Grundlagen der Informationstheorie

Konzepte und Prinzipien

Digitale AV Technik, MIB 5

Was ist Information?

- Information ist die **Reduktion von Unsicherheit**.
- Claude Shannon definierte Information als eine Größe, die quantifizierbar ist.

Was enthält Information?

- **Nachrichten** enthalten Information.
  - Eine Nachricht kann z.B. ein Text in deutscher Sprache sein.
  - Ein Binärcode enthält nur Nullen und Einsen, diese bilden eine Nachricht.

Formale Definition einer Nachricht

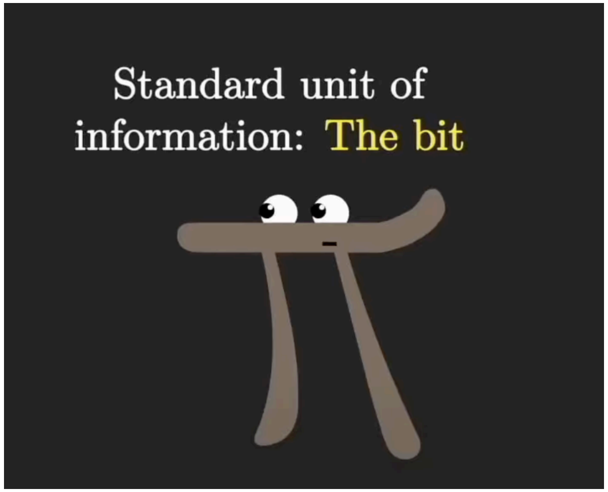
- Eine Nachricht besteht aus beliebig vielen Zeichen hintereinander. Also zum Beispiel:
  - AABACAABADDA
  - 001110101001
  - Hallo, das ist eine Nachricht.
- Eine Nachricht muss mit einem beliebigen Alphabet codiert werden, um übertragen werden zu können.

Nachrichtentechnik

- **Nachricht:** Folge von Zeichen, die von Sender (Quelle) an Empfänger (Senke) übermittelt wird.
- **Information:** intuitiv: alles was durch eine Nachricht übermittelt wird, Bedeutung der Nachricht
- **Informationstheorie nach Shannon:** Versuch, den Begriff der Information rein statistisch zu erfassen

Was ist ein Bit?

Wie viel Information steckt in einem Bit?



[image from 3Blue1Brown]

Informationsgehalt

Der **Informationsgehalt eines Zeichens** x mit einer Auftrittswahrscheinlichkeit $P(x)$ ist definiert als

$$I(x) = \log_a\left(\frac{1}{P(x)}\right) = \log_a(1) - \log_a(P(x)) = -\log_a(P(x))$$

- mit a als Kardinalität des Alphabets
  - also der Anzahl der möglichen Zeichen
  - also $a = 2$ bei binären Codes.

Beispiel: Informationsgehalt eines Buchstaben

In deutschen Texten tritt das B mit der Wahrscheinlichkeit 0.016 auf und wir wollen es binär codieren, also

$I(b) = \log_2(\frac{1}{0.016}) \approx 5.97$ Bit

Bemerkung: $\log_b(x) = \log_{10}(x)/\log_{10}(b)$

Maßeinheit des Informationsgehalts: Bit (Binary Digit)

- 1 Bit = Entscheidung zwischen zwei gleich wahrscheinlichen Alternativen
  - z.B. Ja/Nein, 0/1, True/False
- Pseudoeinheit für die Anzahl der Stellen in Binärdarstellung
  - daraus folgt $P(x) = 2^{-x} = \frac{1}{2^x}$

Beispielrechnung

- Eine 8 Bit Darstellung ermöglicht 256 verschieden Zeichen.
- Bei gleicher Wahrscheinlichkeit tritt jedes Zeichen x mit $P(x) = 2^{-8} = \frac{1}{2^8} = \frac{1}{256}$ auf.
- Der Informationsgehalt ist daher $I(x) = -\log_2(P(x)) = -\log_2(\frac{1}{256})$ $= -\log_2(\frac{1}{2^8}) = \log_2(2^8) = 8$ Bit.

Shannons Entropie

- **Entropie (H)**: Maß für den mittleren Informationsgehalt pro Zeichen einer **Nachricht**.
- Je höher die Entropie, desto **mehr Information** trägt eine Nachricht.
  - **Achtung:** es geht nur um den statistischen Informationsgehalt, nicht um die Bedeutung.

Formel der Entropie

$$H(X) = \sum_{i=1}^{n} P(x_i) I(x_i) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

- H: die Entropie einer (binären) Nachricht X
- $P(x_i)$: Wahrscheinlichkeit des Auftretens des Zeichen x_i

Beispiel: Entropie einer Nachricht

- Beispiel: Ein Alphabet mit 4 Buchstaben (A, B, C, D) und der Nachricht AABBCDCD
  - gleichverteilte Wahrscheinlichkeit: $P(A) = P(B) = P(C) = P(D) = 0.25$
- Entropie:

$$H = -\left(0.25\log_2 0.25 + 0.25\log_2 0.25 + 0.25\log_2 0.25 + 0.25\log_2 0.25\right) = 2\ \text{Bits}$$

- Um diese Nachricht optimal zu codieren, benötigt man 2 Bit pro Zeichen.

Beispiel 2: Entropie einer ähnlichen Nachricht

- Beispiel: Ein Alphabet mit 4 Buchstaben (A, B, C, D) und der Nachricht AAAABCCD
  - nicht gleich verteilte Wahrscheinlichkeiten: $P(A) = 0.5, P(B) = P(D) = 0.125, P(C) = 0.25$
- Entropie:

$$H = -\left(0.5 \log_2 0.5 + 0.125 \log_2 0.125 + 0.25 \log_2 0.25 + 0.125 \log_2 0.125\right) = 1.75 \text{ Bits}$$

- Um diese Nachricht optimal zu codieren, benötigt man 2 Bit pro Zeichen.

Entropie intuitiv

Anzahl der Bits, die notwendig sind, um eine Nachricht in einem (bzgl. der Wortlänge optimalen) Code binär zu codieren.

- hängt von der Auftrittswahrscheinlichkeit der Zeichen ab
  - hier Wahrscheinlichkeit = relative Häufigkeit

Informationsquelle und Redundanz

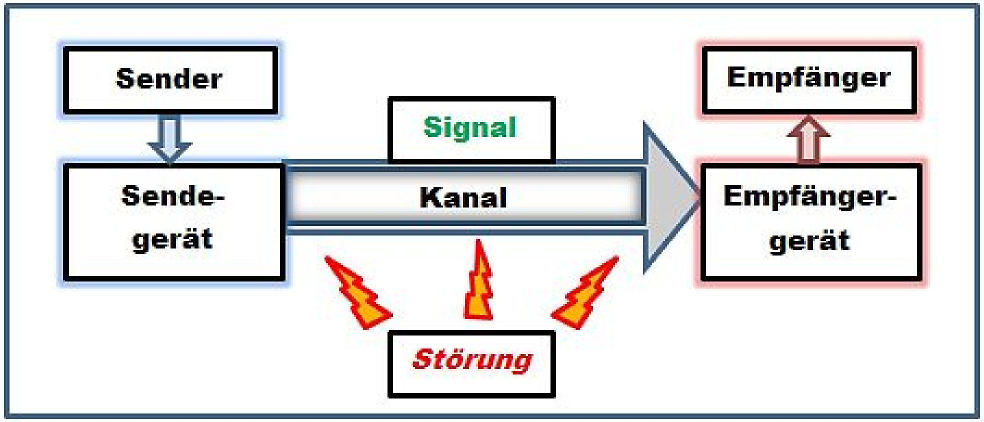
- **Redundanz**: Wiederholung von Information zur Fehlervermeidung.
- Nachrichten enthalten oft **mehr Daten**, als zur Übertragung nötig sind.
- Redundanz hilft bei der **Fehlerkorrektur** und **Kompression**.

Folgen der Redundanz

- Eine wiederholte Nachricht ("AAAAA") hat geringe Entropie, weil sie wenig Unsicherheit enthält.
- Nachrichten mit **hoher Entropie** sind schwerer zu komprimieren.

Sender - Empfänger Modell

- Das Modell beschreibt die **Übertragung von Informationen** in einem System.
- Benannt nach Shannon und Weaver, 1948 veröffentlicht in "A Mathematical Theory of Communication" von C. Shannon



Shannon-Weaver Modell Komponenten

- **Informationsquelle**: Erzeugt die Nachricht
- **Sender**: Kodiert die Nachricht in Signale
- **Kanal**: Überträgt die Signale (z.B. Kabel, Funk)
- **Empfänger**: Dekodiert das Signal
- **Ziel**: Erhält die Nachricht
- **Störungen**: Fehler oder Rauschen während der Übertragung

Kanal-Kapazität

- **Kanal-Kapazität (C)**: Maximale Datenmenge, die über einen Kommunikationskanal übertragen werden kann.

Shannons Kanal-Kapazitätstheorem

Die Kapazität eines Übertragungskanals berechnet sich als $C = B \log_2 \left(1 + \frac{S}{N}\right)$, wobei

- B, die Bandbreite des Kanals ist und
- $\frac{S}{N}$ das Signal-Rausch-Verhältnis (SNR)

Bedeutung des Theorems

- Die Kapazität gibt an, wie viel Information ohne **Fehler** übertragen werden kann.
- Höheres Signal-Rausch-Verhältnis verbessert die Kapazität.

Beispiel: Kanal-Kapazität

- **Mobilfunknetz**: Bandbreite von 10 MHz, Signal-Rausch-Verhältnis von 15.

$$C = 10 \times \log_2 (1 + 15) \approx 40 \text{ Mbit/s}$$

- Diese maximale Kapazität kann nur erreicht werden, wenn die **Fehlerkorrektur** optimal funktioniert und das Rauschen minimal ist.

Datenkompression

- **Ziel**: Reduktion der **Datenmenge**, ohne Information zu verlieren.
- Komprimierte Daten haben eine **niedrigere Redundanz**.

Verlustfreie Kompression

- Alle Daten bleiben vollständig erhalten (z.B. ZIP, PNG).

Verlustbehaftete Kompression

- Redundante oder unwichtige Daten werden verworfen (z.B. MP3, JPEG).

Beispiel

- Eine Nachricht mit hoher Redundanz kann stärker komprimiert werden als eine zufällige Nachricht.

Fehlerkorrektur

- In der Praxis gibt es bei der Übertragung oft **Fehler** (z.B. durch Rauschen).
- **Fehlerkorrigierende Codes** helfen, diese Fehler zu erkennen und zu beheben.

Shannons Theorem zur Fehlerkorrektur

- Claude Shannon zeigte, dass es **theoretisch möglich** ist, eine Nachricht **fehlerfrei** zu übertragen, wenn die Datenrate unterhalb der **Kanal-Kapazität** liegt.

Grenzen der Fehlerkorrektur:

- Bei hoher Rauschbelastung wird die Übertragung schwieriger.
- Es gibt eine **Grenze**, ab der auch Fehlerkorrekturmethoden nicht mehr helfen können.

Anwendungen der Informationstheorie

- **Datenübertragung**: Internet, Mobilfunk, Satellitenkommunikation.
- **Datenkompression**: MP3, JPEG, ZIP.
- **Kryptographie**: Sicherung der Kommunikation durch Verschlüsselung.
- **Fehlerkorrektur**: DVDs, CDs, moderne Speichertechnologien.

Claude Shannon an der Tafel

Zusammenfassung der Konzepte

1. **Information und Entropie**: Maß für den Informationsgehalt.
2. **Redundanz**: Hilft bei der Fehlervermeidung und Kompression.
3. **Kanal-Kapazität**: Maximale Übertragungsrate ohne Fehler.
4. **Datenkompression**: Reduktion der Datenmenge.
5. **Fehlerkorrektur**: Methoden zur fehlerfreien Übertragung.