

Zusammenfassung Elektronik – Informationsgehalt und Redundanz

Alphabete: $A = \{A_i \mid i = 1 \dots N\}$ N ist die Anzahl der unterscheidbaren Zeichen

Informationsvorrat Q: $Q = N^L$ Informationsvorrat = Anzahl der Zeichen hoch Wortlänge
 Beispiel: Dezimalzahlensystem mit Wortlänge 3:
 $N = 10, L = 3 \quad Q = 10^3$

Wortlänge L: $L = \log_b N$ Wortlänge = Logarithmus der Basis zur Anzahl der Zeichen
 (bei fester Wortlänge) Beispiel: Binärcodierung eines Zeichensatzes mit 50 Zeichen
 $B = 2, N = 50 \quad L = \log_2 50 = 5,64 \rightarrow$ Wortlänge ist 6 Zeichen

Alphabete als „endliches Schema“ (= Wahrscheinlichkeitsfeld):

$$A = \begin{pmatrix} A_1 & A_2 & A_3 & \dots & A_N \\ P(A_1) & P(A_2) & P(A_3) & \dots & P(A_N) \end{pmatrix}$$

A = Zeichen, P(A_i) = Auftretenswahrscheinlichkeit des Zeichens

$$\sum P(A_i) = 1 \quad (100\%)$$

Summe aller einzelnen Wahrscheinlichkeiten ergibt 1

Beispiel: $A = \begin{pmatrix} A & B & C & D \\ 0,25 & 0,125 & 0,5 & 0,125 \end{pmatrix}$
 $0,25 + 0,125 + 0,5 + 0,125 = 1$

Informationsgehalt H:

$$H(A_i) = \text{Id} \frac{1}{P(A_i)}$$

Informationsgehalt H eines Zeichens = Zweier-Logarithmus von Eins durch die Wahrscheinlichkeit des Zeichens

Beispiel:

$$H(A) = \text{Id} 1/0,25 = \text{Id} 4 = \text{Id} 2^2 = 2 \text{ bit (d.h. ich brauche zwei Binärstellen, um das Zeichen A in der Datenquelle zu codieren!)}$$

$$H(B) = \text{Id} 1/0,125 = \text{Id} 8 = \text{Id} 2^3 = 3 \text{ bit}$$

$$H(C) = \text{Id} 1/0,5 = \text{Id} 2 = \text{Id} 2^1 = 1 \text{ bit}$$

Mittlerer Informationsgehalt = Entropie

$$H(A) = \sum P(A_i) * H(A_i)$$

Mittlerer Informationsgehalt = Summe aus allen einzelnen Wahrscheinlichkeiten mal alle einzelnen Informationsgehalte

Beispiel:

$$\begin{aligned} H(A) &= P(A) * H(A) + P(B) * H(B) + P(C) * H(C) + P(D) * H(D) \\ &= 0,25 * 2 + 0,125 * 3 + 0,5 * 1 + 0,125 * 3 \\ &= 1/2 + 3/8 + 1/2 + 3/8 = 1,75 \text{ bit} \end{aligned}$$

→ Man sollte nach einer Codierung suchen, die 1,75 bit nicht wesentlich überschreitet. In diesem Falle wäre das eine Codierung mit 2 Binärstellen pro Zeichen:

Zeichen:	A	B	C	D
Code 1:	00	01	10	11
Alternativer Code 2:	01	001	1	000

(Codierung nach H(A_i))

Wenn alle Zeichen gleich wahrscheinlich sind (= Datenquelle hat maximale Entropie), dann gilt: $H(\mathbf{A}) = H(A_i) = \text{Id } N$
 Beispiel: Datenquelle mit 64 Zeichen und maximaler Entropie:
 $H(\mathbf{A}) = \text{Id } 64 \rightarrow 2^6 = 64 \rightarrow H(\mathbf{A}) = H(A_i) = 6 \text{ bit}$

Mittlere Wortlänge:

$$L_m = \sum P(A_i) * L(A_i)$$

Mittlere Wortlänge = Summe aus allen einzelnen Wahrscheinlichkeiten mal alle einzelnen Wortlängen

Beispiel:

$$L_m = 0,25 * 2 + 0,125 * 3 + 0,5 * 1 + 0,125 * 3 = 1,75$$

Renundanz R (= ungenutzter Informationsgehalt):

$$R = L - H(\mathbf{A})$$

Renundanz = Mittlere Wortlänge minus Mittlerer Informationsgehalt

Beispiel: $R = 2 - 1,75 = 0,25 \text{ bit Renundanz pro Zeichen bei Code 1}$

$R = 1,75 - 1,75 = 0 \text{ bit Renundanz pro Zeichen bei Code 2}$

Wann tritt Renundanz 0 auf?

- Wenn das Auftreten aller Zeichen gleich wahrscheinlich ist:

$$P(A_i) = \frac{1}{N}$$

N = Anzahl der Zeichen

Beispiel: $N = 4 \rightarrow P(A_i) = \frac{1}{4}$

- Wenn die Wortlänge der Potenz b des Zeichens entspricht:

$$L = b$$

b = zur Codierung benötigte bit

Beispiel: $H(B) = 3 \text{ bit}$ Codierung: $B = 001$

Renundanz R: Länge 3 - Informationsgehalt 3 = 0

- Wenn

Der „schnelle Informationsgehalt“:

$$H(A_i) = \text{Id } \frac{1}{P(A_i)}$$

Beispiel: $P(B) = 1/8$

$$H(A_i) = \text{Id } \frac{1}{1/8} = \text{Id } 8 = 3$$

Schneller: Nimm Wahrscheinlichkeit des Zeichens: 1/8

Davon Kehrwert: 8

Davon Id: $2^3 = 8 \rightarrow \text{Id } 8 = 3$