
CHAPTER 1

INFORMATION THEORY / CODING

1	Information Theory	2
2	Huffman	4
3	Lempel-Ziv	4
4	Kanalmodell	4

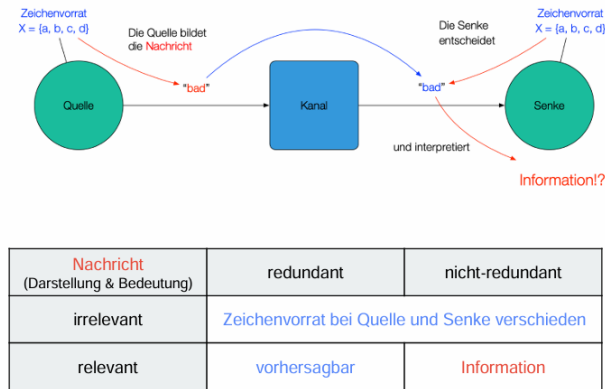


Figure 1.1: Information Theory Term Overview

1 Information Theory

Definition 1.1: Decision Content / Entscheidungsgehalt

The measure of the effort required to form a message or to decide on a decision of a message is the decision content.

$$H_0 = \log_2(N)[\text{bit}] \quad (1.1)$$

Definition 1.2: Decision Flow / Entscheidungsfluss

The formula measures the rate of information flow necessary to make a decision over time.

$$H_0 = \frac{\log_2(N)}{\tau} \quad (1.2)$$

Where τ is the time required for the transmission of a word.

Definition 1.3: Information Content

The information content of a sign indicates how many elementary decisions have to be made to determine this sign.

$$I(x_k) = \log_2\left(\frac{1}{P(x_k)}\right) \quad (1.3)$$

Definition 1.4: Entropy (mittlerer Informationsgehalt)

Entropy is the average information content of the source. It therefore shows how many elementary decisions the source/sink has to make on average per character

$$H(X) = \sum_{k=1}^N P(x_k) \cdot I(x_k) = \sum_{k=1}^N p(x_k) \cdot \log_2\left(\frac{1}{P(x_k)}\right) \quad (1.4)$$

Since we typically use a binary encoding for the discrete values of the source we use:

$$L(X) = \sum_{k=1}^N P(x_k) \cdot I_r(x_k) \quad (1.5)$$

where $I_r = I(x_k)$ aufgerundet (auch tatsächliche Codewortlänge)

Definition 1.5: Source Redundancy (Redundanz der Quelle)

$$R_Q = H_0 - H(X) \quad (1.6)$$

Definition 1.6: Prefix Code

A prefix code is a code that fulfills the Fano condition: No code word of the code is a prefix of another code word. In other words, no code word may be the beginning of another code word. For example, a code with the code words $\{0, 10, 11\}$ fulfills the prefix property, whereas the code with the code words $\{0, 01, 10\}$ does not, as 0 is the prefix of 01.

Theorem 1.7: Shannons Theorem

Für jede beliebige zugehörige Binärcodierung mit Präfixeigenschaft ist die mittlere Codewortlänge nicht kleiner als die Entropie $H(X)$.

$$H(x) \leq L \quad (1.7)$$

Für jede beliebige Quelle kann eine Binärcodierung gefunden werden, so dass die folgende Ungleichung gilt:

$$H(X) \leq L \leq H(X) + 1 \quad (1.8)$$

Definition 1.8: Code Redundancy

$$R_C = L - H(X) \quad (1.9)$$

Definition 1.9: Discrete Source with and without memory

Die mittlere Entropie einer Quelle ohne Gedächtnis ist stets grösser oder gleich der Entropie einer Quelle mit Gedächtnis. In der Quellencodierung sind daher nicht Einzelzeichen zu codieren, sondern stets Zeichenketten.

$$R_Q = H_0 - H_{oG}(X) \leq H_- - H_{mG}(X) \quad (1.10)$$

2 Huffman

Verfahren zur Entwicklung eines kommafreien Codes mit minimaler mittlerer Codewortlänge
Rekursives Verfahren, d.h. der Binärbaum wird nicht von der Wurzel, sondern von den Blättern aus entwickelt

3 Lempel-Ziv

Example : Algorithm

1. Suche in der Tabelle eine möglichst lange Zeichenfolge, die mit den nächsten N zu kodierenden Zeichen übereinstimmt.
2. Bilde ein Token und speichere es.
3. Verschiebe das Fenster um $N + 1$ Zeichen.
4. Wiederhole, bis alle Zeichen kodiert sind.

4 Kanalmodell

TODO