

$$\mathrm{Ws}(b \mid e) = \frac{\mathrm{Ws}(b \cap e)}{\mathrm{Ws}(e)} = \frac{\mathrm{Ws}(b \cap k_1)}{\mathrm{Ws}(b \cap k_1) + Ws(c \cap k_2)} = \frac{0.15}{0.15 + 0.1} = 0.6$$

Lernen Information über zugrundeliegenden Klartext.

- $H(P) = \sum_i p_i \log \frac{1}{p_i} = 1.485$ und $H(P \mid C) = 0.485$.
- D.h. für gegebenen Chiffretext sinkt die Unsicherheit.

Perfekte Sicherheit, das One-Time Pad

Definition Perfekte Sicherheit

Ein Kryptosystem ist perfekt sicher, falls $H(P \mid C) = H(P)$.

One-Time Pad

- Plaintextraum $\mathcal{P}$: $\{0,1\}^n$ mit Ws-Verteilung $p_1, \ldots, p_{2^n}$
- Schlüsselraum $\mathcal{K}$: $\{0,1\}^n$ mit Ws $\frac{1}{2^n}$ für alle Schlüssel
- Verschlüsselung: $c = e_k(x) = x \oplus k$ für $x \in \mathcal{P}, k \in \mathcal{K}$.

Satz One-Time Pad

Das One-Time Pad ist perfekt sicher.

Beweis: Übung.